The Use of Spatial Decompositions for Constructing Street Centerlines

John Radke and Anders Flodmark
Geographic Information Science Center
University of California, Berkeley, CA 94720-2600, USA.

Abstract
Although national data sets are becoming readily available at low cost, scale usually limits their utility for planning and managing small municipalities. As a result, most communities are faced with the construction of their own municipal Geographic Information Systems (GIS), information systems that are critical in handling land-related activities where high accuracy is essential. Most small municipalities cannot afford to begin by commissioning a large scale cadastral map and thus must opt for spatially questionable facsimiles where surveys showing administrative boundaries, property lines and street centerlines are suspect. The accuracy in these data can be enhanced and the results of great value to most city operations. We introduce a new method that applies theoretically based spatial decompositions to automate the generation of street centerlines from spatially corrected block and parcel data. This new centerline data base is a vast improvement over existing data bases for most municipalities.

I. INTRODUCTION

The relatively young field of Geographic Information Science has had significant impact on the development of techniques to capture, measure, store, manage and analyze geographic information within an urban environment. The amount of data that is captured and catalogued in a municipality is now quite voluminous and increasing each day as government departments go completely digital. The invention of technologies such as Remote Sensing, Global Positioning Systems (GPS) and hand held computers or data loggers, produces more data ready for immediate online access than ever before. Much of the data gathered within the municipality can be geographically related rendering location as a unique data field for relating data that might otherwise go unlinked. However, methods of managing these data are often antiquated and land related data held by one agency or department is frequently inaccessible by another. (Zhou, 1995). For example, some departments use one set of base maps for spatial encoding data while others use a different source, often of different scale and accuracy. Some departments catalogue data by tax map, block and parcel, while others rely on street address. To continue these efforts results in no value added through the integration of data and no synthesis can emerge to better plan and manage the municipality.

A Cadastral map or survey showing administrative boundaries and property lines is usually the most detailed and accurate land information available for a municipality and can provide a large-scale base to which other layers of data can be registered or added. (Zhou, 1995) Although the Cadastral map may provide the ultimate registration base, it is generated from an accurate land survey and is usually the most tedious and expensive to produce. For most small to medium size municipalities, starting with a digital Cadastre is not a viable option. It is likely the majority of municipal Spatial Decision Support Systems (SDSS) (Malczewski, 1997) that build upon a base of street centerline, block and parcel information, do not require the accuracy of a Cadastral base and can be effective employing a close facsimile. How then might these municipalities produce a relatively accurate yet cost effective georeferenced data base to facilitate the integration of data and serve the majority of their information management needs?

II. OBJECTIVES

The central objective of this study is to develop a new low cost method which integrates and rectifies non-georeferenced parcel map tiles to an accurate surveyed set of benchmarks and employs theoretically based geometric structures to extract the form of a set of points to automatically define and generate street centerlines. This new centerline generator can be considered an alternative to the existing generators which often rely heavily on human operator intervention, leading to long delays during the construction of municipal SDSS and often postponement. Although there have been several attempts to automate the construction of street centerlines (Ladak and Martinez, 1996; Christensen, 1996; East, 1997), prob-

1082-4006/99/0501-15$3.00
©1999 The Association of Chinese Professionals in Geographic Information Systems (Abroad)
lems still remain correcting for intersections and cul-
deads with most solutions still needing software
ehnancements. The new method proposed here char-
acterizes the endoskeleton (Radke, 1988) of a set of
points by exploiting notions of relative proximity and
neighborliness. These notions are present in two op-
erations in the method, the construction of the
Delaunay Triangulation (DT) and the more rigorous
and quantifiable β-skeleton which delineates a wide
spectrum of possible skeletal structures to construct
the street centerline.

Before we propose a method which has foundations
in computational morphology (Toussaint, 1980), we
summarize descriptors of shape that decompose point
sets into summary bounding shapes, shape hulls
(Toussaint, 1980) or exoskeletons, and internal line
structures referred to as skeletons by Toussaint (1980),
but better described here as endoskeletons. Next the
theoretical models that define neighborhood and con-
struct endoskeletons as part of the process to deline-
ate street centerlines are developed and explained.
Finally, the method developed here is described and
applied to a municipal data set in the City of Berke-
ley, California to demonstrate its sensitivity and ro-
bustness. We argue the resultant data base provides
an inexpensive but accurate skeleton upon which a
robust municipal SDSS can be constructed to serve
most of the City’s application needs.

