# Road Network Reconstruction for Organizing Paths

Daniel Chen* Leonidas J. Guibas* John Hershberger† Jian Sun*

**Abstract**

*We consider the problem of reconstructing a road network from a collection of path traces and provide guarantees on the accuracy of the reconstruction under reasonable assumptions. Our algorithm can be used to process a collection of polygonal paths in the plane so that shared structures (subpaths) among the paths in the collection can be discovered and the collection can be organized to allow efficient path similarity queries against new query paths on the same road network. This is a timely problem, as GPS and other location traces of both people and vehicles are becoming available on a large scale and there is a real need to create appropriate data structures and data bases for such data.*

## 1 Introduction

Finding shared structure in a collection of objects is a fundamental problem in both pattern matching [25] and similarity search [15, 19]. For the case of discrete strings, there are many classical solutions to this problem [17, 22, 24, 25]. However, there has not been much previous work on finding shared structure in path data embedded in Euclidean space. With path data such as GPS or other location traces becoming available on a large scale, it is important to develop algorithms and data structures to organize such geometric data. To facilitate this goal, we develop an algorithm to extract important shared structure in planar paths.

Many real world collections of path data seem to follow an underlying "road network." This assumption holds not only for the obvious example of vehicular GPS traces, but also for natural examples such as animal migration patterns [16] for which an explicit representation of an underlying road network may not be available. Our method of finding shared structure begins by considering what properties make a road network easy to reconstruct from sample data. Then, given reasonable assumptions about the underlying road network, our reconstruction algorithm can extract important shared structure in the paths. We further show how to use this information to implement a simple data structure for partial matching of paths.

Road network reconstruction is closely related to curve reconstruction, which is the problem of computing a polygonal curve that approximates a continuous curve sampled by a given point set (when the ordering of the points along the curve is unknown). Many algorithms have been proposed for reconstructing curves from a point set [2, 9, 10]. The crust algorithm presented by Amenta et al. [2] uses Voronoi-based techniques to reconstruct smooth and closed curves from densely sampled points. Dey et al. [9] followed up with the conservative-crust algorithm, which reconstructs smooth curves with endpoints. Again based on the Voronoi diagram, Dey and Wenger [10] proposed a method to reconstruct curves with sharp corners. Even though path traces typically come with temporal ordering for the samples, road network reconstruction is in fact a more difficult problem, since road networks contain many challenging non-manifold structures such as crossings. Chazal et al. [7] proposed a sampling theory to recover the topology of any compact set in Euclidean space for a set of sample points based on the distance function. But since a collection of sample paths intrinsically contains more information than a collection of sample points, we would also like to recover the geometry of the underlying road network. However, the road network reconstruction problem may have ambiguities, as we will see in Section 2, especially at crossings or when two roads come close to each other. Our strategy is to recover the geometry of the sections of the road network that are unambiguous and the connectivities between these sections.

There have been several interesting techniques developed in the computer vision community for extracting road networks from satellite and aerial imagery. Baumgartner et al. [3] and Zhao et al. [27] propose semiautomatic road extraction methods requiring manual intervention. These systems process digital images and mark the pixels corresponding to the roads. Mena et al. [18] propose an integrated method to automatically extract a skeleton of the road network from images. There has also been work [5, 12] using artificial intelligence techniques to recover a road network from massive amounts of GPS traces. However, these meth-

*Computer Science Department, Stanford University, Palo Alto, CA 94305.

†Mentor Graphics Corp., 8005 SW Boeckman Road, Wilsonville, OR 97070.

ods do not attempt to establish prior-free guarantees of accuracy. Our work is the first, to our knowledge, to approach this problem as a reconstruction problem and to prove guarantees without stringent assumptions on the distribution of input paths.

## 2 Road Network Model

In this paper, the underlying road network is modeled as a geometric graph $G = (V, E)$ embedded in $\mathbb{R}^2$ with edges drawn as polygonal curves. For clarity, we will refer to these embedded edges as *road fragments*. Furthermore, the road network is also associated with a road width $\epsilon$. This representation is both useful and natural, as it contains information about both the original road surface, which is the $\epsilon/2$-fattening of the road network $G$, and possible sample paths, which should be close to embedded paths on the graph under a natural distance measure. Formally, the *$\delta$-fattening* of a point set $A$ is the Minkowski sum $A + B(0, \delta)$, denoted $A^\delta$, where $B(x, r)$ is the ball of radius $r$ centered at $x$.

For convenience, we use vertices and their associated embedded points interchangeably. Moreover, we associate a path $P = (v_1, v_2, \ldots, v_k)$ on an embedded graph with a curve $C : [1, k] \to \mathbb{R}^k$ where $C(i) = v_i$ and $C$ restricted to the domain $[i, i+1]$ for $i = 1, \ldots, k-1$ is the arc length parameterization of the embedded edge $(v_i, v_{i+1})$. This allows us to use curves and paths interchangeably. A *subcurve* of a curve $\ell : [a, b] \to \mathbb{R}^2$ is the restriction of the map $\ell$ to a sub-interval of $[a, b]$. The arc length of a curve $\ell$ is denoted $\mathrm{Len}(\ell)$.

We use the Fréchet distance $d_F$ to measure the distance between two curves; see [1] for its formal definition. An intuitive definition of the Fréchet distance between two curves is the length of the shortest leash that allows a man to walk on one curve and his dog to walk on the other curve from the beginning of the curves to the end, without either one backtracking. The Fréchet distance is more natural than the Hausdorff distance because it takes into account the course of the curves. We restrict the allowable sample paths to be polygonal curves $P$ such that $d_F(P, Q) \leq \epsilon/2$, where $Q$ is a subcurve of an embedded path in $G$.

It is common in the literature to model a road network as a geometric graph [4, 8, 11, 23, 26]. However, we observe that given a real road network system, finding a geometric graph representation may require many arbitrary choices. For example, consider the situation in the top pictures in Figure 1, in which the yellow road and the orange road merge. Both underlying geometric graphs on the left and right sides seem to be reasonable representations of the road network, as their fattenings both seem to approximate the road surfaces well and a car moving through the intersection might be able to follow the path on either graph closely. Therefore, given sample traces of car movement on such a road network, it would be impossible to determine which geometric graph is the underlying truth unless further, possibly unreasonable assumptions are made about the distribution of the deviation of sample paths from the underlying graph skeleton. Moreover, such ambiguities extend beyond the geometric embedding of the graph. The bottom pictures in Figure 1 show that even the graph and its connectivity can be ambiguous. Both the black and blue graphs seem to be equally suitable representations of the illustrated road surface.

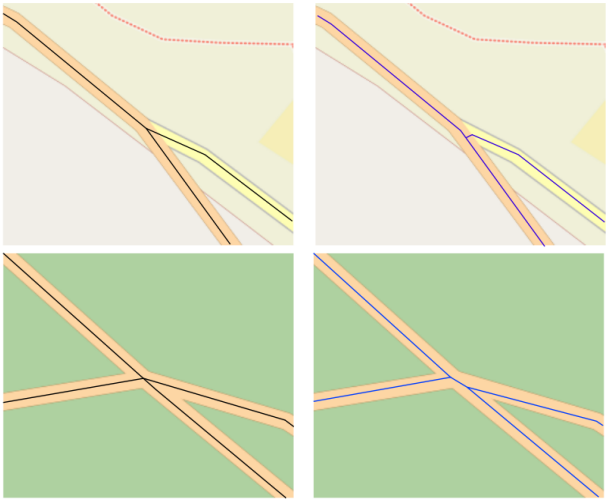


Figure 1: Ambiguities.

