An Integrated Approach for Visual Analysis of a Multi-Source Moving Objects Knowledge Base

Niels Willems, Willem Robert van Hage, Gerben de Vries, Jeroen Janssens, Véronique Malaisé
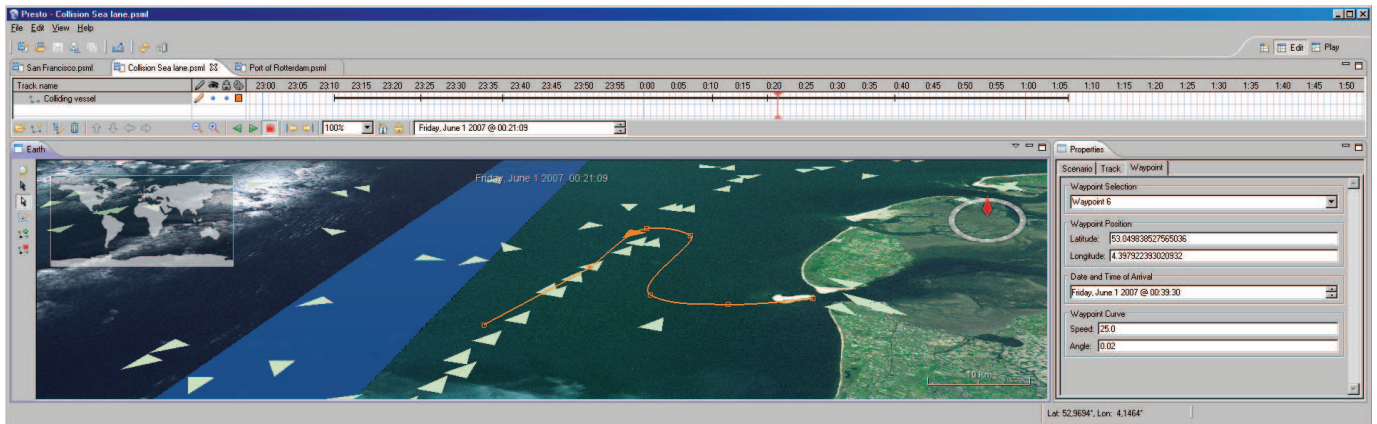
Fig. 1. Presto enables maritime-domain experts to insert artificial vessel trajectories in tracked AIS data. An artificial track, in orange, is added in the opposite direction of a Dutch sea lane that is visible by a trail of green triangles, which indicate tracked vessels.

Abstract—We present an integrated and multi-disciplinary approach for analyzing behavior of moving objects. The results are ongoing research of four different partners in the Dutch Poseidon project [3] where we aim for new developments for Maritime Safety and Security (MSS) systems to monitor vessels. We focus on the following requirements for an MSS system: abstraction of large amounts of trajectory sensor data, fusion of multiple heterogeneous data sources, and visual analysis of the combined data sources. We start by extracting segments of consistent movement from trajectory data, which we store as instances of the Simple Event Model (SEM), an event ontology represented in the Resource Description Framework (RDF). Then we add data from the web about vessels to enrich the sensor data. This additional information is integrated with the representation of the vessels (actors) in SEM. The enriched trajectory data is stored in a knowledge base, which is queried by a visual analytics tool to search for spatio-temporal patterns. Although our approach is dedicated to MSS systems, we expect it to be useful in other domains.

Index Terms—Trajectory Generation, Trajectory Generalization, Event Modeling, Semantic Web, Visual Analytics.

✦

1 INTRODUCTION

Since the early fifties [4], spatio-temporal patterns [1] have been studied in the trajectories of moving objects. Current technology can be used to register behavior on a large scale: many objects are tracked for long periods of time. By using data other than the geometry of a trajectory, we may find different spatio-temporal patterns [8]. For instance, when vessels are tracked, we may look up their type on the web and compare trajectories per type, and conclude that passenger ships use other routes than cargo ships.

