Technologie-Zentrum Informatik

Technical Report 42

Workshop on Behaviour Monitoring and Interpretation BMI´07

In Conjunction with
30th German Conference on Artificial Intelligence
Osnabrück, 10th September 2007

Björn Gottfried (ed.)

TZI-Bericht Nr. 42
2007

Universität Bremen

TZI-Berichte

Herausgeber:
Technologie-Zentrum Informatik
Universität Bremen
Am Fallturm 1
28359 Bremen
Telefon: +49-421-218-7272
Fax: +49-421-218-7820
E-Mail: info@tzi.de
http://www.tzi.de

ISSN 1613-3773

Preface

Over the last years great advances have been made concerning the development of sensor technologies. What has been the realm of video monitoring for a long time is complemented by new technologies. In particular, the simplicity of devices equipped with GPS, infrared, bluetooth, RFID and other technologies make it easy to deploy them for the purpose of determining the locations of objects in space and time. As a consequence, researchers from different fields have been recognised how these technologies improve their work; sometimes, even new application fields are discovered. While video analysis requires very sophisticated methods for analysing the locomotion of objects, other sensors are often restricted to record positional data. That is, possibilities of tracking automatically objects in space and time are rediscovered, and a new class of problems arises: which sensors are to be used in order to monitor objects and how to interpret the recorded data, in order to solve a given problem? Being a constituent part in a number of application fields, such as in Smart Homes or in the context of location based services, it is worth putting one's attention to the problem class of behaviour monitoring and interpretation.

The first workshop on behaviour monitoring and interpretation, BMI for short, is held in conjunction with the 30th German Conference on Artificial Intelligence, in Osnabrück, 10th September 2007. The submitted papers show that researchers from different fields, such as computer science, geography, architecture, or biology are interested in these issues. While the call for papers says that the workshop is devoted to monitor single objects or small groups of objects, the scenarios described in the submitted papers concern individuals tracked in their car, the observation of the spatial behaviour of students on their campus, the analysis of behaviour patterns of mice in a maze, and the monitoring of the behaviour of pedestrians. These observations aim at determining regularities or irregularities in the spatial behaviour of single people or groups of people or animals, up to the recognition of intentions based on these observations. Behaviours are monitored at different scales, extending from table top scale via the scale of gestures via that one of a soccer pitch, the scale of a campus via the scale of a city to even larger scale spaces in geography. Besides the consideration of different scenarios, a number of methods for the interpretation of recorded data are investigated. These methods include qualitative spatial and temporal reasoning, formal languages, logic based inferences, and visual analytics.

The aim of BMI consists in providing a forum for people from different areas who are all interested in monitoring and analysing objects in space and time. While researchers from different areas frequently develop their own methods, it is the idea of BMI to foster the interchange of ideas, methods, and approaches in the context of behaviour monitoring and interpretation.

September 2007

Björn Gottfried
Program Chair
BMI'07

Organisation

Workshop Organisation

Björn Gottfried
Centre for Computing Technologies (TZI)
University of Bremen
Am Fallturm 1
28359 Bremen
Germany

Proceedings

Arne Schuldt	Centre for Computing Technologies, University of Bremen, Germany

Program Committee

Timothy D. Adlam	University of Bath, UK
Stephen Balakirsky	National Institute of Standards and Technology, USA
Roland Billen	University of Liege, Belgium
Christophe Claramunt	Naval Academy Research Institute, France
Christian Freksa	University of Bremen, Germany
Antony Galton	University of Exeter, UK
Hans Werner Guesgen	Massey University, New Zealand
Karin Klabunde	Philips Research, Germany
Paul McCullagh	University of Ulster at Jordanstown, UK
Gerhard Navratil	Technical University Vienna, Austria
Bernd Neumann	University of Hamburg, Germany
Christoph Schlieder	University of Bamberg, Germany
Sabine Timpf	University of Würzburg, Germany
Nico Van de Weghe	Ghent University, Belgium
Howard D. Wactlar	Carnegie Mellon University, USA
Stephan Winter	University of Melbourne, Australia
Stefan Wölfl	University of Freiburg, Germany
Yaolong Zhao	University of Tsukuba, Japan

