In Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming (IWGS '14). ACM, New York, NY, USA,

Temporal Coverage Aggregates Over Moving Region Streams

Mark McKenney, Roger Frye, Zachary Benchly & Logan Maughan
Department of Computer Science
Southern Illinois University Edwardsville
{marmcke,rfrye,zbenchl,lmaugha}@siue.edu

ABSTRACT

A temporal coverage operation computes the duration that a moving object covers a spatial area. We extend this notion into temporal coverage aggregates, in which the spatial area covered for a maximum or minimum amount of time by a moving region, or set of moving regions, is discovered.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Spatial databases and GIS

General Terms

Algorithms

Keywords

Spatiotemporal data, Moving Regions, Data Models

1. INTRODUCTION

The growth of sensing platforms coupled with advances in cyber-infrastructure allowing the integration, storage, and availability of data has resulted in a rapid growth of spatiotemporal data that fits the paradigms of streaming data and stream analysis. Spatiotemporal data with extent, ie., data items represented as regions that change shape and position over time, provide significant opportunities in terms of i) analysis of the spatiotemporal attribute data associated with the data items, and ii) in terms of the information naturally encoded in the spatial and temporal behavior of the geometric representations of the data items themselves. Such data items, often referred to as *moving regions*, also present unique challenges: i) aggregate operations over moving regions are not fully defined and ii) algorithms to compute such operations tend to be complex since notions of a region's topological properties (ie., its interior, exterior and boundary) tend to have more distinct semantic differences in terms of data interpretation than do moving points or moving lines. Thus, the full algebra of analysis operations, especially at the implementation level, remains incomplete.

IWGS '14 November 04-07 2014, Dallas/Fort Worth, TX, USA
Copyright 2014 ACM 978-1-4503-3139-5/14/11 ...$15.00
http://dx.doi.org/10.1145/2676552.2676555.

In this paper, we consider the notion of purely temporal aggregate operations over moving regions. For example, given a set of moving regions, such that each moving region depicts the extent of hurricane force winds for a hurricane as it changes shape and position over time, an interesting operation is to discover the spatial areas covered by all hurricanes for the longest amount of time. Such areas represent places that have been impacted directly by hurricanes for the most amount of time, as opposed to places affected by the greatest number of hurricanes; for example, a single slow moving hurricane could inflict more damage on an area than several fast moving hurricanes. This type of query is agnostic of the attribute values associated with the moving regions, considering only spatiotemporal properties. We denote this operation the *max temporal coverage aggregate* (maxTCA) for a set of moving regions. Clearly the maxTCA has value in certain data sets, as does its converse, the *min temporal coverage aggregate* (minTCA), but such aggregates are not developed in moving region models. Furthermore, such aggregates have compelling use in stream processing in which, for example, the max time aggregate over the last hour of a data stream can be continually maintained. The contributions of this paper include: 1) formally defining the concept of temporal coverage aggregates for moving regions, 2) identifying specific aggregate operations, and 3) providing an algorithmic framework for computing these operations.

2. RELATED WORK AND DATA MODEL

Data Model: In this paper, we consider temporal coverage aggregates over *complex regions*. Complex regions, as defined in [6], are real data types representing a region containing one or more components, denoted *faces*, such that each face may contain zero or more *holes*. For example, the country of Italy can be represented as a single complex region in which the mainland and islands are each individual faces, with the face representing the mainland containing a single hole representing the Vatican City (which does not belong to Italy). In [2], the type of *moving regions* is defined as a mapping from time to the type of complex regions. Let $[R]$ be the set containing all complex regions and τ be the set of all time instants (represented as real numbers), a moving region M is then defined as $M : \tau \rightarrow [R]$, with some restrictions. We denote the set of all moving regions as $[M]$. Moving regions are able to represent regions that change in shape and position over time; for example, an area representing the extent of hurricane force winds for a particular hurricane is naturally represented as a moving region. At the discrete level, [4] represents a region as a collection of straight line segments that define the boundary of the region. Further, [4] defines a moving region as a collection of *moving segments*. A moving segment describes the motion of a line segment on the boundary of a region as it travels across a fixed time interval. Moving segments are typically represented as triangles in 3D space, showing a line segment as it contracts to a point over a time interval, or vice versa; we denote such a triangle a *delta triangle*. A collection of moving segments describing the motion of a region across a single time interval forms an *interval region*. Figures 1-4 depict interval regions, although delta triangles for segments that do not change shape or orientation over a time interval are merged into rectangles to reduce clutter. A collection of interval regions then forms a moving region. In this paper, we use the discrete model of moving regions, focussing on interval regions, to define temporal coverage aggregates.