We present a method that enables behavioral analysis of moving objects using multiple heterogeneous data sources. We demonstrate an integrated approach of multi-disciplinary research in the Dutch Poseidon project [3], which aims for new methods to monitor vessels for ensuring safety and security in coastal areas with so called Maritime Safety and Security (MSS) systems. Figure 2 shows an overview of our System of Systems (SoS); a collection of independent systems,

---

- *Niels Willems is with the Visualization group at Eindhoven University of Technology, E-mail: c.m.e.willems@tue.nl.*
- *Gerben de Vries is with the Theoretical Computer Science group at University of Amsterdam, E-mail: g.k.d.devries@uva.nl*
- *Jeroen Janssens is with Tilburg Centre for Creative Computing at Tilburg University, E-mail: j.h.m.janssens@uvt.nl*
- *Véronique Malaisé and Willem R. van Hage are with the Web & Media group at VU Amsterdam, E-mail: {vmalaise,wrvhage}@few.vu.nl*

which result in a system with more capabilities when put together. We have three low-level data sources: AIS messages from tracked vessels, generated AIS messages from Presto (see Sec. 2), and web data. The two former sources are converted into meaningful Simple Event Model instances (see Sec. 3.1) by Piecewise Linear Segmentation (PLS) (see Sec. 2.2). The web data are added to these event instances in an SWI-Prolog based RDF store (see Sec. 3.2). Rules can be used to define behavior on top of the movement events (see [13]). The visual analytics tool displays movements in a trajectory contingency table, enabling human users to find spatio-temporal patterns in the knowledge base (see Sec. 4). Conclusions and future work are discussed in Section 5.

2 VESSEL TRAJECTORY DATA

We focus on large vessels as moving objects, since they have to transmit their status using the Automatic Identification System (AIS) [6]. AIS is an advanced Global Positioning System (GPS) device that frequently broadcasts messages with data from the vessel and its movement. The data contains many attributes; for vessels we have identification numbers, a name, dimensions, and a type, e.g. passenger ship or tanker. The attributes for the movement are, for instance, position, time, velocity, destination, draught, and navigational status, e.g. at anchor, moored, or fishing. The latter comes from a controlled vocabulary in the AIS protocol. Captains use AIS to receive the status of neighboring vessels to prevent collisions, but AIS can also be received using a sensor network connected to an MSS system for real-time vessel monitoring and long-term analysis.
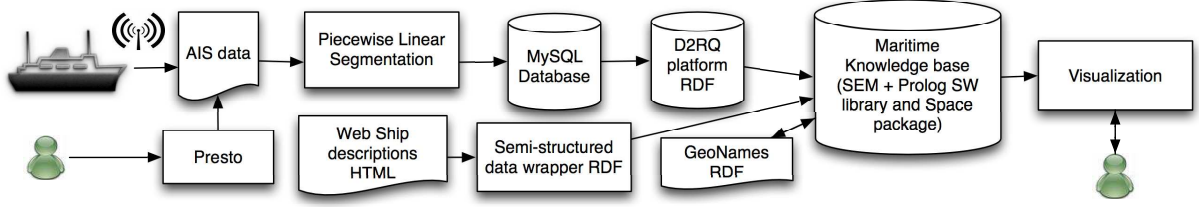
Fig. 2. Our approach as a System of Systems architecture for analysis of a multi-source moving object knowledge base.

In general, the continuous movement of an object, such as a vessel, is captured by sampling its trajectory at various moments in time. A sample contains all information available at that very moment about the movement of an object and about the object itself. We model a trajectory T as a sequence of n-tuples $\boldsymbol{\alpha}_i = (\alpha_0^i, \alpha_1^i, \ldots, \alpha_{n-1}^i)$, where the j-th element of the i-th tuple contains attribute α_j^i. Each tuple includes the attributes time t_i, position $\mathbf{p}_i$, velocity v_i, and object identifier o_i.