In spite of these ambiguities, it seems that under certain conditions, it should be possible to recover well-separated road fragments from sample paths. Therefore, given an underlying geometric graph representation of a road network, we seek to identify sections of the graph for which reconstruction is a plausible goal, and also give an algorithm that reconstructs these sections of the graph with bounded error under reasonable assumptions. To do this, we introduce the following notation:

Definition 2.1. ($\alpha$-good Points) *A point $p$ on $G$ is* $\alpha$-good *if $B(p, \alpha\epsilon) \cap G$ is a 1-ball that intersects the boundary of $B(p, \alpha\epsilon)$ at two points. Let a point $p$ be $\alpha$-bad if it is not $\alpha$-good.*

Intuitively, a good point on the road network graph corresponds to a point whose neighborhood in the graph is a single curve. We will argue that we are able to find such sections of the road network with bounded error. Moreover, portions of the vehicle routes traveling through these sections of road tend to be the most important. For example, if the region under

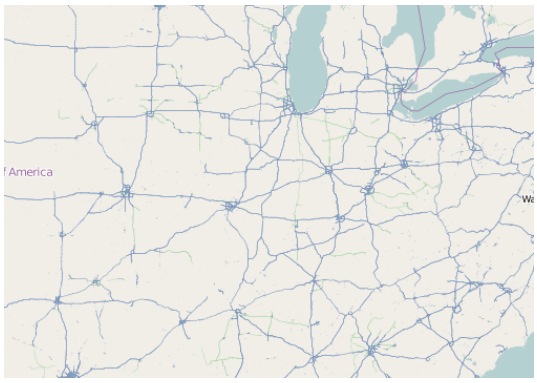


Figure 2: Major roads between cities.

consideration is a large portion of a nation, as in Figure 2, then we are most likely to be interested in paths that connect multiple cities. In this case, the course of a path through city streets is probably less important than the highway routes traversed.

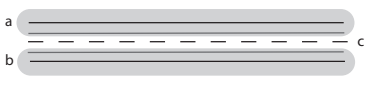
## 3 Assumptions

In this section, we state the assumptions on both the road network and the input sample paths that are necessary to guarantee the performance of our algorithm. However, as Section 7 shows, the algorithm produces reasonable results even when these assumptions are not fully met. We will also explain why the assumptions model real road networks well. The following assumptions on the road network use a parameter $b$, which is set to $2+\sqrt{6}$ to guarantee proper behavior of the algorithm:

1. Each road fragment $\gamma$ is a polygonal curve $(p_1, p_2, \ldots, p_n)$ such that $||p_i p_{i+1}|| \geq 3b\epsilon$, $\angle p_{i-1} p_i p_{i+1} \geq \pi/2$, and for any point $p \in \gamma$, $B(p, (b+2)\epsilon) \cap \gamma$ is a 1-disk.

2. For each road fragment $\gamma$, there exists a subcurve $\beta$ of $\gamma$ such that $\text{Len}(\beta) \geq 9b$ and any $p \in \beta$ is $(b+2)$-good.

3. If the points $p_1, p_2$ are from road fragments $\gamma_1, \gamma_2$ respectively and $d(p_1, p_2) < 3\epsilon$, then $\gamma_1$ and $\gamma_2$ share a node $v$, and the Fréchet distance between the two subcurves of $\gamma_1$ and $\gamma_2$ joining $v$ to $p_1$ and $p_2$ is less than $3\epsilon$.




Assumption (1) states that road fragments do not take sharp turns and do not come close to self-intersecting. It is less restrictive than the smoothness assumption used in many curve reconstruction algorithms [2, 9], as it allows for corners. Furthermore, if a section of road requires sharp turns, it can be split into multiple road fragments. Assumption (2) states that each road fragment has a good section that is long enough that sample paths can be used to justify its existence. Assumption (3) is needed to rule out pairs of nonintersecting road fragments that are so close together that their sample paths are indistinguishable from those belonging to a single road fragment. For example, in the figure above, road fragments $a$ and $b$ are both sampled by the gray paths. However, these sample paths could also arise from a single road fragment $c$. Assumption (3) ensures that two road fragments that do not meet at an intersection do not get close enough together to be mistaken for a single road fragment. Note that we do not make any assumptions on how long two road fragments can be close to each other.

For the input sample paths, we make the following two assumptions:

4. Each input path is within Fréchet distance $\epsilon/2$ of a subcurve of a path on $G$.

5. For each road fragment $\gamma$, there exists an input path $\ell$ such that $\ell^{\epsilon/2}$ contains $\gamma$.

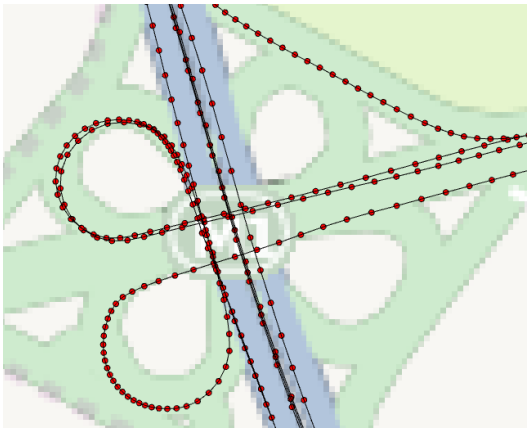


Figure 3: GPS trace data from [20].

Assumption (4) simply says that a path should stay within the road surface. Assumption (5) ensures that each part of the road network graph is well sampled. In fact, Assumption (5) is stronger than necessary. In particular, for the $(b+2)$-good sections of the road fragments, it is enough to assume that for each point, there is an input path passing through the ball of radius $(b+2)\epsilon$ centered that point. However, for simplicity of exposition, we state Assumption (5) in its stricter form. Figure 3 shows that the assumptions are met in GPS trace data from Open Street Map [20]. In fact, real GPS trace data appear to be very densely sampled.

## 4 Algorithm

Our reconstruction algorithm recovers the $(b+2)$-good sections of road fragments and their connectivities. The algorithm starts by obtaining a set of sample points forming a metric $b\epsilon$-net of the input paths. (A metric $b\epsilon$-net is a set of points $\mathcal{S}$ selected from the input paths such that all points on the input paths are within distance $b\epsilon$ of at least one point in $\mathcal{S}$ and no two points in $\mathcal{S}$ are within distance $b\epsilon$ of one another.) The algorithm then computes a Voronoi diagram using these sample points as sites. From a subgraph of the dual Delaunay graph, the algorithm builds a structure graph useful for organizing the input paths. The structure graph is the basis for building a reconstruction graph that represents the underlying road network graph with bounded error. In some sense, the algorithm recovers the road network by organizing the input paths. In the remainder of the section, we describe the structure graph and the reconstruction graph, and then outline the reconstruction algorithm that builds them.

### 4.1 Structure graph and reconstruction graph

To describe the structure and reconstruction graphs, we begin by introducing several concepts: Given the Voronoi diagram of the sample points, we extract a subgraph of the dual Delaunay triangulation (called the *restricted Delaunay triangulation*) by selecting Delaunay edges whose corresponding Voronoi edges intersect one or more input paths. We define the *degree* of a Voronoi cell to be equal to the degree of the corresponding node in the restricted Delaunay triangulation. A curve *cuts through* a bounded region if it intersects the region but both of its endpoints are outside the region.