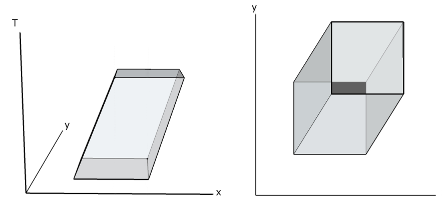
Aggregates and Geometric Algorithms: [5] presents a survery of spatiotemporal aggregate literature in which aggregates are categorized as spatial, temporal, or spatiotemporal. However, spatiotemporal aggregates in that paper focus mainly on the aggregation of attribute values related to spatiotemporal objects; this differs from our focus on aggregating purely temporal information from moving regions.

Additional research on aggregate operations over spatiotemporal data includes [3, 7, 8]. [7] uses histograms to compute aggregates from a database, rather than streaming-style data. [8] treats aggregation of sequenced spatiotemporal data in the form of points and lines, but not regions. Kim [3] views the changes to regions as a continuously changing phenomenon with respect to time and generates a space-time volume from known discrete positions of the region, similar to the approach described in this paper. However, Kim does not perform aggregates with respect to temporal coverage.

The algorithm described below requires computational geometry operations to compute the max (or min) depth with regards to the time extent in a three-dimensional volume. Chazelle [1] describes computing the maximum distance between a pair of points and the minimum distance between a point and a parallel plane through another point contained within a point set in three-dimensional space. While this is the closest computational geometry operations found, they do not match the requirements here.

3. TEMPORAL COVERAGE AGGREGATES

In this section, we build a framework upon which the temporal coverage aggregates are defined. The advantage of this framework is that it can be extended to define additional aggregates. We begin with the concept of a *temporal coverage mapping* (TCM) operation. The TCM takes an interval region and returns a set of temporally static spatial objects paired with a *duration*, a time value indicating the duration that the area represented by the spatial object was covered by the interval region. For example, Figure 1 depicts an interval region consisting of a rectangle that translates in space over a time interval. In the top-down view of the interval region, the shaded portion indicates where rectangles at the two ends of the interval region overlap if they are projected out of the temporal dimension into the spatial dimensions. This area is clearly covered by the interval region over the entire time interval. Thus, the TCM for that interval region will contain a rectangle corresponding to that shaded area, with a duration indicating the length of the time interval. Furthermore, the interval region translates away from the



(a) **(b)**

Figure 1: (a) An interval region consisting of a rectangle that translates over a time interval. (b) a top-down view of (a); effectively the moving segments of (a) projected out of the temporal dimension.

leftmost line segment in Figure 1, causing that line segment (among others) to be covered for only an instant during the time interval; thus, the TCM will include a line segment, representing that line in space, and a duration indicating that it was covered for a single instant. Similarly, each spatial area covered by a distinct duration will exist in the TCM result set. More formally, let $[\alpha]$ be the set of all instances of complex points, lines, and regions, and $[D]$ be the set of all duration values (typically $[D] = \mathbb{R}$):

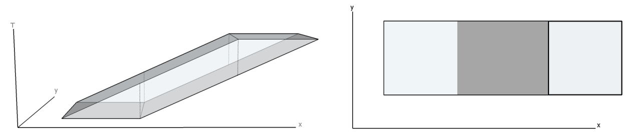
$$TCM : [M] \rightarrow 2^{[\alpha]\times[D]}$$

Temporal coverage operations (TCOs) are then defined by applying operations to the result of a temporal coverage mapping. Due to space limitations, we define two TCOs: *maxTCA* and *minTCA*. The *maxTCA* TCO returns the set of spatial areas covered for the maximum amount of time by an interval region; likewise, the *minTCA* operation returns the set of spatial areas covered for the minimum (non-zero) amount of time by an interval region. The syntax and semantics of the operations are as follows:

$maxTCA : 2^{[\alpha]\times[D]} \rightarrow 2^{[\alpha]}$
$maxTCA(TCM(m)) \rightarrow \{x_1|$
$(x_1, d_1) \in TCM(m) \wedge \forall (x_2, d_2) \in TCM(m) : d_1 \geq d_2\}$
$minTCA : 2^{[\alpha]\times[D]} \rightarrow 2^{[\alpha]}$
$minTCA(TCM(m)) \rightarrow \{x_1|$
$(x_1, d_1) \in TCM(m) \wedge \forall (x_2, d_2) \in TCM(m) : d_1 \leq d_2\}$