2.1 Creating Artificial Vessel Trajectories with Presto

We consider two types of AIS data: real-world trajectory messages emitted by vessels and manually created trajectory data. An MSS system aims at guiding a surveillance operator to find maritime anomalies, such as vessel traffic violations, illegal fishing activities, and drug smuggling. In real-world data, serious maritime anomalies rarely occur. Because the MSS system might be deployed in any maritime circumstance, we need to evaluate if arbitrary anomalies can be detected properly. To this end, we have developed a desktop application with NASA World Wind [9], called *Presto* (see Fig. 1) that enables maritime domain-experts to easily create artificial scenario's, including the above mentioned anomalies. In contrast to existing simulation applications, such as VR-Forces [15], which impose restrictive behavior models, our application gives the expert full control over the vessel trajectories, by means of defining way points. The new vessel trajectories are exported in AIS data format and fused with real-world AIS data, such that the artificial data cannot be distinguished up front from the original data by the system in the further processing.

2.2 Piecewise Linear Segmentation

Vessels frequently broadcast messages and are constrained in their movement. Consequently, they sail often along predictable courses, and therefore many subsequent messages $\boldsymbol{\alpha}_i$ do not semantically differ. In the analysis of logged AIS data, these messages are not meaningful for the movement pattern definition, hence we can abstract to meaningful trajectory segments by removing these redundant messages with a trajectory compression algorithm, like [7].

The temporal component is the main characteristic that differentiates a trajectory from a polyline, which is known to be efficiently compressed by the Douglas-Peucker (DP) line generalization algorithm [2]. Figure 3 illustrates a generic version of the DP algorithm that compresses a sequence of tuples $\boldsymbol{\alpha}_i$ by inspecting the largest error $\varepsilon_{max} = E(\boldsymbol{\alpha}_{max}, \boldsymbol{\alpha}'_{max})$ between a tuple $\boldsymbol{\alpha}_{max}$ and the reconstructed tuple $\boldsymbol{\alpha}'_{max} = R(\boldsymbol{\alpha}_{start}, \boldsymbol{\alpha}_{max}, \boldsymbol{\alpha}_{end})$ at tuple $\boldsymbol{\alpha}_{max}$, for some error function E and reconstruction function R. If ε_{max} is larger than some given error ε, then we select $\boldsymbol{\alpha}_{max}$ and recursively compress the sequence $\boldsymbol{\alpha}_{start} \ldots \boldsymbol{\alpha}_{max}$ and $\boldsymbol{\alpha}_{max} \ldots \boldsymbol{\alpha}_{end}$, otherwise the algorithm stops and accepts $\boldsymbol{\alpha}_{start} \ldots \boldsymbol{\alpha}_{end}$. For polylines, E is the Euclidian distance and R is the projection of $\mathbf{p}_{max}$ on the line $\overline{\mathbf{p}_{start}\mathbf{p}_{end}}$.
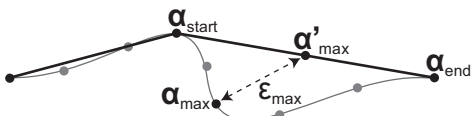

Fig. 3. A compressed trajectory after one step with the DP algorithm.

We are investigating modifications of the DP algorithm for trajectories. Our data is compressed with a two-stage modification, where in the first stage tuples $\boldsymbol{\alpha}_i$ are selected using the generic DP algorithm to reconstruct all velocities v_i within error ε_v, and then, between each pair of selected points, the positions are compressed as a polyline. The velocity v'_{max} is reconstructed with R being the linear interpolation between v_{start} and v_{end} at t_{max} and the error E is the absolute difference. The segments of the compressed trajectories are stored in a MySQL database and constitute the first level of abstraction for meaningful movements, which can be agglomerated in complex behavior definition using the Simple Event Model (see Sec. 3.1).

3 SEMANTIC WEB TECHNOLOGY

Behavior can be derived from low-level movements, such as the segments obtained from the PLS in Section 2.2. Behavior definition requires the combined knowledge about an item's position, movement characteristics and its type information: the same movement pattern may have different implications across object types. For instance, if a vessel shuttles between two locations, it may be a dredger or a ferry. If we know that a vessel is a dredger, then one of the locations is a dredging place. Oppositely, if one of the locations is a dredging place, then the vessel is a dredger. In order to reason about information coming from different sources, it needs to be integrated. For this purpose, we have developed the Simple Event Model.