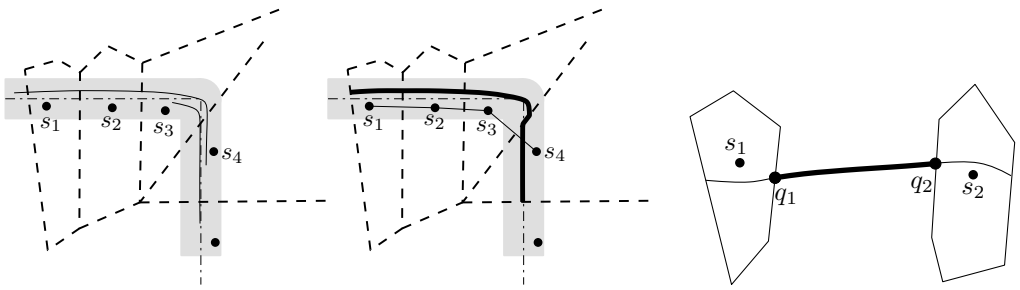
Definition 4.1. (Clean Voronoi Cell) *A Voronoi cell $\mathcal{V}(s)$ is* clean *if it is of degree two, there is at least one input path cutting through the cell, and any path cutting through the cell intersects two distinct Voronoi edges if and only if it also intersects the ball $B(s,\epsilon)$.*

We define the *clean graph* to be the subgraph of the restricted Delaunay triangulation induced by the sites of clean Voronoi cells. Every connected component in the clean graph is either a simple path or a simple cycle as each clean Voronoi cell is of degree two. Define a *primitive chain* to be the Voronoi cells in a connected component equipped with an Euler path $(s_1,\ldots,s_m)$. For a connected component equipped with $(s_1,\ldots,s_m)$, $(\mathcal{V}(s_1),\ldots,\mathcal{V}(s_m))$ is the corresponding primitive chain. We also refer to the Euler path $(s_1,\ldots,s_m)$ as the *skeleton*. We note that it is possible that the skeleton $(s_1,\ldots,s_m)$ can be a single point (i.e., $m=1$) or a simple cycle (i.e., $s_1=s_m$). With each primitive chain $c=(\mathcal{V}(s_1),\ldots,\mathcal{V}(s_m))$, we associate a polygonal curve as follows: for each $s_i$, choose a path cutting through $\mathcal{V}(s_i)$, let $\ell_{s_i}$ be the subcurve in $\mathcal{V}(s_i)$, and then connect consecutive $\ell_{s_i}$ and $\ell_{s_{i+1}}$ using the segment joining their endpoints on the Voronoi edge common to $\mathcal{V}(s_i)$ and $\mathcal{V}(s_{i+1})$; see Figure 4(a). We use this polygonal curve, denoted $\ell_c=(\ell_{s_1},\ldots,\ell_{s_m})$, to approximate the section of the road fragment covered by the primitive chain. This is a more accurate reconstruction than using the polygonal curve $(s_1,s_2,\ldots,s_m)$, as we allow a turn as large as $\pi/2$ within a road fragment; see Figure 4(a).

Definition 4.2. (Link) *A* link *between two clean Voronoi cells (not necessarily distinct) is a subcurve of an input path that intersects both cells at two distinct Voronoi edges. Furthermore, the part of the subcurve not in these two cells does not intersect $B(s,\epsilon)$ for any sample point $s$ corresponding to a clean Voronoi cell and has endpoints on different Voronoi edges.*

As we can see, the part of the link not in these two cells connects them; we call this the *vital part* of the link. Let the endpoints of the link be those of its vital part; see Figure 4(b). We say that two links are of the same *type* if their endpoints are on the same Voronoi edges. We also say a link connects two primitive chains of it touches one cell from each primitive chain. Note that no link connects a primitive chain with a cyclic skeleton, and therefore cyclic primitive chains are isolated. A primitive chain with a simple path skeleton has two *boundary edges* where endpoints of links may reside. A link always connects two different boundary edges. We further define a *clean link* be a link whose boundary edges do not touch links of a different type.

The *structure graph* is obtained from the restricted Delaunay triangulation by collapsing primitive chains into supernodes; the vertices of the structure graph are primitive chains and the edges are the different types of links, with one edge for each type of link present in the restricted Delaunay triangulation. Consider the subgraph of the structure graph obtained by keeping only the clean links. As in the clean graph, each connected component in this subgraph is a simple path or cycle, and we can again equip each connected component with an Euler path. We call such an Euler path a *chain*, denoted $(c_1,c_2,\ldots,c_m)$, where $\{c_i\}_{i=1}^m$ are the component primitive chains. As in a primitive chain, when $c_1=c_m$, the chain $(c_1,c_2,\ldots,c_m)$ is a cycle and no link connects it to any other chain. This means the road fragments covered by such a chain form an isolated cycle, which is very unlikely to happen in the real world. Our algorithm assumes that no chain is a cycle: every chain must end at a boundary edge that either does not connect to any link or connects to at least two links.




(a) (b)

Figure 4: (a) $(\mathcal{V}(s_1), \mathcal{V}(s_2), \mathcal{V}(s_3), \mathcal{V}(s_4))$ is a primitive chain and the bold curve is its associated polygonal curve, which approximates the road fragment (dash-dotted line) better than the polygonal curve $(s_1 s_2 s_3 s_4)$. (b) Illustration of a link. The bold part is the vital part of the link, and $q_1$ and $q_2$ are the endpoints of the link.

We obtain the *reconstruction graph* from the structure graph in the following manner: The vertices of the reconstruction graph correspond to boundary edges at the ends of chains. A chain $(c_1, c_2, \ldots, c_m)$ forms an edge between two vertices. This edge is embedded as a polygonal curve formed by connecting the polygonal curves associated with the primitive chains to the vital parts of the interleaved links by adding segments between the chain/link vertices that lie on common boundary edges. The number of the edges in the reconstruction graph is the same as the number of chains in the structure graph.

**4.2 Outline** Our reconstruction algorithm consists of the following seven steps:

1. Find a $b\epsilon$-net ($b\epsilon$-cover and $(b\epsilon/2)$-packing) $\mathcal{S}$ of the input paths where $b = 2+\sqrt{6}$ and $\epsilon$ is an input parameter denoting the road width.
2. Compute the Voronoi diagram $\mathcal{V}(\mathcal{S})$.
3. Mark clean Voronoi cells, and compute the clean graph and primitive chains.
4. Compute a representative polygonal curve for each type of link and build the structure graph.
5. Mark clean links and find the chains.
6. For each chain, check if it has a clean Voronoi cell $\mathcal{V}(s)$ such that every path cutting through $B(s, 5\epsilon)$ must intersect $B(s, \epsilon)$. If not, mark all the Voronoi cells in the chain not clean.
7. Update the structure graph, and then compute the reconstruction graph.

Sections 5 and 6 establish the performance guarantees of the algorithm. Theorem 5.7 shows that for each subcurve of the $(b+2)$-good section of a sufficiently long road fragment, there is an edge (a polygonal curve) in the reconstruction graph whose $\epsilon/2$-fattening covers most of the subcurve. However, short road fragments may create many chains in step (5) of the algorithm. The sample points in these extra chains must sample the 3-bad sections of the road fragments. Since we do not seek guarantees on the recovery of 3-bad sections, it is fine to replace these chains with links, and that is what step (6) does. However, the algorithm cannot remove all extra chains: there are cases in which the algorithm cannot tell whether input paths sample one road fragment or multiple road fragments. Nevertheless, Theorem 6.6 proves that the number of the chains, and thus the number of edges in the reconstruction graph output in step (7), is at most three times the number of road fragments.

