Discovering Relative Motion Patterns in Groups of Moving Point Objects

Patrick Laube, Stephan Imfeld, Robert Weibel

Department of Geography

University of Zurich

Winterthurerstrasse 190

8057 Zurich, Switzerland

telephone: +41 1 635 51 31

fax: +41 1 635 68 48

{plaube, imfeld, weibel}@geo.unizh.ch

December 4, 2009

Abstract

Technological advances of position aware devices lead to a wealth of data documenting the lifelines of GPS-tracked animals, people, or vehicles. The integration of knowledge discovery techniques in GIS is thus an important research field with respect to the emerging volumes of tracking data. The motivation of this research is to find patterns in the way such moving point objects (MPOs) move in space and time. The main objective is developing generic spatio-temporal knowledge discovery and data mining techniques to overcome the limitations of static GIS. The proposed RElative MOtion (REMO) concept transforms the geospatial lifelines of MPOs into an analysis matrix which allows motion pattern matching. The paper proposes a REMO pattern description formalism adopting elements of the commonly used regular expression formalism (regex) and of basic mathematical logic. An application prototype featuring REMO data mining algorithms has been implemented and tested with two use cases: tracked soccer players and data points moving in an abstract ideological space. In both use cases a set of non-trivial and meaningful motion patterns could be identified. The results indicate that knowledge discovery techniques applied to lifeline data reveal much more motion patterns than simple eye-balling.

Keywords: *Moving point objects, spatio-temporal data mining, knowledge discovery in databases, pattern matching, relative motion, temporal granularity*

1 Introduction

Moving point objects (MPOs) are a frequent representation for a wide and diverse range of phenomena: for example animals in habitat and migration studies (e.g. Ganskopp 2001, Sibbald *et al.* 2001), vehicles in fleet management (e.g. Miller and Wu 2000), agents simulating people for modelling crowd behaviour (e.g. Batty *et al.* 2003) and even tracked soccer players on a football pitch (e.g. Iwase and Saito 2002). All those MPOs share motions that can be represented as *geospatial lifelines*: a series of observations consisting of a triple of *id*, *location* and *time* (Mark 1998, Hornsby and Egenhofer 2002).

Gathering tracking data of individuals has become much easier nowadays due to substantial technological advances in position-aware devices such as GPS receivers, navigation-systems and mobile phones. The increasing number of such devices will lead to a wealth of data on space-time trajectories documenting the spatio-temporal behaviour of animals, vehicles and people for off-line, and potentially even on-line, analysis. These collections of geospatial lifelines present a rich environment to analyse individual and group behaviour. (Geographic) knowledge discovery may detect patterns

and rules to gather basic knowledge about dynamic processes or to design location based services (LBS) (Mountain and Raper 2001, Smyth 2001, Miller 2003).

Whereas the early days of Haegerstrand's time geography were limited to a data-poor and computation-poor environment, nowadays spatio-temporal knowledge discovery is fostered by data-rich and computation-rich environments. Knowledge discovery in databases (KDD) and its component data mining are reasonable responses to the huge data volumes in operational and scientific databases. Data mining is able to distill data into information and KDD turns information into knowledge about the monitored world. The central belief of KDD is that information is hidden in very large databases in the form of interesting *patterns* (Miller and Han 2001). This statement is equally true for the spatio-temporal analysis of geospatial lifelines and is thus a key motivator for this research.

The long tradition of data mining in spatio-temporal databases is well documented, for an overview see Roddick *et al.* (2001). The Geographic Information Science (GISc) community has also recognized the potential of KDD. Miller (2003, pg. 450) remarks:

'GIS can allow capture, representation, analysis and exploration of massive STA (space-time-attribute) databases, potentially leading to unexpected new knowledge about the interactions between people, technologies and urban infrastructures'.

Unfortunately, most GIS are based on a static place-based perspective and are thus still notoriously weak in providing tools for handling the temporal dimensions of geographic information (Mark 2003). Some GISc approaches integrate time in using the concepts of Haegerstrand's time geography with its space-time prisms (e.g. Miller 1991, Miller and Wu 2000, Hornsby and Egenhofer 2002). The space-time prism is a useful concept to analyse the motion constraints of a few individuals, but is limited to very small numbers of individuals. Miller postulates expanding GIS from the place-based perspective to encompass a people-based perspective. Therefore he identifies the development of a formal representational theory for dynamic spatial objects and of new spatio-temporal data mining and exploratory visualisation techniques as key research issues for GISc (Miller 2003).

In this paper work is presented which refines and extends a previously developed concept to analyse relative motion patterns for groups of MPOs (Laube and Imfeld 2002). The research allows the formalisation and identification of generic motion patterns in tracking data and extracting instances of these formalised patterns. The remainder of this paper is structured as follows. Section 2 gives a short review of the RElative MOtion (REMO) analysis concept. Section 3 proposes an object-oriented approach to model and analyse the motion of many MPOs and places special emphasis on dealing with varying temporal granularities. Section 4 introduces a formalism to describe and match motion patterns on geospatial lifelines. Section 5 illustrates an application prototype that implements the REMO analysis concept. Section 6 reports on tests with two use cases. Section 7 discusses the concept, compares it with comparable approaches and identifies strengths and open problems. Finally, conclusions are presented in section 8.

2 The basic REMO analysis concept

The basic idea of the REMO analysis concept is to compare the motion attributes of point objects over space and time, and thus to *relate* one object's motion to the motion of all others (Laube and Imfeld 2002). Suitable geospatial lifeline data consist of a set of MPOs, each featuring a list of fixes. The REMO concept is based on two key features: Firstly a transformation of the lifeline data to a REMO matrix featuring motion attributes (i.e. speed, change of speed or motion azimuth). Secondly formalised patterns are matched on this REMO matrix (figure 1).

(Insert figure 1 about here.)

Two simple examples illustrate the REMO analysis concept: Let the geospatial lifelines in figure 1(a) be the tracks of four GPS-tracked deer. Deer O_1 is moving with a constant motion azimuth of 45° during an interval t_2 to t_5, i.e. four discrete time steps of length Δt. It is showing a *constance.*

In contrast, four deer performing a motion azimuth of 45° contemporaneously t_4 show *concurrence*. The REMO concept allows construction of a wide variety of motion patterns, for example:

- *Constance:* Sequence of equal motion attributes for r consecutive time steps (e.g. deer O_1 with motion azimuth 45° from t_2 to t_5)
- *Concurrence:* Incident of n MPOs showing the same motion attributes value at time t (e.g. deer O_1, O_2, O_3 and O_4 with motion azimuth 45° at t_4)
- *Change:* Change in an MPO's motion attributes of value v over r time steps (e.g. deer O_4 changes its motion azimuth from 90° to 0° during t_3 to t_5)
- *Trend-setter:* One trend-setting MPO anticipates the motion of n others. Thus, a trend-setter consists of a *constance* linked to a *concurrence* (e.g. deer O_1 anticipates at t_2 the motion azimuth 45° that is reproduced by all other MPOs at time t_4)

For simplicity we focus in the remainder of this paper on the motion attribute azimuth, even though most facets of the REMO concept are equally valid for speed or change of speed as well as other attributes that might be determined.

3 Modelling motion attributes of MPOs

The REMO analysis concept deals with MPOs and events in their geospatial lifelines. The object nature of MPOs, events and instances of patterns is evident. Thus, the REMO world is an object-oriented world, and so is its data model.

3.1 Imperfect data and temporal zooming

The representation of the continuously changing position of an MPO is inherently associated with uncertainty (Pfoser and Jensen 1999). How can the gaps between known positions be filled (Wentz *et al.* 2003)? The problem increases if the fixes are not sampled regularly but irregularly with changing intervals and sometimes missing values - a typical property of data originating from position aware devices. Thus, the MPO data model should offer ways of handling gaps to produce regularly sampled analytical derivatives such as a REMO matrix.

'Although most physical theories are spatially and temporally continuous the evaluation of these theories almost always requires the existence of entities which are spatially and temporally discrete' Raper and Livingstone (1995, pg. 362). Raper and Livingstone call this discretisation an 'entification at a particular granularity'. Whereas space is usually dicretised using the vector or raster models, time is normally discretised as a series of events or recordings. It is quite obvious that the shorter the temporal intervals, the more accurate the representation will be (Kemp and Kowlaczyk 1994).

In the discretisation information may be lost and a fragmentary picture of the real world process may emerge. Whereas the term resolution refers to the least detectable difference in a measurement, granularity stands for the level of detail, selectable by the user. Thus, the finest granularity need not be the most adequate for the analysis of a phenomenon. Consequently changing the temporal granularity, referred to as temporal zooming, is an important requirement for many scientific questions (Hornsby 2001). In spatio-temporal knowledge discovery changing between more and less detailed views is an essential feature to solve a problem or to uncover information (e.g. Bettini *et al.* 2000). Support of temporal zooming is an important requirement on a data model for MPOs.

3.2 Data model

The most prominent characteristic of the developed REMO data model is the strict separation of the MPO MODELLING DOMAIN and the ANALYSIS DOMAIN (figure 2). This separation allows the original tracking data to be managed separately from the analysis task. Thus, MPOs always keep their exact lifeline data and compute their motion attributes only on request for the analysis phase.

The class `MPO` maintains a list of `Fixes`, each holding a location `Point`. Based on these data, the MPO answers any spatio-temporal request on its (motion) attributes. The `REMOMatrix` requests the motion attributes of the MPOs in accordance with its motion parameters, granularity and temporal extent. *REMOPatterns* search their realisations (matches) on the *REMOMatrix*.

(Insert figure 2 about here.)

Figure 3 shows the cooperation of the MPO MODELLING DOMAIN with the ANALYSIS DOMAIN. The `REMOMatrix` requests the motion attributes at its specified temporal sampling rate. For a finer sampling rate the `REMOMatrix` requests many motion attribute values, hence the involved `MPOs` may have to interpolate between fixes. For a coarser sampling rate the `REMOMatrix` requests only few motion attribute values, hence `MPOs` may have to aggregate motion attributes. The imperative of handling irregularly sampled fixes and variable temporal granularities therefore implies two problems: (1) Interpolation of motion attributes in the case of irregular fixes or gaps of ignorance; (2) aggregation of motion attributes in the case of a coarse analysis granularity over a much finer fixing rate.