3.1 Simple Event Model

The Simple Event Model (SEM) [13] provides a set of classes and properties to define events and their context. SEM can be used to describe trajectories as events. SEM's classes are organized in three groups (see Fig. 4). SEM Core classes describe the *classic* parts of an event: *What is happening?* (sem:Event, e.g., anchoring, fishing), *Who is doing something?* (sem:Actor, e.g., a vessel), *Where?* (sem:Place, e.g., in Rotterdam), and *When?* (sem:Time, e.g., during year 2008). SEM Types can be linked to each of the Core classes, which typically come from domain-specific vocabularies, like the IMO thesaurus [5] or a vessel classification, such as the vessel types in [6, 14]. SEM Constraints express Temporary, Authoritative and Role constraints on the properties between SEM Classes. Roles can apply to the participation of an Actor in an Event (an Actor *vessel* has the Role *coalition leader* during an Event *NATO joint operation*), but also to the Location of the Event (Place *Rotterdam* has the Role *journey destination* in a trip event). Time and Place can have symbolic (URIs) or concrete values (e.g., coordinates), to fit the representation of most data on the Web. All the classes and properties of SEM are optional and duplicable: SEM can integrate the partial information provided by different sources as different aspects of an event. It is possible to create instances of sem:Actors without specifying explicitly the sem:Event they participate in: we can describe vessels (e.g., information from websites) independently from their behavior description.

The instances of SEM are either accessed from the MySQL database via the D2RQ platform [12] (for AIS events) or directly form the SWI-Prolog based triple store [11] (for events extracted and converted from the web) by the rules defining ship types and ship behavior. The rules are written in SWI-Prolog and the triple store is a knowledge base that enables us to link the different parts of an event represented in SEM with semantic web and spatial reasoning.
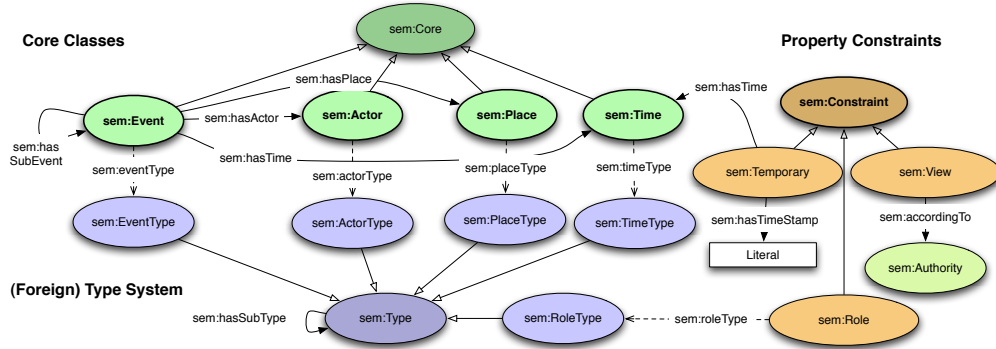
Fig. 4. The Simple Event Model. Ellipses represent classes, arrows represent properties between them and square boxes represent datatypes. Open-ended arrows represent subclass properties, dotted arrows are typing properties and the regular arrows' semantics is given by their labels.