III. SPATIAL DECOMPOSITIONS

The spatial decomposition of data is becoming more
common (Li, 1984; Radke, 1988, Okabe et al, 1992)
with many tools now packaged in popular software.
The locational characteristics of observations are ab-
stracted and encoded as spatial data models with
points, lines and polygons (cells) characterizing the
spatial extent of data (Goodchild, 1992). Further de-
marcation within these data models is valuable and
can take the form of vertical or horizontal spatial data
analysis (Gong, 1994), common practice in the pro-
cessing of data in geographic information science.

Decomposing a point set into simple line or polygon
sets can result in perceptually meaningful shapes or
structures that better describe the original point set’s
morphology. These new structures can provide bet-
ter descriptors of form, or anthropomorphic decomposi-
tions as they are referred to by Pavlidis (1977), be-
tween neighboring points. Essential or extreme
points in the pattern can be used to decompose and detect
the geometrical properties of the points set under-
study.

This paper is about understanding the neighborliness
and characteristics of points, lines and polygons to
aid in the automatic generation of street centerlines.
We undertake a number of spatial processes decom-
posing polygons to lines and then into points, regen-
erating simple polygons based on notions of neigh-
borly, decomposing those simple polygons into lines
and then essential points, and finally decomposing
these essential points into lines which form, for the
most part, the street centerlines. Streets can vary in
complexity from straight to those with extreme curves.
Points demarcating street centers possess the same
characteristics which make it difficult for a single
shape descriptor to be considered the best for all pos-
sible applications. We generate a number of solutions
that describe the internal set of essential points and
form descriptors of the street centerline. The method
is robust and can consider a variety of street curves
while generating street centerline best fits. The
method draws from both internal (Endoskeleton-
graphs) and external (exoskeleton-hulls) theoretically
based shape descriptors.

Exoskeleton-hulls

The simplest exoskeleton decompositions of a set of
points describe very general geometric constructs.
The minimum bounding box, minimum bounding
circle (Freeman and Shapiro, 1975; Toussaint and
Bhattacharya, 1981) or its generalized minimum
bounding ellipse (Kirkpatrick and Radke, 1985) all
provide a crude first approximation of the global shape
of a point set. A more sensitive descriptor of global
shape is the convex hull or the minimum convex poly-
gon that contains the entire point set (Toussaint,
1980). A generalization of the convex hull introduced
by Edelsbrunner et al (1983) introduces a parameter-
ized notion of a family of α-hulls where shapes, es-
pecially cruder and finer than the convex hull, can be
defined.

Endoskeleton-graphs

The simplest endoskeleton decompositions of a set of
points that could be considered a shape descriptor is
the nearest neighbor graph (NNG) which most often
results in an unconnected graph with many spatial
subsets. If we connect all the subsets in the NNG
with the minimal total edge length, the minimum
spanning tree (MST) results which can be considered
the minimal skeleton of the point set. Increasing the
complexity of the link structure and allowing circuit
disks to form, gives the skeleton a more expansive
shape by connecting more essential neighbors. One
such endoskeleton, the relative neighborhood graph
(RNG) (Lankford, 1969), produces edges linking rela-
tive neighbors V₁ and V₂ if their lune, the region of
influence formed by the intersection of two circles of
radius \( d(V_i, V_j) \) and centered at \( V_i \) and \( V_j \), is empty. A conceptually similar graph, the Gabriel graph (GG) (Matula and Sokal, 1980), links Gabriel neighbors \( V_i \) and \( V_j \) if their disc, the circle of influence with radius \( d(V_i, V_j) / 2 \) which passes through both \( V_i \) and \( V_j \), is empty. All of these internal shape descriptors are subsets of the Delaunay Triangulation, as \( \text{NNG} \subseteq \text{MST} \subseteq \text{RNG} \subseteq \text{GG} \subseteq \text{DT} \) (Figure 1).