The REMO data model resolves these issues with *detached attribute functions*. `MPOs` offer *attribute functions* to describe their motion based on their lifeline data (e.g. `getAzimuth()`). This can happen at any desired query time and with any desired sampling rate; the function is *detached* from the actual fixes. Thus the functionality for temporal and/or spatial interpolation or aggregation of motion attributes and associated questions of fix uncertainty is encapsulated in the MPOs. This architecture allows the integration of various detached attribute functions, using nearest neighbour functions, focal functions or curve fitting approaches. Nearest neighbour functions compute the motion attributes considering simply the two nearest neighbouring fixes of query time t. The focal functions use a moving window of length Δt around query time t to smooth inaccurate or incomplete data. Yet another approach is to first fit a smooth curve to the possibly scattered fixes and derive motion attributes from this smoothed lifeline. In short, the MPO MODELLING DOMAIN defines *how* to interpolate/aggregate motion attributes, the ANALYSIS DOMAIN defines *when* this happens (figure 3).

(Insert figure 3 about here.)

4 Pattern description formalism

The REMO analysis concept follows the syntactic pattern detection approach. After Jain *et al.* (2000) syntactic pattern recognition adopts a hierarchical perspective where a pattern is viewed as being composed of simple sub-patterns, the primitives. Complex patterns are represented in terms of interrelationships between primitives. A formal analogy can be drawn between the structure of patterns and the syntax of a language. The patterns are viewed as sentences belonging to a language, primitives are viewed as the alphabet of the language and the sentences are generated according to a grammar. Thus, in principle any arbitrarily complex pattern can be described by a set of primitives and grammatical rules (Jain *et al.* 2000).

The following section describes the formal language used to describe the primitives and the complex patterns of the REMO analysis concept.

4.1 Scope of the formalism

The REMO analysis concept is designed to be a flexible and intuitive tool for researchers. Thus it is an important precondition that users can compose patterns in a simple and flexible way. To allow flexible exploratory data analysis, the patterns need parameters and descriptors to adjust their size and shape. Not only the basic motion patterns described in Section 2 shall be formalised, but also user-definded arbitrary patterns. This goal is best achieved with the development of a generic REMO pattern description formalism. A pattern description formalism furthermore is a prerequisite for the automation of the pattern detection mechanism. For the REMO analysis concept we propose

a formalism related to the commonly used regular expression formalism and mathematical logic to describe the REMO patterns.

4.2 Using concepts of regular expressions

The commonly used regular expressions (regex) are a way of describing a set of strings without having to list all the strings in the set (Wall *et al.* 1996). Regex are basically used to determine whether a string matches a particular pattern. Commonly, regex are used to search, edit and manipulate string data (e.g. Wall *et al.* 1996, Friedl 2002). Regex are widespread in UNIX programs (grep), editors (emacs, vi) and programming languages (perl, java, Tcl, Python). As an example the regular expression $a\{3,5\}$ matches the bold characters in the following string:

*bcbaabc**aaa***cbc**aaaaa***ccc*

There are three major differences between pattern matching in strings using regex and REMO pattern matching. First, and most obviously, the REMO analysis concept requires two-dimensional pattern matching on a two-dimensional matrix whereas regex is normally used to match patterns on one-dimensional strings. Second, pattern matching on the REMO analysis matrix is confronted with two different types of dimensions. Whereas the temporal axis is an interval scale, and thus comparable to sequences of strings, the object axis implies no order among the objects. Thus, adjacency among objects is arbitrary (figure 1). Thirdly, the elements constituting a pattern are numbers on a continuous scale rather than characters and numerals as with regex. These three differences lead to the differences from basic regex in the formation of the REMO pattern matching expressions.

Nevertheless, many features of regex are very similar to the requirements for a REMO pattern matching formalism. Whenever possible, the REMO formalism uses familiar regex structures to express REMO patterns. That applies, for instance, to the pattern descriptors referring to the pattern's length and width (see Sections 4.3.2 and 4.3.3). Slight changes had to be made to describe the pattern depth and the range of attribute values building a pattern. The most prominent changes emerged from the need to combine *simple patterns*, in either the temporal or the objects dimension, to build the *complex patterns* extending across dimensions, such as for a trend-setter.

4.3 Structural elements of the REMO pattern formalism

The most important principle of language design says that easy things should be easy and hard things should be possible (Wall *et al.* 1996). This rule also applies for the development of the REMO formalism. The easy tasks are expressing the simple patterns akin to known regular expressions; the harder problems involve the formalization of the complex patterns over two matrix dimensions.

4.3.1 Simple patterns over time or across objects

First of all the dimension of the simple patterns in the analysis matrix must be specified. A pattern P over time describes a sequence S of motion attribute observations A_m (1). A pattern P across objects describes an incident I of a set of concurrent motion attribute observations A_m (2).

$$P = S(A_m) \tag{1}$$

$$P = I(A_m) \tag{2}$$

The REMO formalism allows formalism of identical patterns in different but synonymous forms. Quantifiers allow specifying A_m. Attribute values v are given in 'brackets'. REMO patterns span time and across objects, their extent is expressed with quantifiers in 'braces'. Patterns over time have a lenght l, patterns across objects have a width n (3 and 4).

$$P = S(v_1, v_2, v_3, \ldots, v_l) = S([v]\{l\}) \tag{3}$$

$$P = I(v_1, v_2, v_3, \ldots, v_n) = I([v]\{n\}) \tag{4}$$

A change C is a special form of a pattern over time. It is formalized using an indicator for a starting attribute value v in 'brackets', for the value of change δv and a length l in 'braces' (5). The starting attribute value indicator $[v]$ is optional, since a starting attribute value is not always desired.

$$P = C([v]\{\delta v\}\{l\}) \tag{5}$$

Change may be directional, some attribute sets are cyclically closed (e.g. the compass rose for the motion azimuth). Increasing change is indicated using $+\delta v$, decreasing change using $-\delta v$ (6 and 7).

$$P = C([v]\{+\delta v\}\{l\}) \tag{6}$$

$$P = C([v]\{-\delta v\}\{l\}) \tag{7}$$

Examples A single deer heading north-east for four consecutive time steps is formalised as $P = S(45, 45, 45, 45)$ or $P = S([45]\{4\})$ (figure 4a). In contrast four deer all heading north-east at the same time are formalised as $P = I(45, 45, 45, 45)$ or $P = I([45]\{4\})$ (figure 4b). A motion azimuth change of one deer switching from east (90°) to north (0°) within three time steps would be formalised as $P = C([90]\{-90\}\{3\})$ (figure 4c).

(Insert figure 4 about here.)

4.3.2 Quantifiers

Much like the single characters in strings do in regex, REMO pattern elements can have quantifiers in the REMO formalism (table 1). In our case the quantifiers are used to describe the pattern extent in the temporal and the objects axis. The pattern *length* (time axis) can be stated as *fixed* $\{l\}$ (exactly l time steps), *open* $\{l, \}$ (at least l time steps) or as a *range* $\{l, k\}$ (between l and k time steps). The pattern *width* (object axis) is formalised in an analogous way: *fixed* $\{n\}$ (exactly n individuals), *open* $\{n, \}$ (at least n individuals) or as a *range* $\{n, m\}$ (between n and m individuals). Thus, a constance pattern can either have a fixed, open or range length.

A question mark (?) makes the preceding pattern element optional. A star ($*$) refers to a pattern element that may appear never or many times. A plus sign ($+$) describes pattern elements that appear once or many times.

(Insert table 1 about here.)

Examples Analysing lifelines of deer one might look for individuals moving constantly towards north-east for a certain time period. Depending on the exact scientific question this could be stated with a fixed pattern $P = S([45]\{3\})$ (case e in figure 4), with an open pattern $P = S([45]\{2, \})$ (cases a, d, and e), or with an range pattern $P = S([45]\{2, 3\})$ (cases d and e only).

As a further example one might want to identify individuals heading first north-east (45°) and subsequently turning through east (90°) to southeast (135°). Quantifiers allow every single step of this motion to be stated more precisely. The expression $P = S(45, 90?, 135)$ makes the eastward motion step optional and matches cases d and e in figure 4. With $P = S(45, 90*, 135)$ an arbitrary series (including 0 times) of eastward motion steps can be interposed, again d and e match. $P = S(45, 90+, 135)$ requires at least one eastward motion step, which excludes d from the matches.

4.3.3 Pattern depth descriptors

Since the pattern elements are numbers and not characters or numerals, the REMO formalism has its own pattern depth descriptors allowing to express value ranges and relational operators (see table 2). Exact motion attribute values are expressed in brackets $[v]$. The pattern depth descriptor can take the common relational operators less than ($<$), greater than ($>$), less than or equal to ($\leq$), greater than or equal to ($\geq$). Ranges are expressed with a hyphen ($-$), exclusion with a caret ($\wedge$).

The OR operator allows specification of a choice of pattern elements, using a vertical bar (|). Since a pattern element is unique, the AND operator (&) is omitted.

(Insert table 2 about here.)

Examples One possible use of pattern depth descriptors is to find an individual deer that first moves exactly north-east and then moves on in a direction between north-east and south-east. This pattern is formalized as $P = S(45, [45-135])$. One match amongst many others in figure 4 is instance f.

4.3.4 Unbound Patterns

So far all the patterns consisted of clearly specified motion attribute observations A_m. However, some users might not know in advance which exact values the REMO patterns in their data will have. There might, for instance, lurk constance patterns of u as well as v or w in a data set. In this case the constance pattern has a shape but no defined content, the pattern is *unbound*. A *bound* pattern in constrast has defined motion attribute observations. The REMO formalism features the wildcard [♯] to express unbound patterns. Pattern P in 8 matches any sequence of l consecutive equal motion attribute observations. Pattern P in 9 matches any set of n concurrent motion attribute observations.

$$P = S([\sharp]\{l\}) \tag{8}$$
$$P = I([\sharp]\{n\}) \tag{9}$$

Examples $P = I([\sharp]\{4\})$ matches any set of 4 deer concurrently moving in the same direction. Two possible instances of P are highlighted as b and h in figure 4. As will be seen in the implementation section, unbound pattern matching has been chosen to keep the prototype as generic as possible. Diverse filtering procedures turn unbound pattern matching back into bound pattern matching.