3.2 Moving Object Knowledge Base

Instances of SEM events from AIS are accessed as RDF [17] data from the D2RQ platform, and are available for reasoning in the knowledge base. To be able to do efficient reasoning about ship behavior, more information and data sources than the AIS messages have to be taken into account. For example, some websites contain information about ship properties, some others contain information about geographic places. Some information is already present as RDF on the web (`DBpedia.org` and `GeoNames.org`), while most is still exposed as HTML. We give a simple example of how different dimensions about a single SEM event can be collected from the web and used together within a Prolog rule that defines the notion of *ship from the European Union*. Other rules can then build on this notion of *European ship* and define behavior patterns at this new level of abstraction. AIS messages contain some identifying information about the broadcasting ship: name, call sign, MMSI and IMO [5]. We use the two latter to automatically query the web [14] for additional information about the ship, and convert them semi-automatically to RDF. One piece of such information is the flag of the ship: the name of the country in which it is registered. The flag is captured as a mere string, but this string is in turn used to automatically query the GeoNames' RDF database, to retrieve the URI and RDF knowledge that corresponds to this flag. The SWI-Prolog Semantic web package [11] can then be used to this enriched data. A Prolog rule combining RDF graph query and spatial hierarchical reasoning retrieves the parent feature of the URI corresponding to the ship's flag, in GeoNames, and adds a new triple to the knowledge about the corresponding SEM event's Actor: the ship sending the AIS message. As the parent feature of a country in GeoNames is the continent it belongs to, the new triple added is the fact that the ship's flag belongs or not to a country from the European Union.

SEM is the central piece for aggregating knowledge and resources at different levels of abstraction, that can be used in complex rules defining ship behavior. These complex events are also multidimensional data that can be explored by humans, to find manually patterns relevant to a certain domain or situation. A powerful visualization tool is needed to take advantage of the richness of the knowledge and present it in a useful way to a human user.

4 VISUALIZATION WITH TRAJECTORY CONTINGENCY TABLES

The knowledge base can be browsed with an interactive Trajectory Contingency Table (TCT). This visualization aims for expert users, which want to discover relations between attributes in terms of spatial patterns, such as the change over *time* for different *types* of vessels.

The visualization retrieves trajectories by querying the knowledge base with SPARQL [18] queries (Fig. 5), which convert segments to trajectories. First, we determine which elements and types the tuples α_i have (Fig. 5a), by searching for owl:DatatypeProperty properties that contain data and rdf:datatype properties that contain the data types, such as xsd:int and xsd:string. Then, for each vessel identified with an MMSI (Fig. 5b), a trajectory $\alpha_0, \ldots, \alpha_{N-1}$ is reconstructed (Fig. 5c) by gath-

ering all attributes of the Actor with the current MMSI, as illustrated with ais:name, and the attributes of the accompanying events as illustrated with sem:hasBeginTimeStamp. Since the location is a sem:Place class it is treated separately, and allows us to use the external WGS84 vocabulary [16]. To avoid duplications, attributes for Events and Actors are stored separately.

In a contingency table [10] two categorical attributes are put on the axes of a table by assigning (groups of) outcomes to a row (or column). In a cell, the number of data items are shown that contain the outcomes of the corresponding row and column for the attributes assigned to the axes. Furthermore, each axis has an additional row that shows for each column the sum of the items. Instead of showing a number in the table, we display a map with the parts of tracks that satisfy the constraints of the cell (see Fig. 6). For instance, by putting *time* and *type* on the axes, we see the change of spatial usage over time for various types.

```
1  # A. Get attributes with their types
2  SELECT DISTINCT ?attr ?type WHERE {
3    ?attr a owl:DatatypeProperty .
4    ?attr rdf:datatype ?type }
5
6  # B. Get all MMSI ship identifiers
7  SELECT DISTINCT ?mmsi WHERE { ?ship ais:mmsi ?mmsi }
8
9  # C. Get the trajectory for a certain [MMSI]
10 SELECT DISTINCT ?time ?lat ?lon ?name WHERE {
11   ?ship ais:mmsi ?mmsi . FILTER ( ?mmsi = '[MMSI]' )
12   ?event sem:hasActor ?ship .
13   ?event seg:hasBeginPlace ?place .
14   ?place wgs84:lat ?lat .
15   ?place wgs84:lon ?lon .
16   # An example of a trajectory attribute
17   ?event sem:hasBeginTimeStamp ?time .
18   # An example of a ship attribute
19   ?ship ais:name ?name } ORDER BY ?time
```

Fig. 5. SPARQL queries to reconstruct trajectories from SEM instances.