The Delaunay Triangulation, the maximal planar description of internal structure in a point set, is a popular decomposition that along with its combinatorial dual, the Voronoi diagram, has had one of the greatest unifying effects of all graphs studied in computational geometry and has many interesting properties and applications (Shamos and Hoey, 1975; Getis and Boots, 1978; Okabe et al, 1992). The DT can be efficiently computed in \( O(n \log n) \) time as it is made up of edges that join all Voronoi neighbors embedded in a plane (Toussaint, 1980). Two points \( V_i \) and \( V_j \), from a point set \( s \), are Voronoi neighbors and define an edge of the DT if there exists a point \( x \) in a plane for which \( d(x, V_i) = d(x, V_j) = \min(d(x, V) \mid V \in s) \).

The DT can also be computed, although not as efficiently, using as a generative property the notion of empty neighborhoods similar to those used by the RNG and the GG. Conceptually this method helps explain our cross street sampling strategy employed in this paper. Like the GG disc or circle of influence, each Delaunay triangle is defined by an empty circumsphere (Okabe et al, 1992), an empty circle whose circumference intersects all three points of the Delaunay triangle (Figure 2).

If we generalize the process of using discs or empty circles in a plane, these discs having neighborly properties similar but not equal to those that construct the DT, we can construct a spectrum of endoskeletons, the family of \( \beta \)-skeletons (Radke, 1983; Kirkpatrick and Radke, 1985; Radke, 1988). The neighborly properties which generate these endoskeletons, produce a spectrum of skeletons which include the completely connected graph at one extreme, passing through subgraphs like the RNG and GG, to eventually a graph which equals the point set itself. Since our objective is to construct a better descriptor of street centerline which integrates the essential points in a street center sample point pattern, it is likely that the acceptable skeleton structure will emanate from a fairly narrow range within the overall spectrum.

\( \beta \)-skeletons provide a hierarchy of descriptors of internal shape based on measures of neighborliness. We use both a lune and disc (or circle based method) to search a neighborhood for intervening opportunities in the form of other points from a point set \( V_s \) under-study.

Circle Based Neighborhoods: For a given pair of points \( V_i \) and \( V_j \) we can construct a continuous family of neighborhoods \( N(V_i, V_j, \beta) \) based on a pair of circles which pass through both \( V_i \) and \( V_j \), and are indexed

![Figure 1. NNG\( \subseteq \text{MST} \subseteq \text{RNG} \subseteq \text{GG} \subseteq \text{DT} \)](image)

![Figure 2. The circumsphere method for generating the DT](image)
by a single real value parameter $\beta$, where $\beta \in [0, \infty]$. When $\beta \leq 1$, the neighborhood searched between $V_i$ and $V_j$ is composed of the union of two discs, with radius $\beta(V_i, V_j) / 2$, which pass through both $V_i$ and $V_j$. When $\beta \in [0,1]$ the neighborhood searched for $V_i$ and $V_j$ is the intersection of two discs of radius $d(V_i, V_j) / (2\beta)$ which pass through $V_i$ and $V_j$. This $\beta$-skeleton algorithm generates the Gabriel graph when $\beta = 1$ and the neighborhood searched is $N(V_i, V_j, 1)$ (Figure 3).

**Lune Based Neighborhoods:** In the lune based approach we can also construct a continuous family of neighborhoods $N(V_i, V_j, \beta)$, indexed by a parameter $\beta$, where $\beta \in [0, \infty]$. However, when $\beta \geq 1$, the neighborhood searched is the intersection of the two circles of radius $\beta d(V_i, V_j) / 2$, centered at the points $(1 - \beta/2) V_i + (\beta / 2) V_j$ and $(\beta / 2) V_i + (1 - \beta / 2) V_j$, respectively. When $\beta \in [0,1]$ the neighborhood searched is the intersection of two discs of radius $d(V_i, V_j) / (2\beta)$ which pass through $V_i$ and $V_j$. Like the circle based approach, this $\beta$-skeleton algorithm generates the GG when $\beta = 1$ but also generates the RNG when $\beta = 2$.

No matter what $\beta$ method is chosen, as $\beta \to \infty$, the $\beta$-skeleton generated, except for degenerate point sets, is devoid of edges and as $\beta \to 0$, the $\beta$-skeleton generated becomes the completely connected graph (CCG) where edges occur between all pairs of points.