4.3.5 Complex pattens over time and across objects

The REMO formalism syntax presented so far is closely related to basic regex. The major differences come with the need to express two-dimensional patterns on a two-dimensional analysis matrix. The solution chosen for the REMO formalism follows an intuitive and simple approach. Since complex patterns are defined as a composite of simple patterns, i.e. sequences and incidents, it is straightforward to reproduce this construction principle in the formalism. A complex pattern is formalised as a set of simple patterns that are temporally linked (10).

$$complexPattern_i = \begin{cases} simplePattern_1 & : & interval_1 \\ simplePattern_2 & : & interval_2 \\ simplePattern_3 & : & interval_3 \\ ... & : & ... \\ simplePattern_n & : & interval_n \end{cases} \tag{10}$$

The term $interval_i$ allows indication of the relative temporal order of the linked simple patterns. Although in many cases a complex pattern will either have a common shared start time t_b or end time t_e this is not required to build a complex pattern.

The following set shows three trend-setters with different lengths $\{l\}$, $\{l,\}$, and $\{l, k\}$ respectively (11, 12, and 13).

$$P = \begin{cases} S([v]\{l\}) & : & t_{e-l+1}, \ldots, t_e \\ I([v]\{n\}) & : & t_e \end{cases} \tag{11}$$

$$P = \begin{cases} S([v]\{l,\}) & : & t_{e-j}, \cdots, t_e \mid (l-1) \leq j \leq (e-1) \\ I([v]\{n\}) & : & t_e \end{cases} \tag{12}$$

7

$$P = \begin{cases} S([v]\{l, k\}) & : & t_{e-j}, \cdots, t_e \mid (l-1) \le j \le (k-1) \\ I([v]\{n\}) & : & t_e \end{cases} \qquad (13)$$

Quantifiers can be used not only to describe sets of single REMO matrix cells but also to compose complex patterns. For example the contemporary occurrence of n identical constance patterns of length l could be viewed as a concurrence of n constance patterns and thus be expressed as a nested term (14).

$$P = I(S([v]\{l\})\{n\}) \qquad (14)$$

Examples Investigating group dynamics in a herd of deer one might search for an alpha individual initiating a travel in a north-east motion before all other members of the herd. Such a trend-setter pattern P is shown in figure 4i. Deer O_6 anticipates at time t_4 three time steps in advance the motion of the deer O_7, O_8, and O_{10} (15).

$$P = \begin{cases} S([45]\{3\}) & : & t_{e-2}, \ldots, t_e \\ I([45]\{4\}) & : & t_e \end{cases} \qquad (15)$$

A slight modification of the trend-setter pattern illustrates the potential of the REMO formalism. The pattern of interest shall be called *infection* (figure 4j): A deer starts to move along the motion azimuth $45°$ at t_b. After two time steps two other deer join this emerging group and show the same motion. Another two deer join one time step later. A set of three deer are joining the group at t_{b+5} building an incident of width 8 (16).

$$P = \begin{cases} S([45]\{6\}) & : & t_b, \cdots, t_{b+5} \\ I(S([45]\{4\})\{2\}) & : & t_{b+2}, \cdots, t_{b+5} \\ I(S([45]\{2\})\{2\}) & : & t_{b+4}, \cdots, t_{b+5} \\ I([45]\{8\}) & : & t_{b+5} \end{cases} \qquad (16)$$

The REMO formalism offers a simple and comprehensible way to describe arbitrary REMO patterns. The formalism is used to describe the pattern matching process in the following sections.

5   Implementation

As a proof of concept the REMO analysis appraoch was implemented as an application prototype in Java. Three reasons argued for a stand-alone solution instead of customizing an out-of-the-box GIS. First, there was the need for an object-oriented spatio-temporal data model for MPOs, not present in most of today's commercial GIS. Second, high demands on integrating the pattern matching process with visualisation required an open and flexible environment. Third, only a few traditional GIS functions were required.

To maximise the intuitiveness of use for potential users (e.g. wildlife biologists or social scientists) we refrained from developing an interpreter for the REMO formalism. In contrast we decided to develop an application prototype featuring an easy to use graphical user interface (GUI). Thus, the REMO prototype features modules to manage and preprocess the data, to animate the MPOs in a space-time viewer (figure 8), to control the pattern matching process (figure 5) and to visualise results.

(Insert figure 5 about here.)

5.1   Analysis process

The example of matching the pattern $P = S([\sharp]\{4, \})$ illustrates the exploratory data analysis process (figure 6). Upon user request the `mainController` first calls `create` to build an `MPOGroup`. The `MPOGroup` calls subsequently `create` to construct a `matrix` based on the lifelines of the MPOs. The `matrix` is built according to the properties set up by the `mainController`: the dimensions, the temporal granularity and the motion attribute of interest and its reclassification (`setParameter()`).

Under the control of the users the `mainController` calls `create` to set up (`setParameter()`) a `pattern` to `search` on the `matrix`. The `pattern matching` process locates instances of the searched pattern on the `matrix`. A `matchList` gathers the resulting matches (`create`) for later visualisation (`highlight()`) and further investigations. This process can be repeated with varied matrix and pattern properties.

(Insert figure 6 about here.)

Note that the pattern matching itself is two-tiered. First unbound patterns like $P = S([\sharp]\{4, \})$ are constructed (figure 5) and matched on the matrix. Only in a downstream filtering process users can select patterns of specified values from the `matchList`. This step reduces in this example the matches of $P = S([\sharp]\{4, \})$ to the probably fewer matches of $P = S([45]\{4, \})$.

5.2 Pattern matching algorithms

Matching REMO patterns on the REMO matrix has a certain similarity to the classical pattern matching problem on strings. Given a text string T of length n and a pattern string P of length m, one wants to know first whether P is a substring of T and second where on T this match is located. Decomposing a REMO matrix into its rows (motion attribute arrays) and columns (timeslices) allows use of derivatives of classical string pattern matching algorithms like *Brute-Force Pattern Matching* (BFPM) or *Knuth-Morris-Pratt* (KMP) (Knuth *et al.* 1977). While pattern matching on strings compares characters, REMO pattern matching has to compare numbers which can be associated with relational operators. However the basic principles remain the same. Complex patterns are matched by first matching constitutive simple patterns in either the temporal or object dimension and then testing whether the additional conditions over both dimensions are also met.

Brute-force pattern matching simply tests all the possible placements of P relative to T in the worst case in $O(nm)$ running time (Goodrich and Tamassia 1998). KMP uses a failure function f for the pattern string P which encodes repeated substrings inside the pattern itself to reuse previous comparisons and thus avoid unnecessary comparisons. It achieves in the worst case a running time of $O(n+m)$ (Goodrich and Tamassia 1998). The performance of application prototype did not show worst case behavior and thus did not encounter significant performance problems with the above mentioned algorithms and with the test data used so far. With larger data volumes there may a requirement to develop and use more sophisticated pattern matching algorithms.

6 Use cases

The REMO analysis concept and its implementation prototype have been tested with various data ranging from tracked individuals to moving data points in abstract spaces. Two different use cases illustrate the concept's generic applicability: Football (soccer) players tracked on the pitch and data points in an abstract ideological space.

6.1 Football players

From a non-scientific perspective the motion of football players is probably the most intensively observed and most competently discussed motion of individuals ever created by human culture. However, from a GIScientist's perspective a team of football players is a group of MPOs acting in a structured way on a well-defined space and over a well-defined time period. Thus football players are an ideal use case to evaluate the REMO analysis concept. The motion of a team of football players is highly coordinated. Luring the other team into the off-side trap, for example, requires coordinated motion of all four defenders in a row. Another example might be if a creative striker governs the motion of the row of defenders at the back.

(Insert figure 7 about here.)

The data used emerged from research pursued with the aim of tracking football players using multiple television cameras (Iwase and Saito 2002, 2003). The time frame covers about 33 seconds

of a football game in a Japanese University league, tracking 11 players with a sampling rate of 15 fixes/sec (figure 7). Note the confusing tangle of lifelines produced by such a small group of eleven individuals and such a short time frame of half a minute.

For simplicity and illustration purposes we focus on the motion azimuth in this example. The REMO matrix in figure 7 is resampled at a temporal interval of 1 second to simplify the matrix and to eliminate short term positional noise. Since the pitch's orientation in this example is left to right, the team is attacking to the east (azimuth 90°) and defending to the west (azimuth 270°).

(Insert figure 8 about here.)

First we search for *constance*. Constance with regard to tracked football players describes players running in one direction for some period. This pattern can especially be expected for the left and right wingers as well as for the strikers. The former repeatedly sprint along the side-lines, the latter striking from the midfield in the direction of the opponent's goal. As a first simple example we search for the longest straight tracks in this data sample. Figure 8 illustrates the matches for a constance of at least length 10 seconds ($P = S([270]\{10, \})$). The three players No. 5, 14, and 21 showing this long straight sprint are indeed offensive players with the longest way back and thus forced to take the straight line.

As a second pattern we try to identify a *concurrence*. Common sense and a first inspection of figure 7 suggest that concurrent motion is often seen in the lifelines of a team of football players. For example the maintenance of an effective off-side trap demands highly coordinated motion from the defenders. If we are interested in the degree of coordination in the players' motion we can search for concurrence incidents with as many participants as possible. In the present example a concurrence consisting of ten out of eleven players can be found. The event shown in figure 9 illustrates the reaction of almost the whole team to an attack at time t_{15}; ten MPOs move backwards, with an azimuth of 270°, respectively formalised as ($P = I([270]\{10, \})$).

(Insert figure 9 about here.)

From a football coach's perspective *trend-setters* might be of special interest. A trendsetting football player might anticipate the important moves in the game. For this illustration of the concept trend-setter we focus on the coordinated defending of the team around t_{14}. The general exploratory task is to find trend-setters anticipating this move. In the example in figure 10 we identify a trend-setter of at least length 4 seconds and with at least 8 team mates joining in the backwards move of the leading individual.

$$P = \begin{cases} S([270]\{4, \}) & : & t_{e-3}, \dots, t_e \\ I([270]\{8, \}) & : & t_e \end{cases} \tag{17}$$

(Insert figure 10 about here.)