Table of Contents

Analysing Behaviour Through Vision

Extracting Patterns of Individual Movement Behaviour from a Massive Collection of Tracked Positions

Gennady Andrienko and Natalia Andrienko

Fraunhofer Institute IAIS
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
gennady.andrienko@iais.fraunhofer.de
http://www.ais.fraunhofer.de/and

Abstract. A EU-funded project GeoPKDD develops methods and tools for analysis of massive collections of movement data, which describe changes of spatial positions of discrete entities. Within this project, we design and develop methods of visual analytics, which combine interactive visual displays with database operations and computational methods of analysis. In this article, we demonstrate by example how visual analytics methods can help in acquiring knowledge about the movement behaviour of an individual from a very large set of movement data.

1 Introduction

A EU-funded project GeoPKDD - Geographic Privacy-aware Knowledge Discovery and Delivery (IST-6FP-014915; see http://www.geopkdd.eu) - aims at developing methods and tools for analysis of massive collections of movement data. Movement data describe changes of spatial positions of discrete entities, which preserve their integrity and identity, i.e. do not split or merge. Within this project, we develop methods for supporting human analysts in visual inspection of movement data and detection of characteristic patterns of movement behaviours.

It is commonly recognised that interactive and dynamic visual representations are essential for gaining understanding of spatial and spatio-temporal data and underlying phenomena. However, visualisations alone are insufficient for exploration and analysis of massive data collections. This is not only the matter of technical limitations such as the screen size and resolution or the speed of rendering but also of the natural perceptual and cognitive limitations of the humans who need to view and interpret the visual displays. Hence, it is necessary to combine visualisation with computational analysis methods, database queries, data transformations, and other computer-based operations.

Recently, we have developed a theoretical basis for the creation of methods for visual analysis of movement data (Andrienko and Andrienko 2007). In particular, we have defined the possible types of behavioural patterns that can be detected by analysing movement data alone and in combination with data about other phenomena. Next, we have envisaged the kinds of data transformations, computations, and visualisations that could enable a human analyst to detect these pattern types in truly massive data, possibly, not fitting in a computer's memory. On the basis of the previous works (Tobler 1987; Dykes and Mountain 2003; Laube, Imfeld, and Weibel 2005; and others), we have suggested a set of techniques where a key role belongs to aggregation and summarisation of the data by means of database operations and/or computational techniques.

For a practical verification of this choice of techniques resulting from a theoretical analysis, we started a prototype implementation of a visual analytics (Thomas and Cook 2005) toolkit for movement data. In this article, we demonstrate by example how visual analytics methods can help in acquiring knowledge about the movement behaviour of an individual from a very large set of movement data.

2 The Example Dataset

The example dataset consists of more than 60,000 records of positions of a car, which has been tracked during 5 months. The data have been recorded only when the car moved, i.e. there are no records for stops and still periods. The temporal spacing of the records is mostly 1 second; however, the records corresponding to periods of uniform movement (i.e. with constant speed and direction) are sparser. The data are stored in a relational database.

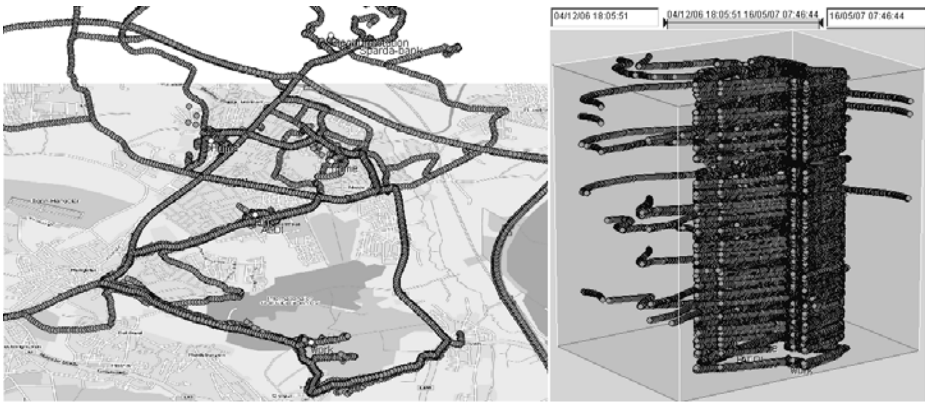


Fig. 1. An attempt to display all individual data items from a large dataset is not productive for data exploration and analysis.