The running time of the algorithm is dominated by the time to find the $b\epsilon$-net $\mathcal{S}$ of the infinite point set of input curves. For $n$ input paths of total complexity $M$, using an incremental Voronoi diagram and tracing the polygonal curves through the incremental diagram to find the next point to add to the $b\epsilon$-net, the time complexity is bounded by $O(k^2 + Mk \log k)$, where $k$ is the complexity of the $b\epsilon$-net. This is because in the worst case, the incremental Voronoi diagram construction takes $O(k^2)$ time [21], each segment of a polygonal curve might traverse $\Theta(k)$ different Voronoi cells, and ray shooting takes $O(\log k)$ time [14]. Of course, in practice, one can use an approximate solution by finding a dense sample of the input paths, and then using Har-Peled's algorithm [13] to find a $b\epsilon$-net of the dense sample in linear time. The rest of the algorithm can be implemented by tracing the polygonal curves through the Voronoi diagram. In our implementation, we use a version of this approach, except that we use a simpler algorithm to find the $b\epsilon$-net. Section 7 describes our implementation and the results of our experiments with it.

## 5 Recovering $(b+2)$-Good Sections

In this section, we show that most of the Voronoi cells whose sites sample the $(b+2)$-good sections of a road fragment are clean and those that sample a connected component of the $(b+2)$-good section form a primitive chain.

A Voronoi cell $\mathcal{V}(s)$ *properly intersects* a curve $\gamma$ if there is a subcurve of $\gamma$ such that its $\epsilon/2$-fattening is contained in $\mathcal{V}(s)$. A primitive chain $c = (s_1, \ldots, s_m)$ *properly intersects* a curve $\gamma$ if one of its composing Voronoi cells properly intersects $\gamma$, or equivalently either $\mathcal{V}(s_1)$ or $\mathcal{V}(s_m)$ properly intersects $\gamma$. We say a point *samples* a curve if the point is in the $\epsilon/2$-fattening of the curve. The following lemma bounding the size of a Voronoi cell in $\mathcal{V}(\mathcal{S})$ is obvious but very useful.

LEMMA 5.1. *Let $s$ be any sample point in $\mathcal{S}$. Then we have*

*(i) $B(s, b\epsilon/2)$ is contained in $\mathcal{V}(s)$.*

*(ii) For any point $x$ in $G^{\epsilon/2} \cap \mathcal{V}(s)$, $d(x,s) < (b+1)\epsilon$.*

Let $X = \{x_i\}_{i=1}^m$ be a set of points sampling any part of a road fragment $\gamma$. One can impose an order on the points in $X$ as follows. For each $x_i$, let $x_i'$ be a point on $\gamma$ within $\epsilon/2$ of $x_i$. The order of $x_i'$s inherited from $\gamma$ induces an order on the $x_i$s. Note that if any two balls of radius $\epsilon/2$ centered at any two different points in $X$ are disconnected, this order does not depend on the choice of $x_i'$s and therefore is well-defined up to reversal. For instance, if $S_\gamma \subset \mathcal{S}$ is the set of sample points in $\gamma^{\epsilon/2}$, then since $\mathcal{S}$ is a $(b\epsilon/2)$-packing, there is an induced ordering on the points in $S_\gamma$. By a slight abuse of notation, we also use $S_\gamma$ for the ordered sequence $S_\gamma = (s_1, s_2, \ldots, s_m)$.

To show that $(b+2)$-good sections are recovered well, we first prove the following lemmas:

LEMMA 5.2. *If a point $w$ samples the $(b+2)$-good section of a road fragment $\gamma$, then $w$ must be in the Voronoi cell of some sample point in $S_\gamma$.*

*Proof.* Let $s \in \mathcal{S}$ be a sample point with $w \in \mathcal{V}(s)$. (We are assuming Voronoi cells are closed, so $w$ might belong to more than one.) By Lemma 5.1, $d(s,w) < (b+1)\epsilon$. Let $w'$ be a $(b+2)$-good point in $\gamma \cap B(w, \epsilon/2)$. We have $d(w', s) < (b+1.5)\epsilon$ and hence $B(s, \epsilon/2)$ is contained in $B(w', (b+2)\epsilon)$. Thus $s$ is in $\gamma^{\epsilon/2}$ as $w'$ is $(b+2)$-good, which proves the lemma. □

LEMMA 5.3. *Suppose a point $w$ on the 1-skeleton of the Voronoi diagram $\mathcal{V}(\mathcal{S})$ samples the $(b+2)$-good section of a road fragment $\gamma$. Consider the set $S_\gamma \cup \{w\}$ with the order induced by $\gamma$. Then $w$ must be in the Voronoi cell of the samples of its two immediate neighbors in this order.*

*Proof.* By Lemma 5.2, $w$ must lie in the Voronoi cells of the samples in $S_\gamma$. As $d(w,s) > b\epsilon/2$ for any $s \in \mathcal{S}$, the road fragment indeed induces an order on the set $S_\gamma \cup \{w\}$. Let $s_{j-1}$ and $s_j$ be the two immediate neighbors of $w$ in the ordered set $S_\gamma \cup \{w\}$. Suppose that $\mathcal{V}(s_k)$ with $k > j$ contains $w$. Let $w'$ be a $(b+2)$-good point on $\gamma$ within distance $\epsilon/2$ of $w$, and let $s_i'$ be a point on $\gamma$ within distance $\epsilon/2$ of $s_i$ for any $i$. Since $d(w, s_k) < (b+1)\epsilon$ by Lemma 5.1, $s_k \in B$ for $B = B(w', (b+1.5)\epsilon)$. By Assumption (1) each line segment in $\gamma$ is longer than $3b\epsilon \geq (b+1.5)\epsilon$, and so there is at most one corner on $\gamma$ between $w'$ and any intersection of $\gamma$ and $B$; see Figure 5(a). Let the angle at this corner be $\theta$ (if there is no corner, $\theta = \pi$).

Consider the triangle $\Delta w s_j s_k$ and define $a = \frac{1}{\epsilon} d(w, s_j)$ and $c = \frac{1}{\epsilon} d(s_j, s_k)$. Let $\alpha$ be the angle $\angle w s_j s_k$. Then $d(w, s_k) \leq a\epsilon$, $b/2 \leq a \leq (b+1)$ and $c \geq b$. From the law of cosines, $a^2 \geq (d(w,s_k)/\epsilon)^2 = a^2 + c^2 - 2ac\cos\alpha$ and thus

$$c^2 - 2ac\cos\alpha \leq 0.$$

On the other hand, consider two lines $\ell_1, \ell_2$ parallel to and distance $\epsilon/2$ from the line segment containing $w'$; see Figure 5(a). The placement of $s_j$ and $s_k$ falls into one of the following three cases: (1) both $s_j$ and $s_k$ are sandwiched between $\ell_1$ and $\ell_2$; (2) neither $s_j$ nor $s_k$ lies between $\ell_1$ and $\ell_2$; (3) $s_j$ is between $\ell_1$ and $\ell_2$ and $s_k$ is not. One can verify that in each case, $\alpha > \theta - \alpha_1 - \alpha_2$ where $\sin\alpha_1 < \frac{\epsilon}{d(w,s_j)} = \frac{1}{a}$ and $\sin\alpha_2 < \frac{\epsilon}{d(s_j,s_k)} = \frac{1}{c}$. Since $\theta \geq \pi/2$, $\cos\alpha < \cos(\theta - \alpha_1 - \alpha_2) < \frac{1}{a} + \frac{1}{c}$. Thus for $b \geq 2 + \sqrt{6}$,

$$c^2 - 2ac\cos\alpha > c^2 - 2c - 2a \geq b^2 - 2b - 2(b+1) \geq 0,$$

a contradiction. This proves the lemma. □

COROLLARY 5.4. *For a road fragment $\gamma$, if $\beta = \{q \in \gamma \mid q \text{ is } (b+2)\text{-good}\}$, then there is no Voronoi vertex in $\beta^{\epsilon/2}$.*