Football scene analysis using multiple TV cameras has a huge potential for strategy understanding and making digest TV programs. Investigating the emerging vast amount of lifeline data from sports applications requires spatio-temporal data mining methods. We identify this field as an opening opportunity to bring in the knowledge and tradition of GIScientists analysing spatio-temporal data.

6.2 Abstract data points

The frequently held popular referendums in Switzerland (approx. 8-10 per year) allow researchers to make detailed inferences about value conflicts within the society. Hermann and Leuthold (2001) developed an inductive approach to discover the basic ideological conflicts in Switzerland. Performing factor analysis on referendum data at the district level of all 158 federal referendums held between 1981 and 1999, they hypothesised a structure of mentality, which was interpreted as being composed of three dimensions: political left vs. political right, liberal vs. conservative and ecological vs. technocratic. The axes of this multidimensional ideological space, taken in pairs, provide a total of three two-dimensional maps of the political landscape of Switzerland (Hermann and Leuthold 2003). In these two dimensional ideological spaces the 185 districts can be localized in intervals of

one year, from 1981 until 1999. Irrespective of their political and social meaning, the districts can be considered as moving points in a two dimensional space (figure 11).

(Insert figure 11 about here.)

Figure 12a shows the REMO matrix for the motion azimuth of the districts. The districts of a Canton (member states of the Swiss Federation) are grouped together in the matrix (figure 12a). Thus, proximity in the matrix corresponds to a certain institutional and cultural similarity.

To show the generic potential of the REMO analysis concept we investigate the old argument as to whether the German and the Latin part (French, Italian and Rhaeto-romanic) of Switzerland are politically diverging, opening the so-called 'Röschtigraben' ('Röschti' being a characteristic Swiss-German food, and 'Graben' being a ridge; literally 'röschti-ridge', the concept used to describe the cultural divide between the two parts of country). Since we are not political scientists it is not our intension to derive any statistically significant political conclusions. In fact we merely want to demonstrate the exploratory potential of our approach, identifying segregation tendencies to motivate further political research on the topic. Such tendencies can be extracted with the REMO concept, detecting for constance patterns to show that there really is a notable sum of districts moving in a specified direction. For this illustration we searched for constance patterns with a length of 7 years, a third of the entire period of 20 years.

(Insert figure 12 about here.)

Figure 12b reveals in the upper two thirds 45 districts showing $P = S([90]\{7,\})$ standing for a rightwards motion of the German speaking districts in the 1990s (36% of all German speaking districts). The lower third of the plot in figure 12c shows 18 constances of $P = S([270]\{7,\})$ illustrating the opposite motion to the left by the francophone districts (39% of the francophone districts). As an interesting feature two German speaking districts could be identified that show a fairly isolated constant leftward drift in the 1980s (in the centre of figure 12c, Sargans SG and Stein SH).

Having found so many constances of length 7 in only one decade points to the assumption that there must be concurrences, representing the divergence of the German and French speaking districts. Those concurrences are expected to spread over a width corresponding approximately to the sum of the constances found above. Indeed, figure 12d shows 6 subsequent concurrences $P = I([90]\{45,\})$ mainly in the German speaking districts and figure 12e a striking over-representation of French speaking districts involved in one of the 9 occurrences of $P = I([270]\{18,\})$.

Finally the question arises as to whether the REMO analysis concept could identify the districts anticipating the left-right divergence. Figure 12f and g give first hints on this issue. It shows all matches to the following trend-setter pattern, again mapping approximately the dimensions of the constances and concurrences found.

$$P = \begin{cases} S([90]\{7,\}) & : & t_0, \ldots, t_6 \\ I([90]\{45,\}) & : & t_6 \end{cases} \tag{18}$$

This trend-setter pattern can be matched a total of 72 times. The real trend-setting districts are those who anticipate the rightward drift the earliest. To identify them the trend-setter pattern is expanded to:

$$P = \begin{cases} S([90]\{12,\}) & : & t_0, \ldots, t_{11} \\ I([90]\{45,\}) & : & t_{11} \end{cases} \tag{19}$$

Figure 12f shows the result. Three districts can very clearly be identified as the trend-setters (Trachselwald BE, Zofingen AG and Gösgen SO). The first two districts anticipate more than 10 years in advance a motion attribute that at least 45 followers later adopt. To identify trend-setters in the leftward drift of the French speaking districts the following pattern has been matched:

$$P = \begin{cases} S([270]\{7,\}) & : & t_0, \ldots, t_6 \\ I([270]\{18,\}) & : & t_6 \end{cases} \tag{20}$$

Figure 12g helps to identify the first two districts heading left (Conthey VS and Entremont VS).

The spatialisation of statistical data by e.g. plotting the annual shift of political entities in a ideological space opens up a chance for motion analysis. GISc may contribute with its analysis and visualisation potential. Both use cases underscore the need for designing generic methods for spatio-temporal knowledge discovery, as proposed with the REMO analysis concept.

7 Discussion

This discussion first evaluates our approach with respect to other approaches providing analysis tools for spatio-temporal data and their potential to analyse motion data. Second we discuss some open problems of modelling MPOs relevant to our approach. Third we conclude with an outlook.

7.1 Evaluation

Descriptive statistics. Common descriptive statistics applied to all available data of a single individual or of a group are unsuitable to investigate motion. Collapsing the data into a set of descriptive measures makes it impossible to detect inter-object relations and spatially or temporally delimited motion patterns. In contrast, the REMO analysis concept allows to investigate the data of many individuals concurrently and thus allows detection of short- and long-term as well as inter-object relationships.

Databases. In the database management systems (DBMS) community numerous approaches exist to extend common query languages to cover the special properties of spatio-temporal data, e.g. SQL/temporal (TQuel, TSQL2) or future SQL (FTL) (e.g. Abraham and Roddick 1999, Snodgrass 1987, Snodgrass and Kucera 1995, Sistla *et al.* 1998). These queries may even involve data about moving entities as well as moving query windows (Raptopoulou *et al.* 2003). Still, the basic task is to retrieve stored objects, collections of objects or their observations from a database according to a query. The REMO analysis concept in contrast focuses on objects representing motion patterns that are not stored *per se* in a database.

Dynamic cartography and exploratory spatial data analysis (ESDA). Another set of analysis tools for tracking data focuses on the visualisation of spatio-temporal data and on analysis based on inter-activity (e.g. Andrienko and Andrienko 1999, MacEachren *et al.* 1999, Edsall *et al.* 2000). ESRI Inc., for instance, offers a tool to visualise and analyse tracking data, the ArcGIS Tracking Analyst extension. It features various symbology options and a sophisticated playback manager. Its exploratory power lies, however, in the functionality to define events and to visualise where and when they occur. Thus, analysis is performed in a exploratory way, depending predominantly on the user's knowledge of the data and their sensitivity to conspicuous features. Most of the ESDA approaches share the *data projection and reduction* step with the REMO analysis concept. However, while dynamic cartography and ESDA depend on an alert user to find eye-catching patterns or trends viewing the data from varying perspectives, the REMO analysis concept provides a quantitative approach. Instead of qualitative exploration it offers the formalisation of expected patterns and their detection in a quantitative and automated, and above all, in an objective and repeatable way. In short, it adopts the syntactic pattern detection approach.

KDD and data mining. KDD and its step data mining in spato-temporal data are mainly focused on cluster detection in changing point distributions. Besides the seminal work of Openshaw (Openshaw 1994, Openshaw *et al.* 1999) on long-term disease data further research was also carried out. For instance, Sadahiro worked on various urban point patterns (Sadahiro 2002). Whereas Openshaw's approaches follow the 'search everywhere for the unusual' philosophy, Sadahiro tracks the local maxima of density surfaces over time to reveal the displacement of disease clusters. However there is a fundamental difference between these approaches and the REMO analysis concept: In the former case the points represent point occurences of, for instance, disease cases, that is point distributions without trackable individuals; in the latter case the points represent successive fixes of trackable individuals. The latter information is of potantially higher information content and of-

fers completely different analysis approaches, e.g. the establishment of spatio-temporal inter-object relationships shown in the REMO analysis concept.

With the REMO analysis concept we propose a comprehensive procedure for spatio-temporal geographic knowledge discovery. Our approach incorporates almost the complete list of KDD functions given by Miller and Han (2001): Data selection, pre-processing, data cleaning, data reduction and projection (i.e. construction of REMO matrix), segmentation (i.e. classification of constance patterns), deviation and outlier detection (i.e. finding trend-setters), trend detection (i.e. finding concurrences), generalisation, characterisation and visualisation.

Patterns are non-random properties and relationships that are valid, novel, useful and ultimately understandable (Fayyad *et al.* 1996). *Valid* means the pattern has to be general enough to apply to new data. The REMO patterns have been tested with such diverse data as GPS-tracked animals, football players on a pitch and moving data points in an abstract ideological space. A *novel* pattern is non-trivial and unexpected. A trend-setter anticipating 12 time steps before 45 followers join a motion feature is without doubt both non-trivial and unexpected. To be *useful* a pattern has to lead to some benefit to the user. Even though a list of REMO pattern matches neither explains the complex behaviour of wild animals nor the sometimes unpredictable paths of football players, nor the complex processes of Swiss society, it provides useful initial insights and gives hints for further scientific investigations. Finally a pattern should be *ultimately understandable*, that is, simple and interpretable by humans. Using obvious and intuitive patterns, a simple analysis matrix and an extension of an established formalism, the REMO analysis concept is easy to understand for its potential users.

7.2 Open problems

The integration of the MPO MODELLING DOMAIN and the ANALYSIS DOMAIN revealed a set of generic problems of describing the motion of tracked individuals. Since similar problems arise in any implementation of MPO models, we discuss them here in detail.