The dataset is too large for a straightforward visualisation of all data items. Fig.1 demonstrates the result of showing all positions of the car on a map (left) and in a space-time cube (right), where the horizontal plane represents the geographical space and the vertical dimension represents the time. The map, in fact, reveals only the street network in the area where the car moves. It is even impossible to see which streets are used frequently and which only occasionally, because the symbols greatly overlap. Temporal filtering and display animation do not help much: the intervals of movement are very short in relation to the 5-months long time period and therefore hard to extract through interactive filtering and hard to detect by viewing an animated display, where most of the time nothing happens. Hence, to be able to extract useful information from this mass of data, it is necessary to summarise it somehow before trying to visualise.

In our work, we put a particular focus on the use of database operations for data summarisation and other data transformations and computations. On this basis, we strive at developing scalable visual analytics methods, which could be applied even to datasets not fitting in the computer memory. Besides database operations and visualisation, we utilise data mining techniques, as will be seen from the following sections.

3 Detecting important places

A temporally ordered sequence of all positions of an entity is not a meaningful object for analysis since the entity does not necessarily move all the time (thus, in our data, the time of movement is much less than the time of stillness). It is reasonable to divide the sequence into trajectories or into movement episodes. A *trajectory* is a sequence of items corresponding to a trip of an entity from one location (source) to another (destination) where the source and destination are defined semantically (e.g. home, work, shop, etc.) or according to the time the entity spends in a location. *Movement episodes* (Dykes & Mountain 2003) are fragments of trajectories where the movement characteristics (speed, direction, sinuosity, etc.) are relatively constant whereas a significant change indicates the beginning of the next episode.

In our exercise on analysing the car movement data, we assume that we have no background knowledge that would allow us to divide the data into trips using semantic criteria. Moreover, it is one of the tasks of our analysis to extract and interpret the sources and destinations of the trips. Therefore, we need to find the sources and destinations on the basis of the temporal criterion, i.e. according to the time spent in a location.

As we have explained, the records of the car positions have been made only when the car actually moved; hence, the stops and periods of stillness are present in the data implicitly as temporal gaps between successive records. It is easy to find such gaps with the use of database operations; however, it is necessary to specify the minimum temporal distance between records to be treated as a "gap". This threshold can be chosen quite arbitrarily. Interestingly, by setting different temporal thresholds, it is possible to find places of different importance for the moving object, i.e. car user in our case. Thus, setting a threshold of several hours should result in finding places where the person spends much time. These will include person's home and work.

Fig.2 presents the spatial positions of the trip starts and ends, which have been extracted from the database using a temporal threshold of 2 hours. The positions are shown as small circles on a map. The map on the left shows all extracted positions. Surprisingly, there are much more different positions than could be expected. The maps in the middle and on the right show the starts and ends separately. It is easy to notice that the starts are much more dispersed in space than the ends. This looks very strange: the start position of a trip should normally coincide with the end position of the previous trip. The reason for the observed discrepancy is that the GPS (Global Positioning System) device, which is used for collecting the data, needs some time for warming up, detecting satellites, and establishing connections with them. Therefore, the device starts recording the positions of the car not from the moment when a trip begins but later. Hence, our data are incomplete, and the real times and positions of trip starts are unavailable. This feature needs to be taken into account when analysing the data.