LEMMA 5.5. *Let $\gamma$ be a road fragment, let $S_\gamma = (s_1, s_2, \ldots, s_m)$, and suppose a point $w \in \gamma^{\epsilon/2}$ lies on the boundary of $\mathcal{V}(s_k)$. If $s_k$ samples the $(b+2)$-good section of $\gamma$ and $1 < k < m$, then in the order induced by $\gamma$ on the set $S_\gamma \cup \{w\}$, the immediate neighbors of $w$ are either $s_{k-1}, s_k$ or $s_k, s_{k+1}$.*

*Proof.* Prove by contradiction. Without loss of generality, assume that $w$ lies after $s_{k+1}$ in the order induced

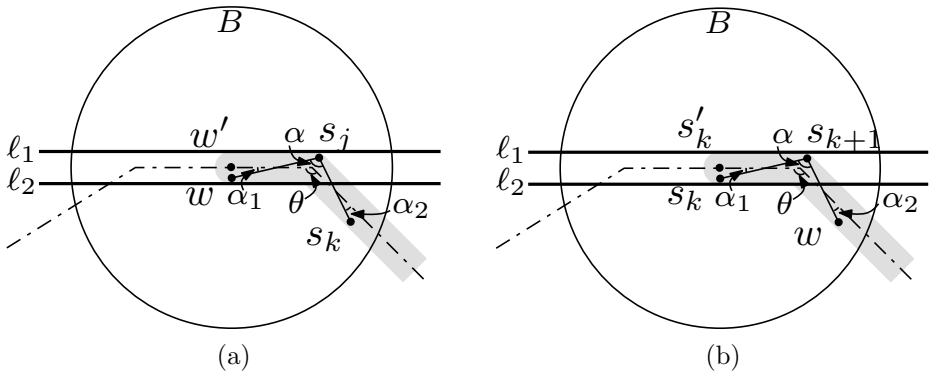



Figure 5: (a) Illustration for Lemma 5.3; (b) Illustration for Lemma 5.5.

by $\gamma$. As in the proof of Lemma 5.3, consider the triangle $\Delta s_k s_{k+1} w$ and define $a = \frac{1}{\epsilon} d(w, s_{k+1})$ and $c = \frac{1}{\epsilon} d(s_k, s_{k+1})$. Then $d(w, s_k) \leq a\epsilon$, $b/2 \leq a \leq (b+1)$ and $c \geq b$. From the law of cosines, $a^2 \geq a^2 + c^2 - 2ac\cos\alpha$ and thus $c^2 - 2ac\cos\alpha \leq 0$. On the other hand, the angle $\alpha = \angle s_k s_{k+1} w$ has the same lower bound as the angle $\angle w s_j s_k$ (see Figure 5(a)) in the proof of Lemma 5.3. Thus similarly we have $c^2 - 2ac\cos\alpha \geq c^2 - 2c - 2a > b^2 - 2b - 2(b+1) \geq 0$ for $b \geq 2 + \sqrt{6}$, which gives a contradiction. □

**Remark:** Although Lemmas 5.3 and 5.5 have similar statements and proofs, they differ in that $w$ samples the $(b+2)$-good section in Lemma 5.3, whereas $s_k$ does so in Lemma 5.5.

LEMMA 5.6. *For any subcurve $\beta$ of the $(b+2)$-good part of a road fragment $\gamma$, if $\mathrm{Len}(\beta) \geq 3b\epsilon$, there must be a sample point $s$ in $\beta^{\epsilon/2}$.*

*Proof.* Let $\beta = (p_1, p_2, p_3, \ldots, p_m)$. Note that by Assumption (1), $\mathrm{Len}(p_i p_{i+1}) \geq 3b\epsilon$ for any $2 \leq i \leq (m-2)$. If $d(p_1, p_2) \geq (b+1)\epsilon$, choose $p \in p_1 p_2$ with $d(p, p_1) = (b+1)\epsilon$. Otherwise, choose $p \in p_2 p_3$ with $d(p, p_3) = (b+1)\epsilon$. Such a $p$ must exist as $d(p_2, p_3) \geq (2b-1)\epsilon$ in this case. Then we know $B(p, (b+1)\epsilon) \cap G$ is a subcurve of $\beta$. From the sampling process, there must be a sample point $s$ in $B(p, (b+1/2)\epsilon)$ as $B(p, \epsilon/2)$ must intersect a sample path. There is a point on $G$ within distance $\epsilon/2$ of $s$. Then this point on $G$ is within distance $(b+1)\epsilon$ of $p$, and therefore is on $\beta$. Thus $s$ must be in $\beta^{\epsilon/2}$. □

We are now ready to prove the reconstruction guarantees on the $(b+2)$-good sections of a road fragment:

THEOREM 5.7. *For a subcurve $\beta$ of the $(b+2)$-good section of a road fragment $\gamma$, let $S_\beta = (s_1, s_2, \ldots, s_m)$ be the ordered set of $\beta$'s samples, and let $S \subseteq S_\beta$ be the subsequence of samples whose Voronoi cells are clean. Then*

- *$S_\beta \setminus S \subseteq \{s_1, s_m\}$, and thus there is a primitive chain $c$ whose skeleton contains $S$ as a sub-path,*
- *if $\mathrm{Len}(\beta) \geq 10b\epsilon$ the $\epsilon/2$ fattening of $\ell_c$ covers a subcurve of $\beta$ no shorter than $\mathrm{Len}(\beta) - 10b\epsilon$.*

*Proof.* Let $s'_i$ be a point on $\beta$ within distance $\epsilon/2$ of $s_i$. Notice that any $\mathcal{V}(s_i)$ is of degree at least two, as Assumption (5) states that the cell is cut through by a subpath in $\gamma^{\epsilon/2}$. Note that $\mathcal{V}(s_i)$ does not intersect the $\epsilon/2$-fattening of road fragments in $G$ other than $\gamma$. This is because for any $x \in G^{\epsilon/2} \cap \mathcal{V}(s_i)$, $d(s_i, x) < (b+1)\epsilon$. As $s'_i$ is $(b+2)$-good, $x$ must sample $\gamma$.

Let $w$ be a point in both $\gamma^{\epsilon/2}$ and the boundary of $\mathcal{V}(s_i)$. For any $2 \leq i \leq (m-1)$, from Lemma 5.5, $w$ is within distance $\epsilon/2$ of a point on the $(b+2)$-good subcurve of $\gamma$ between $s'_{i-1}$ and $s'_{i+1}$. By Lemma 5.3, $w$ must also be in the Voronoi cell of $s_{i-1}$ or $s_{i+1}$, implying that $\mathcal{V}(s_i)$ is of degree two. In addition, by Corollary 5.4, $w$ cannot be a Voronoi vertex, which means that the Voronoi edges dual to $s_{i-1}s_i$ and $s_i s_{i+1}$ cut through $\gamma^{\epsilon/2}$, and sandwich the ball $B(s_i, b\epsilon/2)$ between them. From Assumption (4) on the input paths, any path cutting through $\mathcal{V}(s_i)$ intersects these two Voronoi edges if and only if it also intersects $B(s_i, \epsilon)$. Therefore $s_i$ is clean for $2 \leq i \leq (m-1)$ and $S_\beta \setminus S \subset \{s_1, s_m\}$. It is then obvious that there is a primitive chain $c$ whose skeleton contains $S$ as a sub-path.