Uncertain and missing fixes. Dealing with real tracking data one is often faced with uncertain or missing fixes. For the detection of relative motion patterns intervals with uncertain or missing MPO tracks can be fatal. Preliminary work has been carried out by Pfoser and Jensen (1999) for dealing with uncertain fixes and by Wentz *et al.* (2003) for filling fragmentary tracks. Within the REMO approach the use of detached attribute functions to construct a derivative analysis matrix at arbitrary granularities allows us to circumvent this problem. Nevertheless, real geospatial lifeline data will always be incomplete and error-prone. Thus, dealing with imperfect lifeline data remains an open research problem.

Interpolation issues. Just as in the spatial case (see O'Sullivan and Unwin 2003) radius-limited or nearest neighbour interpolation of motion attributes on lifeline data have two points in common. First, the size of the moving window is arbitrary. Even though users will select the interval of the moving window according the characteristics of the data, different interval lengths may change the results of the data mining process. Secondly, the wider the moving window is chosen, the smoother the lifeline description becomes. In the case of tracking data with a very high sampling rate this may be desired to eliminate the effects of inaccurate fixes. In many other cases excessive smoothing undesirably blurs the crucial rough edges of lifelines.

Sampling frequency. On the one hand the interpolation of high frequency time steps between distant fixes may lead to the problem of granularity as described by (Duckham *et al.* 2003,pg. 79): 'Imprecision leads to granularity, where individual elements within a particular grain cannot be discerned apart'. Adapted to the problem of sampling motion attributes from lifelines, we may sample consecutive time steps that cannot be discerned and show the same parameter values. Dense resampling of lifelines with sparse fixes may create artefacts, e.g. false constance patterns filling the gaps between temporally distant fixes. To avoid semantic mismatches, the tracking data should have the same granularity as the analysis task (Cheylan 2001, Hornsby 2001). Thus, the temporal

sampling interval of the sequences should be chosen at the granularity given by the fix sampling rate. On the other hand the Nyquist-Shannon sampling theorem (Nyquist 1928, Shannon 1949) applies if every single fix must be sampled. Following the theorem the sampling frequency for computing motion attributes must be greater than twice the shortest interval between two fixes. Since the detached attribute functions use interpolation, the Nyquist-Shannon theorem, however, cannot be applied in the REMO approach. Nevertheless, integrating the granularity problem and the sampling theorem a rule of thumb can be stated. The motion attribute sampling rate should have have a similar granularity as the original observation data. In short, the granularity to the input data limits the analysis.

Aggregation. Another crucial issue arises similar to the modifiable areal unit problem (MAUP) (Openshaw 1984). Describing lifelines we consider different temporal aggregations instead of spatial aggregations as with the classical MAUP. The potentially arbitrary aggregation comes with the mapping of the irregular fixes on to the regular REMO matrix, determined by a stepsize and a starting time. If the temporal units were specified differently, we might observe very different patterns and relationships. The effects of different aggregation schemes on lifeline data will therefore have to be studied in detail.

Classification. Yet another granularity effect shows the classification of the motion attributes. The number of matched patterns is highly dependent on the attribute granularity of the pattern matching process. For example the classification of motion azimuth into only the two classes *east* and *west* reveals a lot of presumably meaningless constance patterns. In contrast, every constance pattern found with 360 azimuth classes tends to be highly meaningful. The same is true for the classification mode. Linear, standard deviation or quantile approaches all produce different classification results and thus different patterns and relationships.

7.3   Outlook

In future work we intend to test the REMO approach with an extended set of real use case and synthetic data. In addition to the football players and socio-political data described in the use case section we will include data of GPS-tracked ungulates and sharks. In order to have data of arbitrary temporal and spatial granularity we will furthermore investigate synthetic tracking data emerging from agent based MPO models. We will therefore configure MPO-agents with a set of behavioural constraints. To evaluate the REMO analysis concept we will investigate how accurately we are able to identify the before specified behavioural constraints in emerged patterns in the lifelines.

For a better understanding of the concept's reliability and sensitivity to different spatial and temporal granularities, lifeline interpolation techniques and motion attribute classifications we will use Monte Carlo Simulations.

We intend to address in detail the problem of interpolating and aggregating lifeline data. We will therefore integrate additional interpolation and aggregation techniques in the detached attribute functions and investigate their performance with our use case data. This includes smoothing the lifelines in a preprocessing step using curve fitting approaches such as linear regression, polynomial regression, or splines.

We will address the establishment of a statistical background for our approach to discern pattern from noise. This includes the introduction of statistical measures of interestingness, expressiveness and reliance for REMO patterns (Silberschatz and Tuzhilin 1996, 1995). We will, for instance, use one-dimensional autocorrelation measures to provide a more objective basis for deciding whether or not the found REMO patterns are random, and if the patterns are not random, to quantify how unusual they are.

The REMO analysis approach described here concentrates on relative motions of objects irrespective of their proximity in absolute space. However, behavioural patterns based on absolute positions like *proximity* as well as *divergence* or *convergence* are crucial for the understanding of many motion processes. Thus, the incorporation of spatial relations using not only relative but also absolute spa-

tial measures of the fixes will greatly expand the pattern detection capabilities of the concept. Right now we are integrating a set of 'absolute space' REMO patterns (Laube *et al.* 2004).

8   Conclusions

The main contribution of this paper is the development of a generic approach for spatio-temporal knowledge discovery in geospatial lifeline data. The approach is built on the integration of the following KDD steps:

- *Data reduction and projection:* The transformation of the lifeline data to an analysis matrix featuring motion attributes (i.e. speed, change of speed, motion azimuth) allows comparison of the object's motion across objects and over time.
- *Exploratory analysis and model selection:* The REMO formalism allows description of patterns of relative motion in geospatial lifeline data.
- *Data mining:* Pattern detection algorithms facilitate the automatic search for relative motion patterns within large data sets.
- *Visualization:* Interactive linking of the mined patterns with the object's motion in a MPO-Viewer allows interpretation of the mined patterns in order to suggest avenues for further research.

To demonstrate the generic nature of the approach we used football players tracked on a pitch and data points moving in an abstract ideological space, namely the Swiss districts displacing in an ideological space over 20 years. We demonstrated that our proposed methodology is able to extract a set of valid, novel, useful and understandable motion patterns within these datasets

We showed that REMO pattern detection reveals many more motion patterns than are easily seen through inspection and do so in a more formal and hence repeatable way. Thus the presented methodology may increase our knowledge about various processes in dynamic geospatial data. More generally, pattern matching is an interesting and promising analysis technique for the current increase of spatio-temporal data and the developimg need for their analysis.

9   Acknowledgements

References

ABRAHAM, T., and RODDICK, J. F., 1999, Survey of spatio-temporal databases. *GeoInformatica*, **3**, 61–99.

ANDRIENKO, N. V., and ANDRIENKO, G. L., 1999, Interactive maps for visual data exploration. *Int. J. Geographical Information Systems*, **13**, 355–374.

BATTY, M., DESYLLAS, J., and DUXBURY, E., 2003, The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades. *Int. J. Geographical Information Systems*, **17**, 673–697.

BETTINI, C., JAJODIA, S., and WANG, S., 2000, *Time Granularities in Databases, Data Mining, and Temporal Reasoning*, first edn. (Berlin: Springer).

CHEYLAN, J.-P., 2001, Time and spatial database, a conceptual application framework, In A. Frank, J. Raper, and J. Cheylan (eds), *GISDATA8, Life and Motion of Socio-economic Units*, pp. 35–48 (London: Taylor and Francis).

DUCKHAM, M., KULIK, L., and WORBOYS, M., 2003, Imprecise navigation. *GeoInformatica*, **7**, 79–94.

EDSALL, R. M., HARROWER, M., and MENNIS, J., 2000, Tools for visualizing properties of spatial and temporal periodicity in geographic data. *Computers and Geosciences*, **26**, 109–118.

FAYYAD, U., PIATETSKY-SHAPIRO, G., and SMYTH, P., 1996, From data mining to knowledge discovery in databases. *AI Magazine*, **17**, 37–54.

FRIEDL, J. E. F., 2002, *Mastering Regular Expression*, second edn. (Beijing: O'Reilly).

GANSKOPP, D., 2001, Manipulating cattle distribution with salt and water in large arid-land pastures: a GPS/GIS assessment. *Applied Animal Behaviour Science*, **73**, 251–262.

GOODRICH, M. T., and TAMASSIA, R., 1998, *Data Structures and Algorithms in Java* (New York: Wiley).

HERMANN, M., and LEUTHOLD, H., 2001, Weltanschauungen und ihre soziale Basis im Spiegel der Volksabstimmungen. *Swiss Political Science Review*, **7**, 39–63.

HERMANN, M., and LEUTHOLD, H., 2003, *Atlas der politischen Landschaften. Ein weltanschauliches Portrait der Schweiz* (Zurich: vdf-Verlag).

HORNSBY, K., 2001, Temporal zooming. *Transactions in GIS*, **5**, 255–272.

HORNSBY, K., and EGENHOFER, M. J., 2002, Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, **36**, 177–194.

IWASE, S., and SAITO, H., 2002, Tracking soccer player using multiple views, In *IAPR Workshop on Machine Vision Applications* (MVA Proceedings), pp. 102–105.

IWASE, S., and SAITO, H., 2003, Tracking soccer players based on homography among multiple views, In T. Ebrahimi, and T. Sikora (eds), *Visual Communications and Image Processing* (Proceedings of SPIE), pp. 288–292.

JAIN, A. K., DUIN, R. P. W., and MAO, J., 2000, Statistical pattern recognition: A review. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, **22**, 4–37.

KEMP, Z., and KOWLACZYK, A., 1994, Incorporating the temporal dimension in a GIS, In M. F. Worboys (ed.), *Innovations in GIS 1*, pp. 89–103 (London: Taylor and Francis).

KNUTH, D. E., MORRIS, J. H., and PRATT, V. R., 1977, Fast pattern matching in strings. *SIAM Journal on Computing*, **6**, 323–350.

LAUBE, P., and IMFELD, S., 2002, Analyzing relative motion within groups of trackable moving point objects, In M. J. Egenhofer, and D. M. Mark (eds), *Geographic Information Science, Second International Conference, GIScience 2002, Boulder, CO, USA, September 2002*, vol. 2478 of *Lecture Notes in Computer Science*, pp. 132–144 (Berlin-Heidelberg: Springer).