Computing TOCs: The semantic expression of the TCM described above is not appropriate for computation of TOCs since it may result in an infinite mapping. For example, the result may contain an infinite series of lines that each exist for a single instant (for example, the area traversed by the leftmost moving segment in Figure 1 has a monotonically increasing amount of temporal coverage, resulting in an infinite set of lines each covered by a monotonically increasing amount of time at infinite temporal resolution). In this section, we describe an algorithm to compute the maxTCA and minTCA TOCs using well known geometric operations.

We frame the discussion of the computation of TOCs in terms of the maxTCA operation, since the minTCA operation follows the same template. Recall that moving segments describing an interval region form triangles in 3D space; thus, a moving segment is allowed to travel in a single direction, and at a constant rate, across a time interval. It follows that if any area is covered by an interval region for an entire interval, then that area will be covered by the regions at the interval boundaries. It further follows that the maximum time that an area can be covered by an interval region is limited to the duration of the time interval for which the interval region is defined. These observations lead to the easiest case of determining a maxTCA operation: the case when the regions at both ends of an interval region



(a) (b)

Figure 2: A side (a) and top-down (b) view of an interval region.

overlap when they are projected out of the temporal dimension. Figure 1 depicts just such a case. Clearly, the dark shaded area in Figure 1b is covered by the interval region throughout the entire time interval, so it must be the area of maximum temporal coverage.

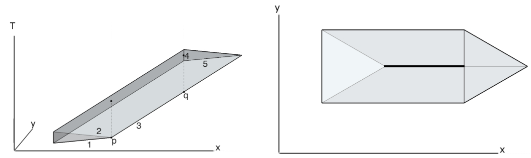
If the two regions at the temporal boundaries of an interval region are disjoint when projected out of the temporal dimension, the computation of TOCs becomes more difficult. The TCM produces a set of spatial objects (geometries) such that each entire object is covered by an interval region for the same duration; geometrically, this definition implies that the intersection of an interval region with any spatial object returned by a TCM that is extended infinitely in the temporal dimension is a 3-dimensional object g with constant height. In other words, the height of g is identical at every point covered by g in the spatial dimensions. For example, the shaded area shown in Figure 1b is the region of maximum temporal coverage, and thus the volume formed by extending that region in the temporal dimension across the duration of the interval region in Figure 1a will result in a 3-dimensional rectangle with constant height. Figure 2 shows a case in which the regions at the ends of the interval are disjoint. Figure 2b depicts a top down view of the interval region in Figure 2a. The shaded area in Figure 2b represents the region of maximum temporal coverage of the interval region in Figure 2a. Clearly, if the shaded region in Figure 2b is extended infinitely in the temporal dimension and intersected with the interval region in Figure 2a, the resulting volume will have constant duration for all points covered in the spatial dimension by the volume. The resulting volume is indicated by the dotted lines in Figure 2a. Note that the volume does not have constant values in the temporal dimension, but does have constant duration.

We make the observation that regardless of the dimension of the geometry of maximum coverage (ie., it may be a point, line, or region), the geometry of maximum coverage will always include a point on the boundary of at least one delta triangle; this follows from the fact that delta triangles are planar by definition. Furthermore, regardless of the dimension of the geometry of maximal coverage, each spatial object (point, line, or region) covered for the maximum amount of time by an interval region will be:

1. A point at which two boundaries of delta triangles intersect when projected out of the temporal dimension, denoted as *coverage points of interest* (CPOIs).
2. A line whose endpoints are CPOIs.
3. A region such that the end points of all line segments on the region's boundary are CPOIs.

We define the term *boundary points* of the geometry of maximum coverage to refer to the end points of line segments forming those geometries (or the point of maximal coverage in the case of a point geometry).

THEOREM 1. *The boundary points of all geometries of maximum coverage that exist for an interval region are always CPOIs.*



(a) (b)

Figure 3: A side (a) and top-down (b) view of an interval region.

Proof: In the case where the regions at either end of an interval overlap when projected out of the temporal dimension, it follows from the definition of regions that the geometry of maximum coverage is a region and that its boundary points are CPOIs. In the case where the regions at either end of the interval are disjoint, there are three cases.