In fact, $\mathcal{V}(s_i)$ is also clean for any $i \in \{1, m\}$ if $B(s_i, (b+1.5)\epsilon) \cap \gamma$ is a subcurve of $\beta$, because in this case any point $x \in \mathcal{V}(s_i) \cap G^{\epsilon/2}$ samples $\beta$. This is due to the fact that $d(s_i, x) < (b+1)\epsilon$ from Lemma 5.1 and thus $B(x, \epsilon/2)$ is contained in $B(s_i, (b+1.5)\epsilon)$. Consider $w$ on the boundary of $\mathcal{V}(s_i)$ and in $G^{\epsilon/2}$. Based on

Lemma 5.3 and Corollary 5.4, an argument similar to that above shows that $\mathcal{V}(s_i)$ is clean.

If $\text{Len}(\beta) \geq 10b\epsilon$, then by Lemma 5.6, we know that $S_\beta$ contains at least three points. If both $\mathcal{V}(s_1)$ and $\mathcal{V}(s_m)$ are clean, we know that the subcurve of $\beta$ between $s'_1, s'_m$ is no shorter than $\text{Len}(\beta) - 6b\epsilon$. If $\mathcal{V}(s_1)$ is not clean, then $B(s_1, (b+1.5)\epsilon) \cap \gamma$ is not a subcurve of $\beta$, in which case one can show that the subcurve of $\beta$ between one of the endpoints of $\beta$ and $s'_1$ has length at most $2b$. By Lemma 5.6, the length of the subcurve between the endpoint and $s'_2$ is at most $5b$. A similar statement holds for $s'_{m-1}$. This proves the theorem. □

## 6 Bounding the number of edges of the reconstruction graph

In this section, we bound the number of edges (embedded as polygonal curves) in the reconstruction graph. Because the number of edges is the same as the number of chains in the structure graph, it suffices to bound the number of chains. Our strategy is to consider the 3-good sections and the 3-bad sections separately. By Assumption (3) on road fragments, the 3-good section of a road fragment is a (connected) curve and the 3-bad section of a road fragment consists of at most two connected curves attached to the nodes to which the road fragment is incident. In order to bound the number of chains, we first prove several useful lemmas:

LEMMA 6.1. *Let $s$ be a point sampling a road fragment $\gamma$. If $\mathcal{V}(s)$ is clean and does not properly intersect the 3-bad section of $\gamma$, then for any input path $\ell$ cutting through $\mathcal{V}(s)$ and intersecting two distinct Voronoi edges of $\mathcal{V}(s)$, its part in between these two Voronoi edges lies in $\gamma^{\epsilon/2}$.*

*Proof.* Note that there exists a point $s'$ on the 3-good section of $\gamma$ with $d(s, s') < \epsilon/2$. Furthermore, since $\mathcal{V}(s)$ is clean, there is a point $q \in \ell$ in between the two Voronoi edges that $\ell$ intersects satisfying $d(q, s) < \epsilon$, and therefore $d(s', q) < 1.5\epsilon$. Since $s'$ is 3-good, $q \in \gamma^{\epsilon/2}$. Since $\mathcal{V}(s)$ does not properly intersect the 3-bad section of $\gamma$, any point on $\ell \cap \mathcal{V}(s)$ is within distance $\epsilon$ of a 3-good section of $\gamma$, and hence samples a 2-good section of $\gamma$. This means that the two intersections of $\ell$ with the Voronoi edges of $\mathcal{V}(s)$ sample $\gamma$, which proves the lemma. □

Consider all clean cells whose sites sample a road fragment but do not properly intersect the 3-bad section of the road fragment. Among them, by Theorem 5.7, those whose sites sample the $(b+2)$-good section are part of a primitive chain. However, the remaining clean cells may form other primitive chains. Nevertheless, the following lemma says that these primitive chains form a single chain.

LEMMA 6.2. *For any road fragment $\gamma$, let $S = (s_1, \ldots, s_m)$ be the subsequence of $S_\gamma$ consisting of the sites $s$ such that $\mathcal{V}(s)$ is clean and does not properly intersect any 3-bad section of $\gamma$. Then there is one and only one chain that contains $\{\mathcal{V}(s_1), \ldots, \mathcal{V}(s_m)\}$.*

*Proof.* Since $B(s_i, b\epsilon/2)$ is contained in $\mathcal{V}(s_i)$, $\mathcal{V}(s_i)$ breaks $\gamma^{\epsilon/2}$ into two pieces: left (containing the sample points in $S_\gamma$ before $s_i$) and right (containing the sample points in $S_\gamma$ after $s_i$). We can classify the intersections of the paths in $\gamma^{\epsilon/2}$ with the boundary of $\mathcal{V}(s_i)$ into two classes; one class of intersections lying on the left, denoted $v_i^-$, and the other class lying to the right, denoted $v_i^+$, see Figure 6(a).

From Assumption (5) on the input paths, there exists a subcurve $\beta$ of an input path in $\gamma^{\epsilon/2}$ connecting a representative of $v_i^+$ to a representative in $v_{i+1}^-$, where both $B(s_i, \epsilon)$ and $B(s_{i+1}, \epsilon)$ intersect $\beta$. Because $\mathcal{V}(s_i)$ and $\mathcal{V}(s_{i+1})$ are clean, $\beta$ intersects each of them at two distinct Voronoi edges. (Regions $\mathcal{V}(s_i)$ and $\mathcal{V}(s_{i+1})$ are not guaranteed to abut, which is the chief difficulty this lemma must handle.) Note that the subcurve of $\beta$ between the representative in $v_i^+$ and the representative in $v_{i+1}^-$ is in the $\epsilon/2$-fattening of the 3-good section of $\gamma$. Thus it does not intersect any ball of radius $\epsilon$ centered at a sample point that does not sample $\gamma$. It follows that if $\mathcal{V}(s_i)$ and $\mathcal{V}(s_{i+1})$ are not in the same primitive chain, $\beta$ is a link connecting them.

Note that both the representatives in $v_i^+$ and the representatives in $v_i^-$ lie on one Voronoi edge for every $i$. Thus there is only one type of link possible between $\mathcal{V}(s_i)$ and $\mathcal{V}(s_{i+1})$, see Figure 6(b).

For $1 \leq i \leq m-1$, consider any link $L$ connecting $\mathcal{V}(s_i)$ with one of its endpoints in $v_i^+$. By Lemma 6.1, we know that $L \cap \mathcal{V}(s_i)$ is in $\gamma^{\epsilon/2}$. From Assumption (4) on sample paths, the vital part of the link $L$ must be in $\gamma^{\epsilon/2}$. Thus the link $L$ cannot connect $\mathcal{V}(s_i)$ to a clean cell other than $\mathcal{V}(s_{i+1})$. Similarly one can show that for $2 \leq i \leq m$ any link connecting $\mathcal{V}(s_i)$ with one of its endpoints in $v_i^-$ must connect to $\mathcal{V}(s_{i-1})$. Therefore $\{\mathcal{V}(s_1), \ldots, \mathcal{V}(s_m)\}$ is contained in one chain.