Since the input data is large and consists of many attributes, the TCT is extended with two views: Attributes and Overview (see Fig. 6). In the Attributes view on the left, all attributes are listed and divided in bins, which are shown as a histogram. The bins are selectable and filter the whole data set. For example, by choosing the *tankers* bin of the *vessel type* attribute only tankers are in the TCT. These attributes can be dragged to the axes of the TCT to quickly choose different settings. In the overview on the right of the TCT, the selection of data made by choosing bins (focus) is highlighted on top of all data (context). In the overview it is possible to brush parts of the focused trajectories to define areas, which can be used to define a new boolean attribute stating whether or not a segment is contained in the area.
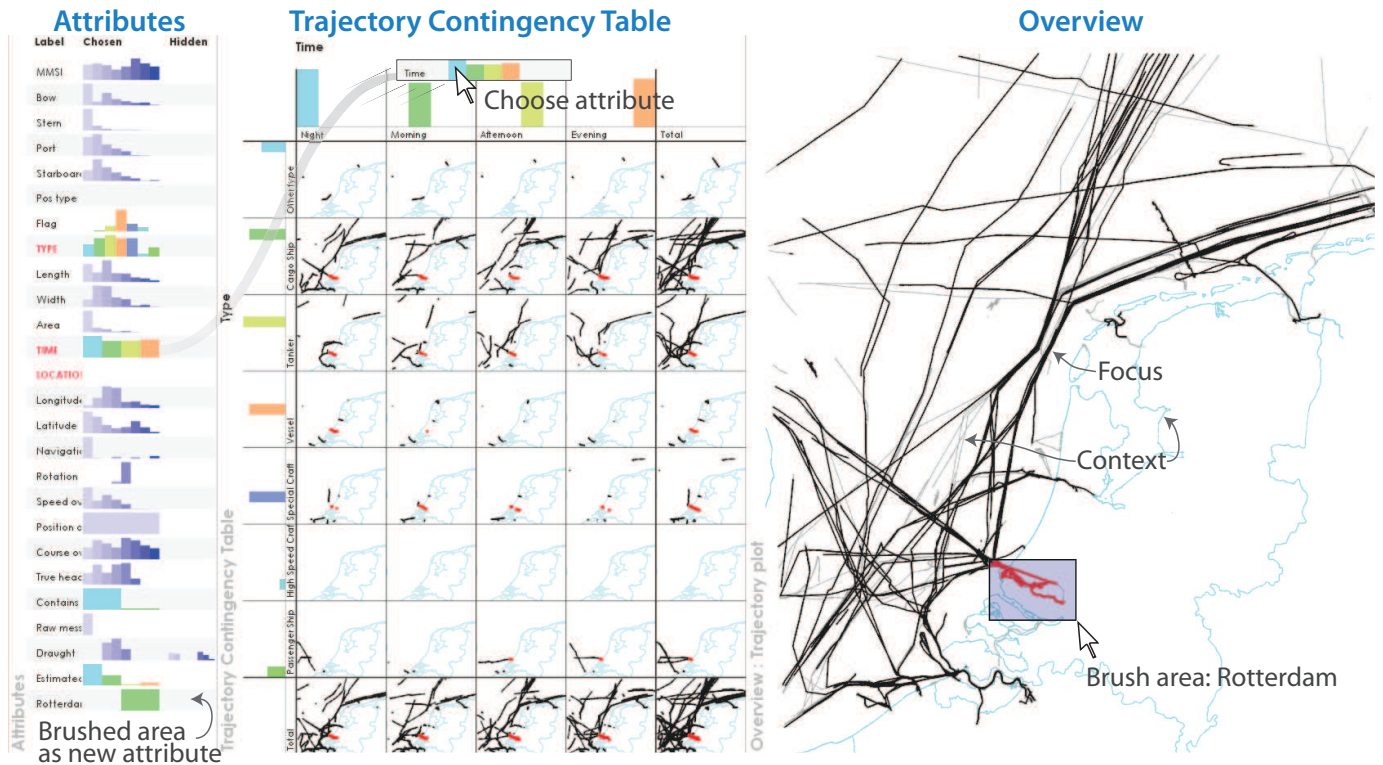
Fig. 6. A Trajectory Contingency Table (TCT) with an Attribute view (left) and an Overview (right). About 200 vessel trajectories in front of the Dutch coast during a single day are shown. In the Attributes view, attributes are divided in bins and listed with two histograms: chosen bins and hidden bins. The draught of selected ships is between 3.1m and 7.8m, for the data in all views. The TCT is displayed with *time* and *vessel type* attributes on the axes. Each cell contains a map with parts of trajectories that satisfy the accompanying row and column labels. From the TCT we may conclude the following: First, vessels have a steady spatial pattern across the temporal period considered. Second, tankers and cargo ships tend to have a similar spatial pattern. Third, passengers ships sail a commuting pattern. To conclude, other vessel types stay close to the coast. In the overview, selected trajectories (focus) are highlighted in dark gray on top of the context containing all data in light gray and a map. By brushing we define areas to define new attributes.