LAUBE, P., VAN KREVELD, M., and IMFELD, S., 2004, Finding REMO - detecting relative motion patterns in geospatial lifelines, In P. F. Fisher (ed.), *Developments in Spatial Data Handling*, Proceedings of the 11th International Symposium on Spatial Data Handling, pp. 201–214 (Berlin-Heidelberg: Springer).

MACEACHREN, A. M., WACHOWICZ, M., EDSALL, R., and HAUG, D., 1999, Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods. *Int. J. Geographical Information Systems*, **13**, 311–334.

MARK, D. M., 1998, Geospatial lifelines, In *Integrating Spatial and Temporal Databases*, 98471 (Dagstuhl Seminars).

MARK, D. M., 2003, Geographic information science: Defining the field, In M. Duckham, M. F. Goodchild, and M. F. Worboys (eds), *Foundations of Geographic Information Science*, chap. 1, pp. 3–18 (London: Taylor and Francis).

MILLER, H. J., 1991, Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information Systems*, **5**, 287–301.

MILLER, H. J., 2003, What about people in geographic information science? *Computers, Environment and Urban Systems*, **27**, 447–453.

MILLER, H. J., and HAN, J., 2001, Geographic data mining and knowledge discovery: An overview, In H. J. Miller, and J. Han (eds), *Geographic data mining and knowledge discovery*, pp. 3–32 (London: Taylor and Francis).

MILLER, H. J., and WU, Y.-H., 2000, GIS software for measuring space-time accessibility in transportation planning and analysis. *GeoInformatica*, **4**, 141–159.

MOUNTAIN, D., and RAPER, J., 2001, Modelling human spatio-temporal behaviour: A challenge for location-based services, In *Proceedings of the 6th International Conference of GeoComputation*.

NYQUIST, N., 1928, Certain topics in telegraph transmissio theory. *Trans. AIEE*, **47**, 617–644.

OPENSHAW, S., 1984, *The Modifiable Areal Unit Problem* (Norwich, UK: Geo Books).

OPENSHAW, S., 1994, Two exploratory space-time-attribute pattern analysers relevant to GIS, In S. Fotheringham, and P. Gogerson (eds), *GIS and Spatial Analysis*, chap. 5, pp. 83–104 (London: Taylor and Francis).

OPENSHAW, S., TURTON, I., and MAGILL, J., 1999, Using geographic analysis machine to analyze limiting long-term illness cencus data. *Geographical and Environmental Modelling*, **3**, 83–99.

O'SULLIVAN, D., and UNWIN, D. J., 2003, *Geographic Information Analysis*, first edn. (Hoboken, NJ: Wiley).

PFOSER, D., and JENSEN, C. S., 1999, Capturing the uncertainty of moving-object representations, In R. H. Gueting, D. Papadias, and F. Lochowsky (eds), *Advances in Spatial Databases, 6th International Symposium, SSD'99, Hong Kong, China, July 1999. LNCS 1651*, pp. 111–131 (Berlin-Heidelberg: Springer).

RAPER, J., and LIVINGSTONE, D., 1995, Development of a geomorphological spatial model using object-oriented design. *Int. J. Geographical Information Systems*, **9**, 359–383.

RAPTOPOULOU, K., PAPADOPOULOS, A. N., and MANOLOPOULOS, Y., 2003, Fast nearest-neighbor query processing in moving-point databases. *GeoInformatica*, **7**, 113–137.

RODDICK, J. F., HORNSBY, K., and SPILIOPOULOU, M., 2001, An updated bibliography of temporal, spatial, and spatio-temporal data mining research, In J. F. Roddick, and K. Hornsby (eds), *Temporal, spatial and spatio-temporal data mining, TSDM 2000, LNAI 2007*, pp. 147–163 (Berlin: Springer).

SADAHIRO, Y., 2002, A graphical method for exploring spatiotemporal point distributions. *Cartography and Geographic Information Science*, **29**, 67–84.

SHANNON, C., 1949, Communication in the presence of noise. *Proc. Institute of Radio Engineers*, **37**, 10–21.

SIBBALD, A. M., HOOPER, R. J., GORDON, I. J., and CUMMING, S., 2001, Using GPS to study the effect of human disturbance on the behaviour of the red deer stags on a highland estate in Scotland, In A. M. Sibbald, and I. J. Gordon (eds), *Tracking Animals with GPS* (Macaulay Insitute), pp. 39–43.

SILBERSCHATZ, A., and TUZHILIN, A., 1995, On subjective measures of interestingness in knowledge discovery., In *Proc. of the 1st Int. Conf. On Knowledge Discovery and Data Mining, Montreal*, pp. 275–281.

SILBERSCHATZ, A., and TUZHILIN, A., 1996, What makes patterns interesting in knowledge discovery systems. *IEEE Transactions on Knowledge and Data Engineering*, **8**, 970–974.

SISTLA, A. P., WOLFSON, O., CHAMBERLAIN, S., and DAO, S., 1998, Querying the uncertain position of moving objects, In O. Etzion, S. Jajodia, and S. Sripada (eds), *Temporal Databases - Research and Practice*, vol. 1399 of *Lecture Notes in Computer Science*, pp. 310–337 (Berlin: Springer).

SMYTH, C. S., 2001, Mining mobile trajectories, In H. J. Miller, and J. Han (eds), *Geographic data mining and knowledge discovery*, pp. 337–361 (London: Taylor and Francis).

SNODGRASS, R. T., 1987, The temporal query language TQuel. *ACM Trans. on Database Systems*, **12**, 247–298.

Snodgrass, R. T., and Kucera, H., 1995, Rationale for a temporal extension of SQL, In *The TSQL2 Temporal Query Language*, pp. 3–18 (Dordrecht, the Netherlands: Kluwer).

Wall, L., Christiansen, T., and Schwartz, R. L., 1996, *Programming Perl*, second edn. (Sebastopol, CA: O'Reilly).

Wentz, E. A., Campell, A. F., and Houston, R., 2003, A comparision of two methods to create tracks of moving objects: linear weighted distance and constrained random walk. *Int. J. Geographical Information Systems*, **17**, 623–645.

Legend to figure 1   The geospatial lifelines of four MPOs (a) are used to derive at regular intervals the motion azimuth (b). In the REMO analysis matrix (c) generic motion patterns are matched (d).

Legend to figure 2   The class diagram of the REMO data model shows the strict separation of the MPO MODELLING DOMAIN and the ANALYSIS DOMAIN.

Legend to figure 3   MPOs feature detached attribute functions to derive motion attributes from the known lifelines at arbitrary query times (e.g. `getAzimuth(t)`). Fine temporal granularities require interpolation between known fixes, coarse temporal granulatities require aggregation using a moving temporal window.

Legend to figure 4   The REMO formalism must be able to describe the basic and arbitrary REMO patterns. This figure shows a hypothetical REMO matrix for 10 deer over 15 time steps. A set of examples REMO patterns is highlighted: constance (a,d,e), concurrence (b,h), trend-setter (i), change (c,f,g), and infection (j).

Legend to figure 5   The pattern descriptors like pattern length, width and depth can be specified in different interfaces. The REMO pattern specified in this example is the constance $P = S([\sharp]\{4, \})$.

Legend to figure 6   Simplified UML sequence diagram of the implemented REMO pattern matching process.

Legend to figure 7   The football pitch shows the geospatial lifelines of 11 players covering a time frame of approximately 33 seconds. The REMO analysis matrix below illustrates the players' motion azimuth at a sampling rate of a second. The most prominent group motion in this interval of the match is a backwards motion of the team establishing a proper defense formation around $t = 15$ seconds.

Legend to figure 8   Three matches of a constant motion azimuth 270° over at least 10 seconds. Please note that the motion azimuth is classified into 8 discrete classes and computed over a temporal interval with a specific length. Hence the tracks of the matched objects 5, 14, and 21 is considered to be straight to the east even if the exact values may be scattering slightly around 270°.

Legend to figure 9   At t_{15} almost the whole team synchronously moves back to the own goal to establish a defense formation. Only player 6 is classified to 315° instead of 270°.

Legend to figure 10   The lifelines in the football pitch illustrate the early backwards motion of players No. 5, 14, and 21. Only at t_{14} the followers No. 1, 2, 4, 8, 10, 24, and 25 join in to complete the trendsetting pattern described in formula (17). Please note that the REMO matrix plot contains 16 overlapping instances of the same trend-setter pattern.

Legend to figure 11   The figure illustrates the motion of the 185 districts in the ideological space developed by Hermann and Leuthold (2003) using the MPO Viewer of the REMO prototype application. The topmost frame contains the abstract ideological space spanning between the dimensions political left (left) vs. political right (right) and ecological (top) vs. technocratic (bottom). Every point chain holds the consecutive fixes of one district. For this screen-shot the time-frame controlling sliders at the bottom were set to the time *1995* and a time-frame length of *4* years. Thus each point chain represents all fixes of its district in the ideological space between 1993 and 1997. The arrow

in the top right corner highlights the motion of the district Meilen to the 'ecological-right' corner of the ideological space.