**IV. DECOMPOSITION METHOD APPLIED**

The City of Berkeley is located within the fully urbanized Eastern Shore of the San Francisco Bay in Northern California. With a population of 100,000 the City extends from the Bay east to the Berkeley Hills and is bounded by the City of Oakland to its south and the cities of Albany and Richmond to the north (Figure 4).

![Image](Figure 4. City of Berkeley in San Francisco Bay Region.)

In an effort to develop a functional parcel map which would become the base for a SDSS for the City, we translated a series of Computer Aided Design (CAD) drawings from a local utility company, East Bay Municipal Utility District (EBMUD), into a GIS data structure (Arc/Info). These CAD drawings contained unique water tap numbers within each parcel which are used for billing purposes and eventually lead to a process where we were able to link water tap number and parcel address. From the County Assessment data base we were able to further link parcel address with parcel number (APN) and produce a digital parcel map with both street address and APN, a critical task for the majority of a city's information management needs. When using a variety of ancillary data from different sources, coding errors always exist. After applying some standard quality control measures which included editing the data where necessary, we were able to assess data integrity of the parcel base map to be 97% accurate.

Although the parcels had originally been scanned from...
paper cadastral maps, the CAD generated parcel maps were not georeferenced. When the parcel maps were projected into a State Plane Coordinate System using the North American Datum (NAD) 83 and overlaid on 6" panchromatic Aerotopia Digital Orthophotography, we found parcel boundaries transecting both structures and street segments with an average displacement error exceeding 10 feet. To reduce this error we applied rubber sheeting algorithms and adjusted the parcel map to physical, accurately geo-positioned monuments, surveyed by the Public Works Engineering Department. Thirty strategically selected monuments, with accuracy within .01 inch, were used as control points and the parcel map georeferenced.

Our process decomposes city blocks or street casings to generate street centerlines. The generation of city blocks in this instance is derived by dissolving the parcel database based on the first seven figures of the APN which are block-unique. This process can return fictitious sliver parcels which will result in a contaminated street database with erroneous street segments. The error is eliminated by assigning sliver or multiple polygons within a block an identical attribute value from which we dissolve and produce a clean city block database (Figure 5).

The center of a street is of course the mid point between two opposing blocks. If the blocks are uniform, measuring perpendicular to the block face might suffice, but where the block faces are curved, a more rigorous sampling is needed. In the City of Berkeley the street curvature varies from straight lines in the flat lands to the southwest, to extreme curved lines in the hill areas to the northeast (Figure 6).

To measure the mid point between two opposing blocks we sample along both block faces and compare each sample point to its two closest sample points on the other block face. From this we can easily produce a set of essential or mid points which make up a subset of the street centerline. We conservatively sample every 9 meters along the block face in order to insure we capture the complexity of the curves in the hill area to the northeast (Figure 7).

We construct the Delaunay Triangulation (DT) of the set of points which serves to connect each sample point to its nearest two points on its opposing block face. Of course the DT constructs this connection for the block face across the "street space" as well as the opposing face which constructs the same block (Figure 8).

Since our interest is in constructing the mid point across the "street space", we buffer each city block polygon by a very small constant value (in this instance 0.3 meters) and eliminate all line segments that lie within. The resultant data base contains individual line segments (Figure 9) whose mid points lie on and can be used to generate the street centerlines for the entire city (Figure 10).
Figure 7. Sample every 9 meters along the block face.

It is important to note that in the case of cul-de-sacs the same block face generates points that become Delaunay neighbors which results in essential or mid points that will eventually result in a fork in the centerline termination point. The few cases where this occurred were eliminated before further processing occurred.

What visually appears to be a simple task, to decompose the midpoint database and create street centerlines, is a complex process to automate. If only the essential points that sample along the centerline for a given street had to be decomposed, the minimum spanning tree (MST) would suffice, however complexity is introduced by the varied and often complicated ways that streets intersect each other. The addition of X, Y and T intersections (Figure 11) call for a decomposition algorithm which can be tuned to integrate the essential points in a street center sample point pattern in order to accurately construct street centerlines. We employ both the Lune and disc based neighborhood methods of the $\beta$-skeletons and apply a fairly narrow range within the overall spectrum.