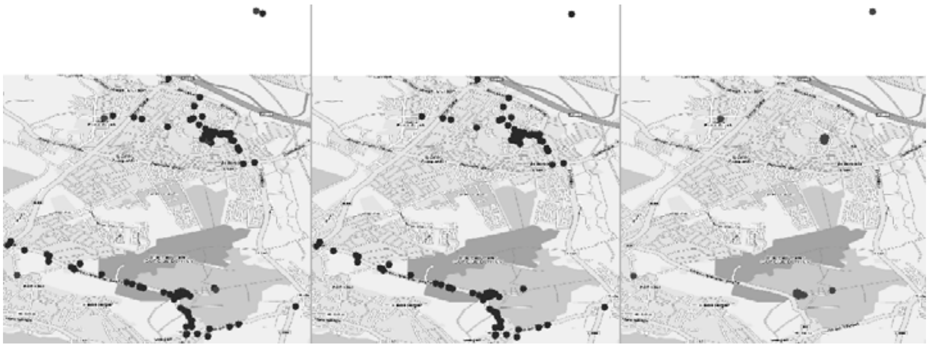


Fig. 2. The positions of the starts and ends of the trips have been extracted by setting a temporal threshold of 2 hours. Left: all extracted positions; middle: starts; right: ends.

It is reasonable to assume that the spatial positions of the trip starts are the same as the destinations of the previous trips. Therefore, we can ignore the extracted starts and look only at the ends, which are more reliable. On the right of Fig.2, we can see 5 different places where the trip destinations are located. Naturally, we are interested first of all in finding frequently visited places, or, in other words, places where the trip destinations are clustered. A map display is not appropriate for this purpose: it is hard to guess how many overlapping circles there are in each place. Instead, we can apply computational techniques for detection of spatial clusters, which are developed in the research area of data mining.

The map on the left of Fig.3 shows the result of applying a clustering tool to the destinations detected with the use of the temporal threshold of 2 hours. The tool has found two clusters; the dark circles mark the corresponding positions. The remaining positions, which are represented by lighter circles, were classified as noise. From the two clusters, the one on the north contains 118 positions and the other, which is on the south, contains 77 positions. It is reasonable to conclude that the larger cluster is located near the home of the car user and the smaller cluster is at the place where the person works.

In the middle of Fig.3, the map presents the results of applying the same clustering tool to the trip destinations extracted using the temporal threshold of 1 hour. Additionally to the two clusters detected before, one more cluster consisting of 11 positions has been found west from the place interpreted as “home”. On the right, the clusters of destinations where the car user spent at least 5 minutes are shown. There are five clusters, including the three previously detected clusters. Two more clusters have appeared in the centre of the map.

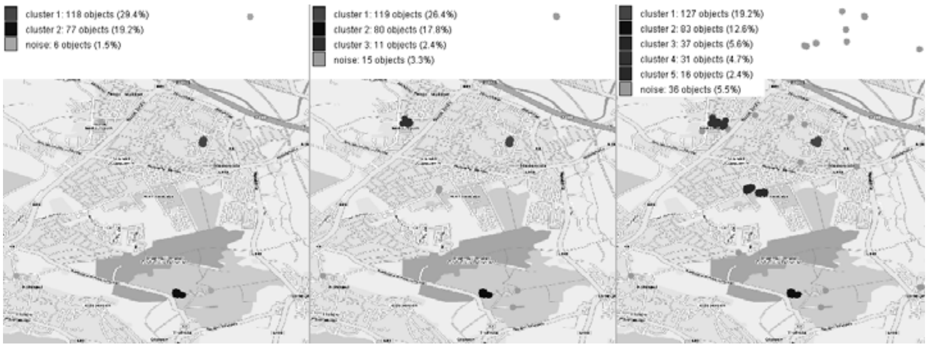


Fig. 3. Results of a clustering method applied to 3 sets of trip destinations extracted using different values for the temporal threshold: 2 hours (left), 1 hour (middle), and 5 minutes (right).

The meaning of the places detected in this way can be established using background knowledge about the territory or information provided by the background map. Unfortunately, the available background map is just an image with a low level of detail. It is only possible to find out that one of the clusters (cluster 3) is located at a shopping centre named “Huma Einkaufspark”. However, we are familiar with the territory and can interpret also the clusters 4 and 5: these are located in a shopping area where several stores are separated by a street.