5 CONCLUSION AND FUTURE WORKS

As result of the Dutch Poseidon project, illustrated by vessel traffic data, we have presented an integrated approach for analyzing moving object data. The approach includes trajectory generation using *Presto*, trajectory compression using *Piecewise Linear Segmentation*, trajectory modeling by the *Simple Event Model*, which is the data model of our knowledge base that can be visualized using a *Trajectory Contingency Table*.

In future research we will include high-level abstractions of trajectories in SEM. Furthermore, we would like to efficiently describe and reason with the relative position of moving ships. The main challenge in the knowledge base will be reasoning with automatically discovered information, such as a location that is described as harbor on the web. The trajectory contingency tables need support to select subsets and compare them for various attributes simultaneously. Finally, we will investigate whether our approach can be used in other domains.

REFERENCES

[1] S. Dodge, R. Weibel, and A.-K. Lautenschütz. Towards a Taxonomy of Movement Patterns. *Information Visualization*, 7(3–4):240–252, 2008.

[2] D. H. Douglas and T. K. Peucker. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. *Cartographica*, 10(2):112–122, October 1973.

[3] Embedded Systems Institute. The Poseidon project: System Evolvability and Reliability of Systems of Systems, www.esi.nl/poseidon.

[4] T. Hägerstrand. *Innovation Diffusion as Spatial Process*. University of Chicago Press, Chicago, Il, 1968. Translation of: T. Hägerstrand, Innovationsförloppet ur Korologisk Synpunkt, 1952.

[5] International Maritime Organization, www.imo.org.

[6] ITU. Technical Characteristics for an Automatic Identification System Using Time Division Multiple Access in the VHF Maritime Mobile Band. *Recommendation ITU-R M.1371-1*, 2001.

[7] J. Gudmundsson *et al.*. Compressing Spatio-temporal Trajectories. *Computational Geometry*, 42(9):825–841, 2009.

[8] N. Andrienko *et al.*. Designing Visual Analytics Methods for Massive Collections of Movement Data. *Cartographica*, 42(2):117–138, 2007.

[9] NASA. World Wind, worldwind.arc.nasa.gov.

[10] K. Pearson. On the Theory of Contingency and its Relation to Association and Normal Correlation. *Drapers' Company Research Memoirs. Biometric Series I.*, 1904.

[11] SWI-Prolog for the (Semantic) Web, www.swi-prolog.org/web.

[12] The D2RQ Platform, www4.wiwiss.fu-berlin.de/bizer/d2rq.

[13] W. R. van Hage, V. Malaisé, G. de Vries, G. Schreiber, and M. van Someren. Combining Ship Trajectories and Semantics with the Simple Event Model (SEM). In *EiMM '09: Proceedings of the 1st ACM international workshop on Events in multimedia*, pages 73–80, 2009.

[14] Vessel Tracker, www.vesseltracker.com.

[15] VR-Forces, www.mak.com/products/vrforces.php.

[16] W3C. Basic Geo Vocabulary, www.w3.org/2003/01/geo.

[17] W3C. Resource Description Framework (RDF), www.w3.org/RDF.

[18] W3C. SPARQL Specification, www.w3.org/TR/rdf-sparql-query.