It remains to show that $S$ is not empty. As there exists a subcurve $\beta$ of the $(b+2)$-good section with length greater than $9b\epsilon$, we have from Lemma 5.6 that $S_\beta$ contains more than three points. Let $s \in S_\beta$ be any point that is not an endpoint of the sequence. From the proof of Theorem 5.7, we know $\mathcal{V}(s)$ is clean. Because $s$ samples the $(b+2)$-good section, it is also easy to show that any path cutting through $B(s, 5\epsilon)$ also intersects $B(s, \epsilon)$ (note that $b+2 > 6$). Thus $S$ is not empty. □

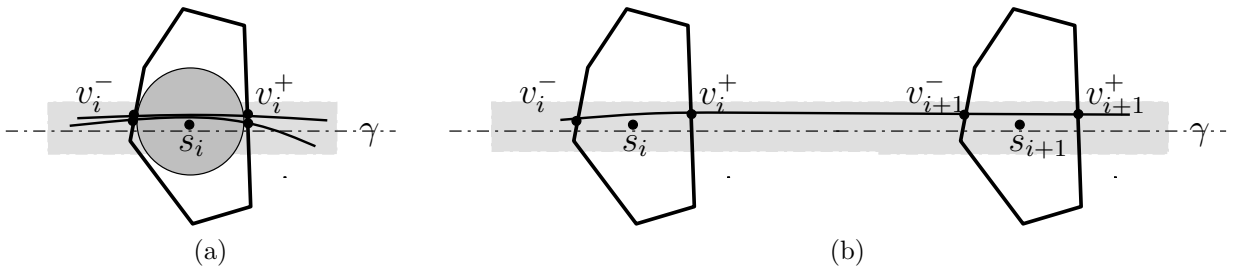



Figure 6: Illustration for Lemma 6.2.

LEMMA 6.3. *Suppose a path $\ell$ fully sampling a road fragment $\gamma$ cuts through a clean Voronoi cell $\mathcal{V}(s)$. Then any path fully sampling $\gamma$ will either cut through $\mathcal{V}(s)$ or stop inside it.*

*Proof.* Let $l \in \ell$ be a point such that $d(s,l) \leq \epsilon$. Let $p \in \gamma$ be such that $d(p,l) \leq \epsilon/2$. Then since the minimum radius of the Voronoi cell is $b\epsilon/2$, $B(p,\epsilon/2) \in \mathcal{V}(s)$. Since any path sampling $\gamma$ also intersects $B(p,\epsilon/2)$, the proof is done. □

The previous lemma allows us to define the chains a road fragment intersects to be the chains intersected by all paths fully sampling the road fragment. Furthermore, since samples are far enough apart, an ordering of a road fragment also induces an ordering on these chains.

LEMMA 6.4. *Let $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ be a connected 3-bad subcurve of a road fragment $r$. We order the vertices such that $\beta_1$ is a boundary point between the 3-bad and 3-good sections of $r$, and $\beta_k$ is an endpoint of $r$. Then this ordering of $\beta$ induces an ordering on the chains $c_1, c_2, \ldots, c_l$ it intersects. Furthermore, $\beta$ intersects $c_i$ on one side and exits it on the opposite side. Let $S(c)$ be the set of connected 3-bad subcurves intersected by a chain $c$. Then $S(c_i) \subseteq S(c_{i+1})$.*

*Proof.* Suppose there is a connected 3-bad subcurve $\gamma$ in $S(c_i)$ but not in $S(c_{i+1})$. Then $c_i$ contains a primitive chain that is intersected by all paths fully sampling $\beta$ and $\gamma$. Moreover, all Voronoi cells in primitive chains are clean. Therefore, every input path that crosses a Voronoi cell $\mathcal{V}(p) \in c_i$ must cross $B(p,\epsilon)$. By the sampling assumption (5), there is such a path in $\beta^{\epsilon/2}$ and another such path in $\gamma^{\epsilon/2}$. Therefore there is a point $b \in \beta$ and a point $c \in \gamma$ such that $d(b,c) \leq 3\epsilon$.

But since $\gamma \notin S(c_{i+1})$ and there is an input path fully sampling $\gamma$ within radius $\epsilon/2$, the distance between $\gamma$ and a Voronoi site $s$ in $c_{i+1}$ is at least $4.5\epsilon$ due to the last step of the algorithm. But as before there is a point $b \in \beta$ within distance $1.5\epsilon$ of $s$. Therefore, $b$ is at least $3\epsilon$ away from $\gamma$, which contradicts the road fragment assumption (3). □

LEMMA 6.5. *If $E$ is the set of road fragments and $\mathcal{C}$ is the set of chains intersecting more than one road fragment, then $|\mathcal{C}| \leq 2|E|$.*

*Proof.* Note that chains in $\mathcal{C}$ must intersect 3-bad subcurves of the road fragments. We assign an arbitrary ordering $\beta_1, \beta_2, \ldots, \beta_m$ to the maximal connected 3-bad subcurves. Because each road fragment may have at most two maximal connected 3-bad subcurves (Assumption (3)), we have $m \leq 2|E|$. We charge the chains of $\mathcal{C}$ to the maximal connected 3-bad subcurves in the following manner: Let a maximal connected 3-bad subcurve $\beta_j$ intersect chains $c_1, c_2, \ldots, c_l$ in the order of intersection given in Lemma 6.4. If there is some $k < j$ such that $\beta_k \in S(c_1)$, we charge $c_1$ to $\beta_j$. Otherwise, we charge the first $c_i$ such that $\beta_j$ is the lowest ordered element in $S(c_{i-1})$ and not the lowest ordered element in $S(c_i)$ to $\beta_j$. Each maximal 3-bad subcurve can be charged to at most once by Lemma 6.4. Furthermore, each maximal chain in $\mathcal{C}$ is charged to at least one maximal 3-bad subcurve. To see this, note that if a maximal chain $c_i$ is connected to $c_j$, then either $c_i$ or $c_j$ must have a second neighbor $c_k$ on the same side as the link $(c_i, c_j)$, or else the algorithm would have merged $c_i$ and $c_j$. Without loss of generality (WLOG) assume $c_i$ has neighbors $c_j$ and $c_k$. Let $\beta$ be the lowest ordered element in $S(c_i)$. Lemma 6.4 implies that $S(c_j)$ and $S(c_k)$ cannot both contain $\beta$; WLOG suppose $\beta \notin S(c_j)$. Then $c_i$ is charged to the lowest ordered element in $S(c_j)$. Therefore, the number of chains in $\mathcal{C}$ is at most $m \leq 2|E|$. □

Finally, we have:

THEOREM 6.6. *The number of edges of the reconstruction graph is at most three times the number of edges in the underlying road network.*

*Proof.* Lemma 6.4 implies that chains intersecting only one road fragment are merged with the chain reconstructing the good section of the road fragment. Combining this with Lemma 6.5, we can bound the number of chains by $3|E|$. □

## 7 Results

In this section, we show the results of our algorithm, implemented using CGAL [6]. For simplicity of implementation, we find an approximate solution by finding a $b\epsilon$-net of a sample of the polygonal curves instead of a $b\epsilon$-net of the polygonal curves themselves. Instead of ray shooting, we use a simple circular search at each Voronoi cell to trace the polygonal curves, as the Voronoi cells are likely to be small in practice. The performance of our implementation was reasonable. A data set of 5000 paths was processed in several minutes on a 2.33GHz Macbook Pro.

We ran our algorithm on two data sets. The first is a synthetic data set in which the input paths are generated by following the edges of an embedded graph. The paths deviate from the edges within an $\epsilon$-fattening of the road network. Figure 7 shows our results on this data set. The colored pieces are the associated polygonal curves of the primitive chains and the black are the links connecting the primitive chains. As we can see, the reconstruction graph recovers the good sections of the road fragment very well and the connectivities between these recovered sections are faithful to those of the original road network.

The second data set consists of 15182 GPS traces from Moscow downloaded from Open Street Map [20]. We overlay the result on a map of Moscow in the following examples, which use only 5000 of the traces. Figure 8(a) is an overview of the reconstructed graph. When zooming in as in Figure 8(b), we can see the recovered road fragments (colored pieces) and the links between them (black pieces). Even in the inner city (the top picture in Figure 8(b)), where the road network is complicated and many road fragments are close to each other, our algorithm recovers the road network quite well, as we can see many colored pieces. Using the lengths of edges as a performance measure, 75% of the graph belongs to the recovered road fragments.