Legend to figure 12 The figure illustrates the political divergence of the German speaking part from the francophone part of Switzerland in the 1990s. (a) gives an overview of the motion azimuths of the 185 districts from 1981 until 1999. (b) and (c) show constance patterns towards the right respectively left political pole. (d) and (e) highlight the concurrence patterns of the trend. Finally (f) and (g) identify the trend-setters of the trend.
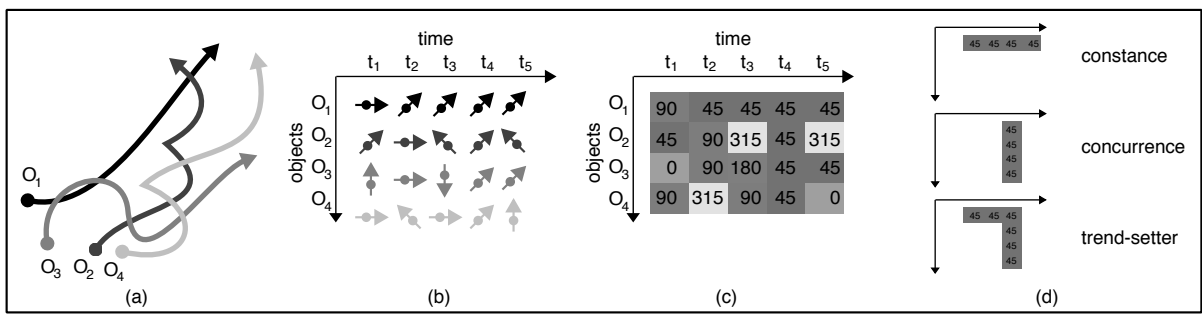
Figure 1: Basic REMO concept
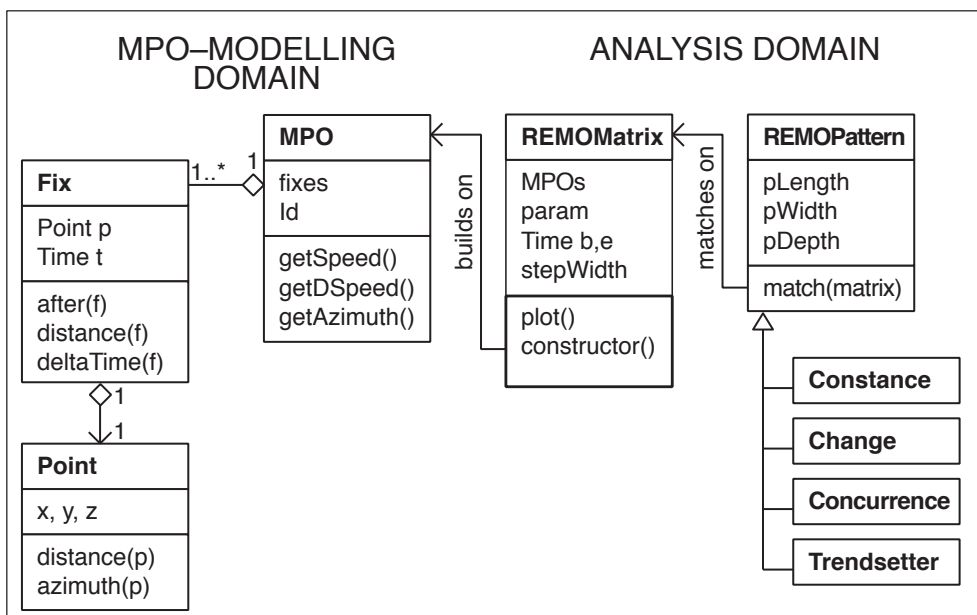
Figure 2: REMO data model
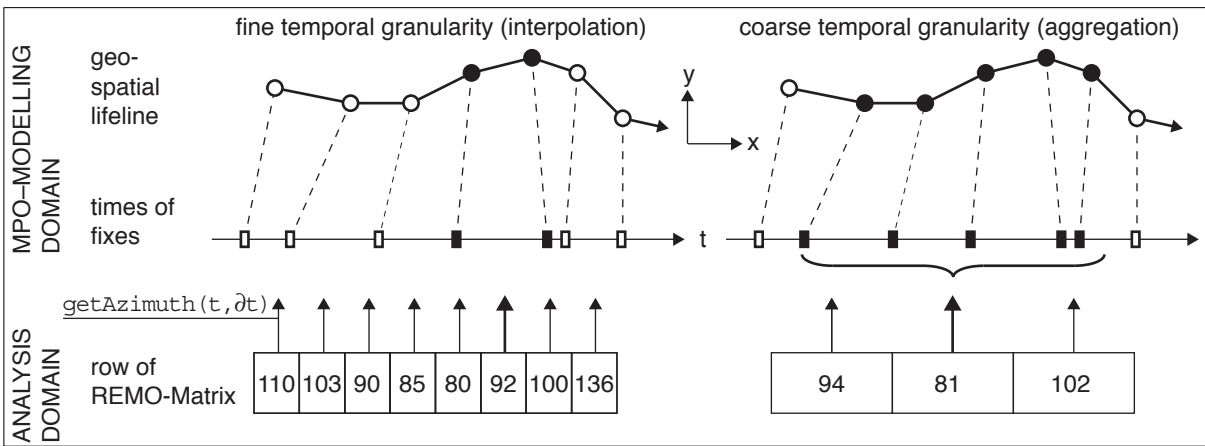
Figure 3: Detached attribute functions

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}
O_1	90	45	45	45	45	90	90	90	135	135	135	90	45	0	0
O_2	45	90	315	45	315	315	0	0	45	135	135	90	45	0	45
O_3	0	90	315	45	315	0	0	45	90	135	90	60	45	90	135
O_4	90	315	90	45	0	0	45	45	90	90	90	45	45	90	135
O_5	45	270	315	315	0	0	45	45	90	135	180	180	225	270	360
O_6	45	90	90	45	45	45	90	90	135	135	90	45	45	90	135
O_7	45	0	0	0	0	45	90	135	90	45	45	45	45	90	45
O_8	0	0	0	0	0	45	90	90	135	135	90	90	90	45	45
O_9	0	0	45	45	90	90	90	45	45	45	45	45	45	360	135
O_{10}	90	45	45	45	90	45	0	0	0	45	45	45	45	0	0

Figure 4: Hypothetical examples

Table 1: Quantifiers

length quantifiers	description
$\{l\}$	exactly l time steps (e.g. 4 in a row)
$\{l,\}$	at least l time steps (e.g. at least 4 in a row)
$\{l,k\}$	between l and k time steps (e.g. between 4 and 6 in a row)

width quantifiers	description
$\{n\}$	exactly n objects at same time (e.g. 4 objects)
$\{n,\}$	at least n objects at same time (e.g. at least 4 objects)
$\{n,m\}$	between n and m objects (e.g. between 4 and 6 objects)

element quantifiers	description
$v?$	motion attribute value v is optional
$v*$	0 or more times motion attribute value v
$v+$	1 or more times motion attribute value v

Table 2: Pattern depth descriptors

depth quantifiers	description
$[v]$	exactly motion attribute value v (e.g. azimuth 45°)
$[\leq v]$	at least motion attribute value v (e.g. azimuth 45°or more)
$[\geq v]$	not more than motion attribute value v (e.g. azimuth 45°or less)
$[> v]$	more than motion attribute value v (e.g. more than azimuth 45°)
$[< v]$	less than motion attribute value v (e.g. less than azimuth 45°)
$[v - w]$	motion attribute value range between v and w (e.g. azimuth between 45°and 135°)
$[^\wedge v]$	all possible motion attribute values except v
$[u \mid v \mid w]$	motion attribute values u or v or w (e.g. 45°or 90°or 135°)
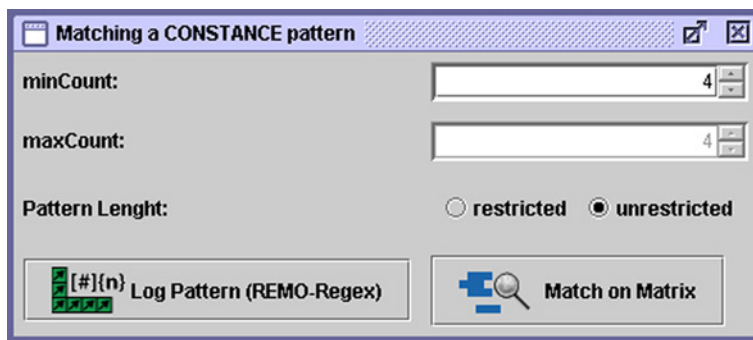
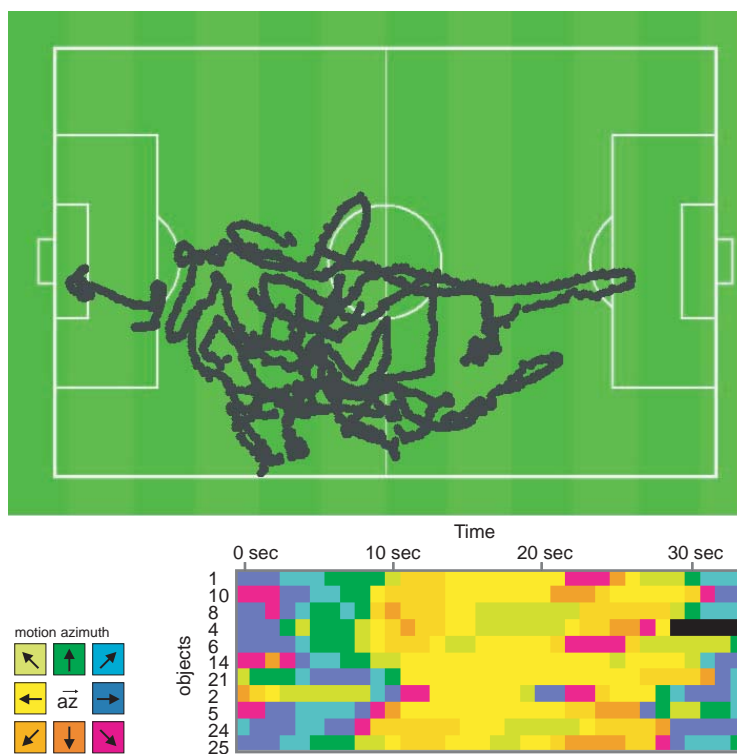Figure 5: Pattern builder interface
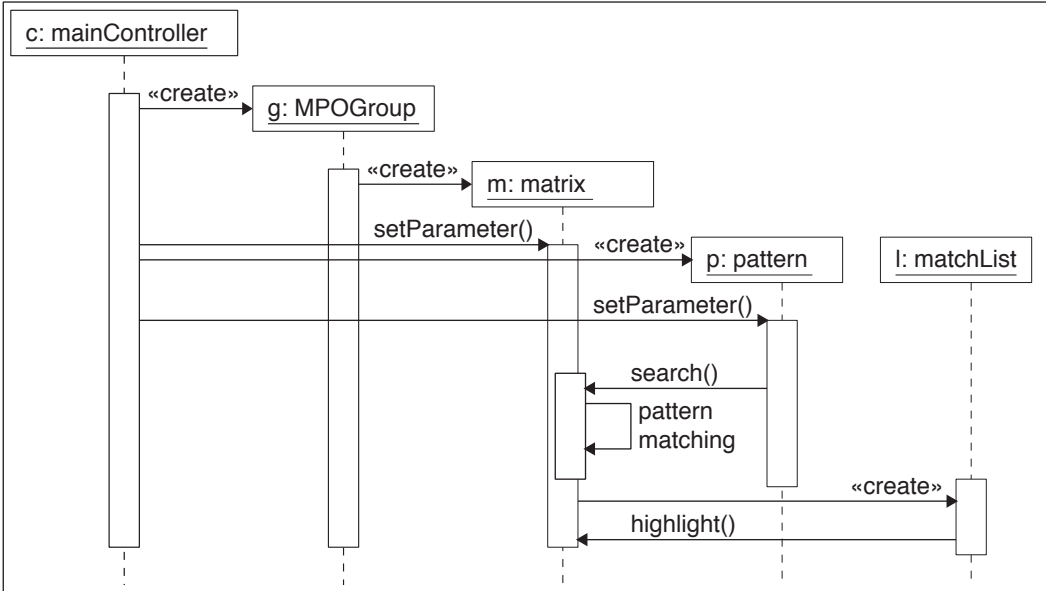
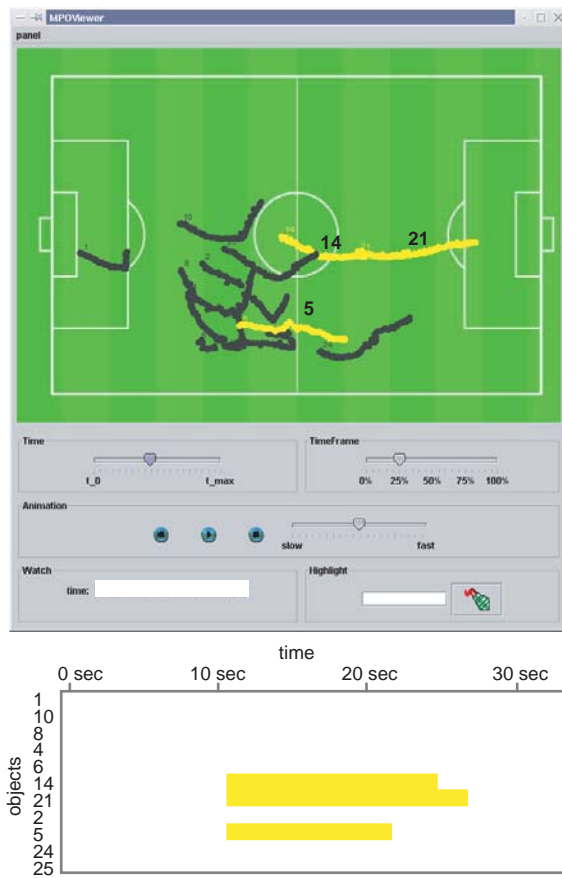Figure 6: REMO data mining process