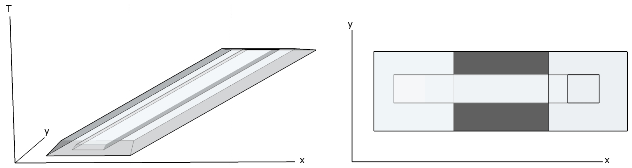
i) *The geometry of maximum coverage is a point.* In this case, the point will correspond to the maximum distance between two 3-dimensional triangles that are non-parallel and overlap when projected out of the temporal dimension, or that meet at a point when projected out of the temporal dimension. If the triangles are non-parallel, and they are furthest apart at a single point, that point must be on the boundary of at least one of the triangles, and thus is a CPOI.

ii) *The geometry of maximum coverage is a line* (Figure 3). Again, since moving segments correspond to planar delta triangles in 3-dimensional space, if the geometry of maximum coverage is a line, then that line must lie on the boundary of at least one moving segment. Furthermore, that line will extend until the boundary of one triangle is no longer over the other, meaning that the boundaries of the triangles will intersect when projected out of time. Thus the boundary points of the line must be CPOIs.

iii) *The geometry of maximum coverage is a region.* This case can only occur when the triangles representing moving segments across an interval are parallel (since the triangles are planar). In such a case, the geometry of maximum coverage is a region that corresponds to the intersection of the projections of the triangle out of the temporal dimension, meaning its boundary points must be CPOIs. □

Theroem 1 implies that to discover the geometry of maximum coverage for an interval region, one must simply look at CPOIs. Once the CPOIs that are covered by an interval region for the longest amount of time are known, the task is simply to merge CPOIs into lines and regions as appropriate.

The process of merging CPIOs into lines and regions of maximal temporal coverage is accomplished via a bookkeeping mechanism. To compute CPIOs, the edges of delta triangles are projected out of the temporal dimension, and the intersection of any of the edges results in a CPIO. Our approach to bookkeeping is to label each triangle edge; when two projected edges intersect, the CPIO is recorded once for each edge, and CPIO copies are labeled with the respective edge labels involved in the intersection. Once the maximal CPIOs are computed, CPIOs are sorted according to their label, and then their position along their assigned edge. If no holes are present in the interval region, adjacent CPIOs on an edge form a line segment. Line segments sharing end points are joined into lines. Finally, lines forming cycles are formed into regions. For example, consider Figure 3 in which some of the segments are labeled with numbers, and the points p and q are labeled. Segments 1, 2 and 3, when projected into the spatial dimensions, intersect at p; thus, the pairs $(p, 1), (p, 2)$, and $(p, 3)$ are recorded. Simi-



(a) **(b)**

Figure 4: A side (a) and top-down (b) view of an interval region.

larly, segments 3 and 4 will intersect at q, resulting in pairs $(q, 3)$ and $(q, 4)$ being recorded. The points are sorted first according to label, then according to position on the line: $(p, 1), (p, 2), (p, 3), (q, 3), (q, 4)$. p and q have the maximum time coverage, determined by extending p and q into the temporal dimension, and computing the length of intersection of the resulting lines with the volume induced by the interval region. Furthermore, p and q are adjacent on line segment 3, indicating that the entire line segment has maximal temporal coverage, and are joined into a line segment.

If an interval region contains holes, two additional steps are required. First, the adjacency of two points on a line segment does not guarantee that the line segment joining them has maximal temporal coverage by the interval region. For example, Figure 4 shows an interval region with a hole. In the top down view (Figure 4b), the left most CPIOs of the shaded region are all maximal coverage CPIOs that involve the same line segment, yet the hole causes a portion of the line segment to not be maximally covered. Thus, the adjacency of maximal coverage CPOIs is not sufficient to ensure a line segment is maximally covered. To verify, a single point along the original line between any two adjacent CPIOs must be tested for temporal coverage. If the coverage of that point is identical to the surrounding CPIOs on the line segment, then the line segment has maximal temporal coverage. Similarly, one cannot assume that a cycle of maximally covered line segments forms a maximally covered region. Instead, a point within such a cycle must be tested for identical time coverage as the cycle.