## 8 Applications and Extensions

One obvious application for the structure graph is to organize paths according to the chains they follow. In such an application, one might build a weighted alphabet in which each character corresponds to a chain. Then paths would simply be strings in this alphabet and many algorithms and data structures for strings could then be applied. As an example, such strings may be inserted in a suffix tree [25] so that partial path queries may be performed in the time required to map the query path into the string. Our algorithm can also be modified for the case of a directed road network graph and for an incremental refinement of a road network graph. Details are omitted due to lack of space. One important and open problem is to detect road widths automatically for different parts of the road network. In practice, one might tune this parameter manually for different parts of the road network, but it would be interesting to consider methods to handle varying road widths automatically.

**Acknowledgements:** The authors acknowledge the support of National Science Foundation grant CCF-0634803, ARO grant W911NF-06-1-0275, and DARPA grant HR0011-05-1-0007.

## References

[1] Helmut Alt and Michael Godau. Computing the Fréchet distance between two polygonal curves. *Int. J. Comput. Geometry Appl.*, 5:75–91, 1995.

[2] Nina Amenta, Marshall Bern, and David Eppstein. The crust and the $\beta$-skeleton: Combinatorial curve reconstruction. *Graph. Models Image Process.*, 60(2):125–135, 1998.

[3] Albert Baumgartner, Stefan Hinz, and Christian Wiedemann. Efficient methods and interfaces for road tracking. In *International Archives of Photogrammetry and Remote Sensing*, page B:28, 2002.

[4] Sotiris Brakatsoulas, Dieter Pfoser, Randall Salas, and Carola Wenk. On map-matching vehicle tracking data. In *Proceedings of the 31st International Conference on Very Large Data Bases*, pages 853–864. VLDB Endowment, 2005.

[5] Rene Bruntrup, Stefan Edelkamp, Shahid Jabbar, and Bjorn Scholz. Incremental map generation with GPS traces. In *Proceedings IEEE Intelligent Transportation Systems*, pages 574–579, 2005.

[6] CGAL, Computational Geometry Algorithms Library. www.cgal.org, February 2008.

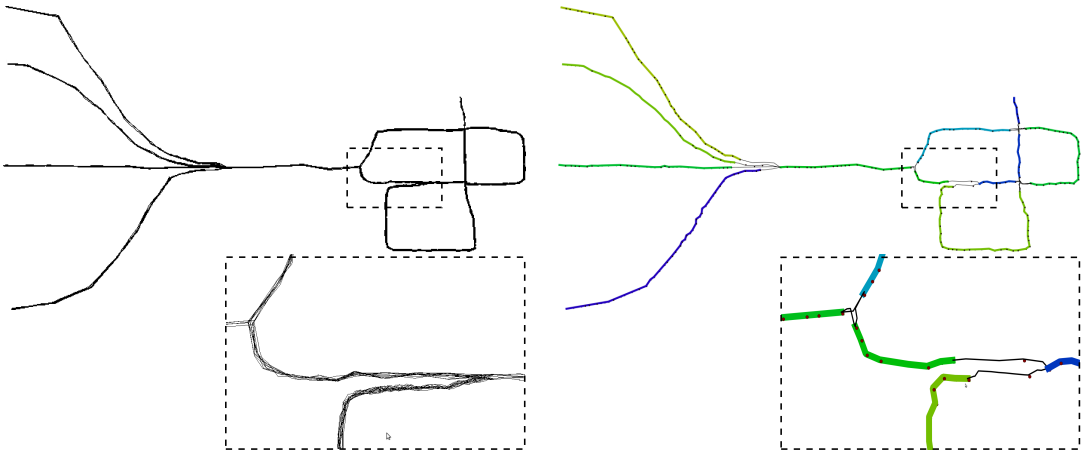
[7] Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in Euclidean space. In *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*, pages 319–326. ACM, 2006.

[8] Marc P. Deseilligny, Herve Le Men, and Georges Stamon. Map understanding for GIS data capture: Algorithms for road network graph reconstruction. In *Proceedings of the Second International Conference on Document Analysis and Recognition*, pages 676–679, 1993.

[9] Tamal K. Dey, Kurt Mehlhorn, and Edgar A. Ramos. Curve reconstruction: Connecting dots with good reason. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry*, pages 197–206. ACM, 1999.

[10] Tamal K. Dey and Rephael Wenger. Reconstructing curves with sharp corners. *Comput. Geom. Theory Appl.*, 19:89–99, July 2001.

[11] David Eppstein and Michael T. Goodrich. Studying (non-planar) road networks through an algorithmic



(a) (b)

Figure 7: (a) The input paths. (b) The reconstruction graph.


(a) (b)

Figure 8: (a) An overview of the reconstructed graph. (b) Reconstruction in two zoomed in regions.

lens. In *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–10. ACM, 2008.

[12] Tao Guo, Kazuaki Iwamura, and Masashi Koga. Towards high accuracy road maps generation from massive GPS traces data. In *Geoscience and Remote Sensing Symposium*, pages 667–670, 2007.

[13] Sariel Har-Peled. Clustering motion. *Discrete Comput. Geom.*, 31(4):545–565, 2004.

[14] John Hershberger and Subhash Suri. A pedestrian approach to ray shooting: Shoot a ray, take a walk. *J. Algorithms*, 18:403–431, 1995.

[15] Uri Keich, Ming Li, Bin Ma, and John Tromp. On spaced seeds for similarity search. *Discrete Appl. Math.*, 138(3):253–263, 2004.

[16] Kenneth J. Lohmann, Catherine M. F. Lohmann, and Nathan F. Putman. Magnetic maps in animals: Nature's GPS. *Journal of Experimental Biology*, 210(4):697–705, 2007.

[17] Edward M. McCreight. A space-economical suffix tree construction algorithm. *J. ACM*, 23(2):262–272, 1976.

[18] J. B. Mena and J. A. Malpica. An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. *Pattern Recogn. Lett.*, 26(9):1201–1220, 2005.

[19] S. Muthukrishnan and Süleyman Cenk Sahinalp. Approximate nearest neighbors and sequence comparison with block operations. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 416–424. ACM, 2000.

[20] OpenStreetMap. www.openstreetmap.org.

[21] Joseph O'Rourke. *Computational Geometry in C*. Cambridge University Press, 1998.

[22] T. F. Smith and M. S. Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195–197, March 1981.

[23] Carsten Steger, Helmut Mayer, and Bernd Radig. The role of grouping for road extraction. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pages 245–256. Birkhäuser Verlag Basel, 1997.

[24] Esko Ukkonen. On-line construction of suffix trees. *Algorithmica*, 14(3):249–260, 1995.

[25] Peter Weiner. Linear pattern matching algorithms. In *Proceedings of the 14th Annual Symposium on Switching and Automata Theory*, pages 1–11. IEEE Computer Society, 1973.

[26] Carola Wenk, Randall Salas, and Dieter Pfoser. Addressing the need for map-matching speed: Localizing global curve-matching algorithms. In *Proc. 18th SSDBM Conf.*, pages 379–388, 2006.

[27] Huijing Zhao, Jun Kumagai, Masafumi Nakagawa, and Ryosuke Shibasaki. Semi-automatic road extraction from high-resolution satellite image. In *Proc. ISPRS Technical Commission III Symposium: Photogrammetric Computer Vision*, page A:406, 2002.