Hence, by extracting and analysing trip destinations, we have found the places where the car user lives, works, and shops. There are two frequently visited shopping areas. The car user spends more time in “Huma Einkaufspark” than in the other shopping area.

Using background knowledge about the territory and/or background map, it is possible to identify also the places visited less frequently, as, for example, post office or bank.

4 Analysing trip directions

After determining the significant places visited by the car user (these places will be henceforth called “places of interest”, or POIs), we would like to know how the person moves between them. Thus, we can see that one of the shopping areas is located between the person’s home and work. Does the person visit it on the way from the work to home and, if so, how often? Does the car user ever go from the work to the other shopping area? Were there any trips from one shopping area to the other?

In order to answer these and similar questions, it is reasonable to count the number of trips between each pair of POIs. However, each position in the dataset is specified only through a pair of coordinates (latitude and longitude), i.e. as a geographical point without any semantics. In order to enable a software tool to treat the starting and ending positions of the trajectories as particular places of interest, it is necessary to specify the places explicitly as named areas surrounded by boundaries. For this purpose, one can use computational functions available in geographic information systems (GIS) or spatial DBMS to build buffer zones around the positions of the trip starts and ends and then assign meaningful names to the zones obtained. Another approach is to encircle the areas on the map manually. Here, we shall use manually defined POIs. Besides encircling the trip destinations we could interpret, we also considered the extracted starting positions (Fig.2 middle) and associated some of them with the most probable trip sources. Thus, the start positions of many trajectories lie on the roads passing near the place of the person's work. As we know, these are false starts. It is reasonable to assume that the real source of the corresponding trips is the place of the work. Hence, we have drawn several shapes enclosing the false start positions and named them "(work)*", "(work)**", and "(work)+". Similarly, we have defined an additional POI named "(home)*" by enclosing the false start positions of the trips starting, most probably, at home. Fig.4 shows the POIs we have specified.

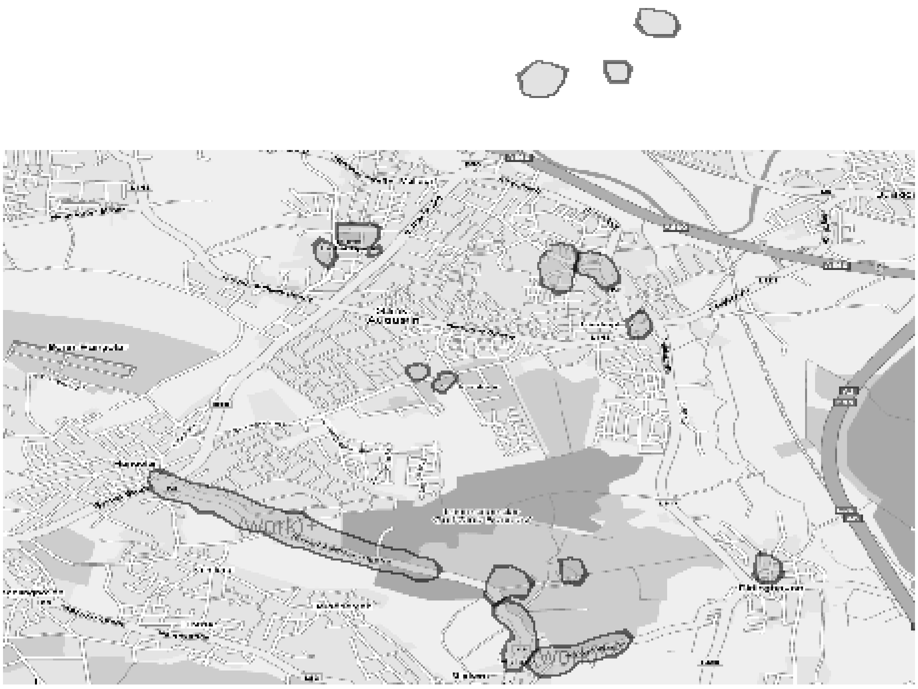


Fig. 4. The places of interest defined by encircling areas on a map.