Figure 8. DT created from the densified block boundary coverage.

V. RESULTS

Table 1 contains the results of a number of $\beta$ values for both the lune and disc based algorithms.

When $\beta = 1$, the Gabriel Graph (GG) is generated by both algorithms and a fully connected street network results which easily solves four-way intersections but over connects at T and Y intersections forcing considerable post-processing to create a satisfactory street network.

Although the disc based neighborhood appears to be the best and is very effective connecting straight and

Figure 9. The isolated DT line segments crossing a street.

Figure 10. Resulting street midpoints for a small section of the city.
Figure 11. X, T and Y intersections defined.

Figure 12. β= 1, the Gabriel Graph (GG) illustrated for a small section of the city.

Figure 13. β> 2 produces some failed intersections.

curved line segments, it too produces some failed intersections when β> 2 (Figure 13).

After an iterative process we found the circle or disc based approach where β= 1.2 to produce the most interesting result (Figure 14).

This iteration solves all four-way or X-intersections, Y-intersections and it creates complete arc segments where there are no intersections. The only remaining problem occurs at T-intersections where the algorithm does not properly connect the street center lines.

Table 1. Contains the results of a number of β values for both the lune and disc based algorithms

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>Circle</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>T, X &amp; Y overconnect.</td>
</tr>
<tr>
<td>1.2</td>
<td>Circle</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Straight segments not complete</td>
</tr>
<tr>
<td>2.0</td>
<td>Circle</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>Straight segments not complete</td>
</tr>
<tr>
<td>3.0</td>
<td>Circle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>Circle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>Circle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.0</td>
<td>Circle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>Lune</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>T, X &amp; Y over-connect.</td>
</tr>
<tr>
<td>1.2</td>
<td>Lune</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>T over-connect.</td>
</tr>
<tr>
<td>2.0</td>
<td>Lune</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>Straight segments not complete</td>
</tr>
<tr>
<td>3.0</td>
<td>Lune</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>Straight segments not complete</td>
</tr>
<tr>
<td>4.0</td>
<td>Lune</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Straight segments not complete</td>
</tr>
<tr>
<td>6.0</td>
<td>Lune</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>Straight segments not complete</td>
</tr>
<tr>
<td>20.0</td>
<td>Lune</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This anomaly occurs as the original block face sampling to produce the DT and eventually the street centerline point sample set, lacks four block corners at T-intersections and thus can only produce connections that deviate considerably from a right angle.

In post-processing it is possible to automatically extend an arc until it intersects another and corrects the T-intersections. These commands are common editing tools in most GIS where a distance can be set to specify how far the algorithm will search for an intersecting arc. This post-processing step could result in a complete data set but more often a few anomalous errors remain which need attention to ensure that all complicated cases, such as circles and complex intersections are represented correctly. Once the street centerline data set is anatomicall correct, we can generalize or omit the redundant vertices in the street centerline data set using weeding algorithms common to most GIS software. Figure 15 illustrates the final street centerline data set processed with a weed tolerance of 2 feet.

VI. CONCLUSIONS

The central objective of this study, to develop a new method that extracts the form of a set of points embedded in a plane and automatically defines and generates street centerlines from parcel-block information, was successfully accomplished. Based on notions of neighborliness, a spectrum of potential centerline solutions are generated which automate the process and provide a better centerline fit, especially where curves exist. This parameterized notion of neighborly provides a flexible and powerful method of tuning the centerline construction to better automate and describe the centerline between parcel-blocks and correct intersection problems common to current automated centerline generators. This method is useful in characterizing centerlines where massive parcel-block data sets prevail and automated systems are their only match.

VII. LIMITATIONS

Although we generalize the process of generating street centerlines, the application is still dependent on the accuracy and completeness of the parcel-block data base. Since we use tax parcel maps as a base data set, right-of-ways are not included. The shape of the block polygons are not always a good representation of how the street is actually aligned. In addition, wide streets with dividers or traffic islands are represented as a single line unless these islands are added in the block coverage.

ACKNOWLEDGMENTS

This research was partially supported by a contract from the City of Berkeley, and a grant from the Beatrice Farrand Fund of the University of California, Landscape Architecture Environmental Planning Department.
REFERENCES