The complete algorithm is shown in Algorithm 1. In summary, the algorithm requires delta triangles to be projected into two dimensions (line 1). For an interval region with n moving segments, this produces $O(n)$ line segments. The intersection of those projected line segments are computed to determine labeled CPIOs (lines 2-3). Line segment intersection is $O(n \lg n + k)$ where k is the number of intersection points. The CPIOs are extended into the temporal dimension, and lengths for which they are contained within the interval region are recorded (lines 4-5). Because a volume representing a moving region is defined mainly as a set of triangles (moving segments), this a segment/triangle intersection problem. Naively, this step takes $O(kn)$, but can be reduced using advanced data structures to $O(k \lg n)$. The CPOIs with maximal length are sorted (line 6) ($O(k \lg k)$). A point between any pair of CPIOs that are adjacent on a line segment is tested to see if the segment is maximally covered (line 7). Again, this is a segment/triangle intersection problem. Segments are formed into lines and cycles (using binary searches on a sorted list in $O(k \lg k) time$. Finally, a point within each cycle is tested to see if the cycle is a maximally covered region (line 8) (another segment/triangle intersection problem). Thus, the algorithm has time complexity bounded by either $O(kn)$ if a naive segment/triangle intersection algorithm is used, or $O(k \lg k)$ otherwise.

Input: 3D line segments defining the edges of triangles that represent the moving segments defining an interval region.

Output: A set containing points, lines, and/or regions representing the areas covered for the largest amount of time by an interval region.

1 $S \leftarrow$ a set of line segments formed by projecting the 3D line segments out of the temporal dimension;
2 $P \leftarrow$ the set of intersection points of all pairs $s_1 s_2 \in S$;
3 $P_l \leftarrow$ For each $p \in P$, create a labeled point for each segment involved in the intersection at p;
4 $S_e \leftarrow$ Extend each point $p \in P_l$ in the temporal dimension;
5 $P_m \leftarrow$ the points corresponding to each segment $s \in S_e$ whose intersection with the interval region volume has maximal duration;
6 Sort P_m by label, then by position along the line segment corresponding to the label;
7 $S_m \leftarrow$ create a set of segments from pairs of adjacent points on a line segment (using the labeled points). Keep only those such that a point on the segment in between the adjacent points has the same temporal coverage as the adjacent points;
8 $R_m \leftarrow$ create cycles from segments with shared end points in S_m. Keep a cycle if a point within the cycle has the same temporal coverage as the cycle boundary, otherwise it forms a hole in a region. Merge cycles that share edges into regions;

Algorithm 1: Algorithm for computing the maximum and minimum temporal coverage of an interval region.

4. CONCLUSION

In this paper, we formally defined temporal coverage aggregates, provided an algorithm to compute the operations, and showed the correctness of the algorithm. Future work includes creating a prototype querying system to test how the algorithm scales over data sets.

5. REFERENCES

[1] B. Chazelle, H. Edelsbrunner, L. Guibas, and M. Sharir. Diameter, width, closest line pair, and parametric searching. In *Symposium on Computational Geometry*, pages 120–129. ACM, 1992.

[2] R. H. Güting, M. H. Böhlen, M. Erwig, C. S. Jensen, N. A. Lorentzos, M. Schneider, and M. Vazirgiannis. A foundation for representing and querying moving objects. *ACM Trans. Database Syst.*, 25(1):1–42, 2000.

[3] K.-S. Kim, K. Zettsu, Y. Kidawara, and Y. Kiyoki. Moving phenomenon: Aggregation and analysis of geotime-tagged contents on the web. In *Web and Wireless Geographical Information Systems*, pages 7–24. Springer, 2009.

[4] J. Lema, L. Forlizzi, R. H. Güting, E. Nardelli, and M. Schneider. Algorithms for moving objects databases. *The Computer Journal*, 46(6), 2003.

[5] I. F. V. Lopez, R. T. Snodgrass, and B. Moon. Spatiotemporal aggregate computation: A survey. *Knowledge and Data Engineering, IEEE Transactions on*, 17(2):271–286, 2005.

[6] M. Schneider and T. Behr. Topological relationships between complex spatial objects. *ACM Trans. Database Syst.*, 31(1):39–81, Mar. 2006.

[7] Y. Tao and D. Papadias. Historical spatio-temporal aggregation. *ACM Transactions on Information Systems (TOIS)*, 23(1):61–102, 2005.

[8] I. Timko, M. Böhlen, and J. Gamper. Sequenced spatiotemporal aggregation for coarse query granularities. *The VLDB Journal*, 20(5):721–741, 2011.