Figure 7: Tracked football players

Figure 8: Constance
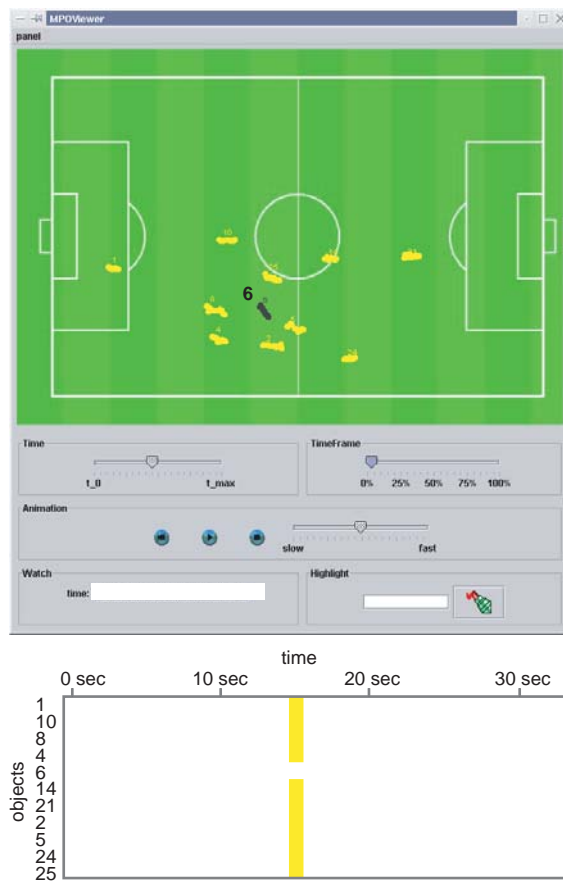
Figure 9: Concurrence
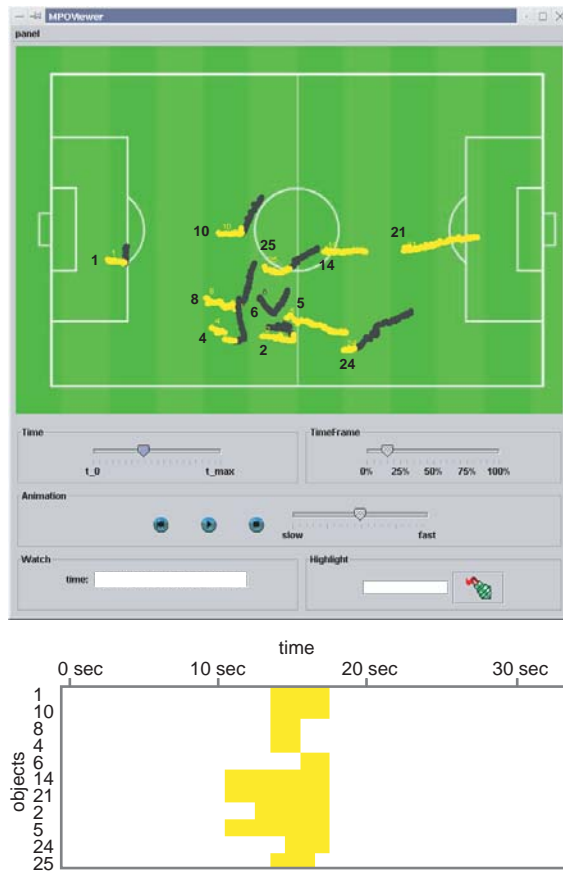
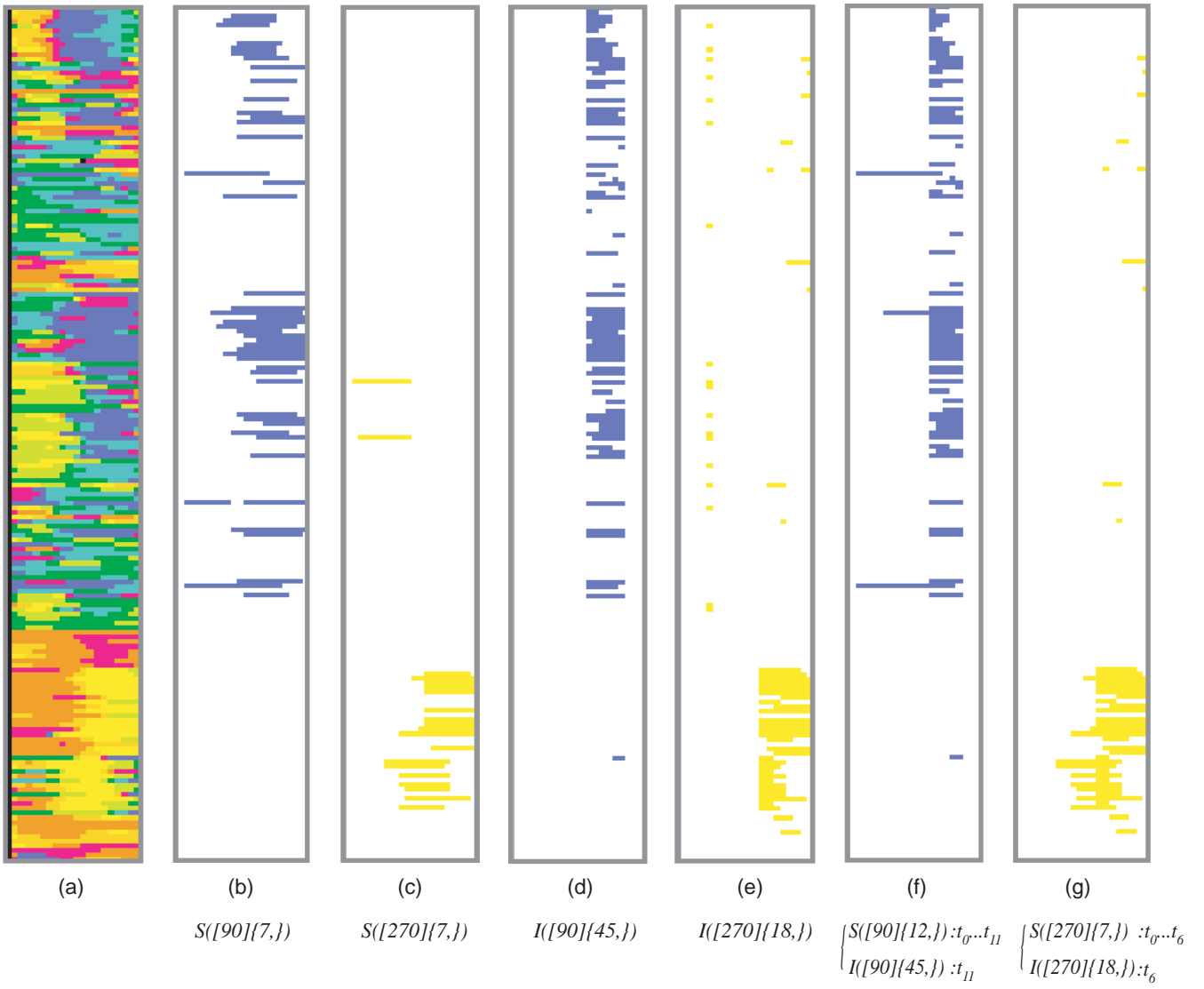Figure 10: Trendsetter

Figure 11: Swiss districts moving in ideological space

Figure 12: Results from mining the districts motion