After the places of interest have been defined, a software tool can attach their names to the positions of the trip starts and ends lying within the areas. Then, it becomes possible to count the number of trips between each pair of places. The resulting counts can be visualised, for example, in an interactive matrix display shown in

Fig.5. The rows of the matrix correspond to the trip sources, the columns to the destinations, and the sizes of the rectangles in the cells encode the numbers of the trips. By putting the mouse cursor on a cell we can learn the exact number of trips made from the corresponding source to the corresponding destination.

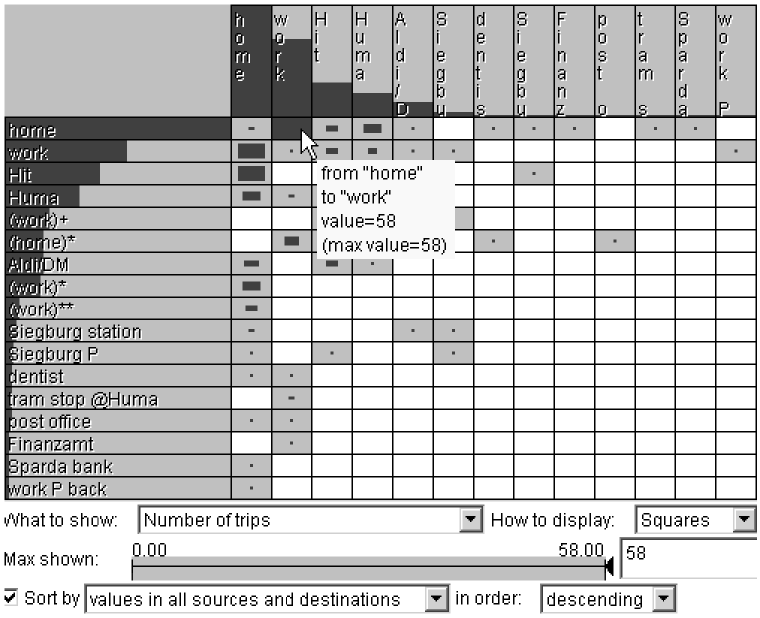


Fig. 5. The numbers of trips between pairs of POIs are represented by proportionally sized rectangles in cells of a matrix where the rows correspond to the trip sources and columns to the destinations.

Thus, during the period under the study, there were 76 trips to work, from which 68 trips were from home (58 from the POI “home” plus 10 from the POI “(home)*” enclosing the false starts of the trips from home). There were 2 trips to work from the shopping area “Huma” and no trips starting in the other shopping area.

Let us now look to which places the person drives from the work. There are several source POIs associated with the place of work. For a more convenient exploration, we can apply an interactive filtering tool to select only the trips starting from any of these POIs. The interactive matrix display reacts to setting the filter by removing irrelevant information (Fig.6). Now, it is more convenient to learn that there were in total 80 trips starting from the work, from which 50 were directly to home, 7 to “Huma” and 19 to the other shopping area (17 to “Hit” plus 2 to “Aldi/DM” on the other side of the street), 2 trips to “Siegburg station” and 2 trips back to work.

As the matrix display lacks the geographical context, it may be useful to complement it with a map display. The map in Fig.7 shows the same summarised information about the trips from the work by vectors (directed lines) connecting the source and destination locations. The widths of the lines are proportional to the numbers of the trips between the respective locations. Unfortunately, the map is not easy to read because of the overlapping of the vector symbols. Still, the major destinations of the trips from the work can be grasped.

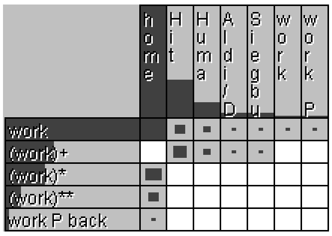


Fig. 6. The matrix display shows only the information about the trips from the work.

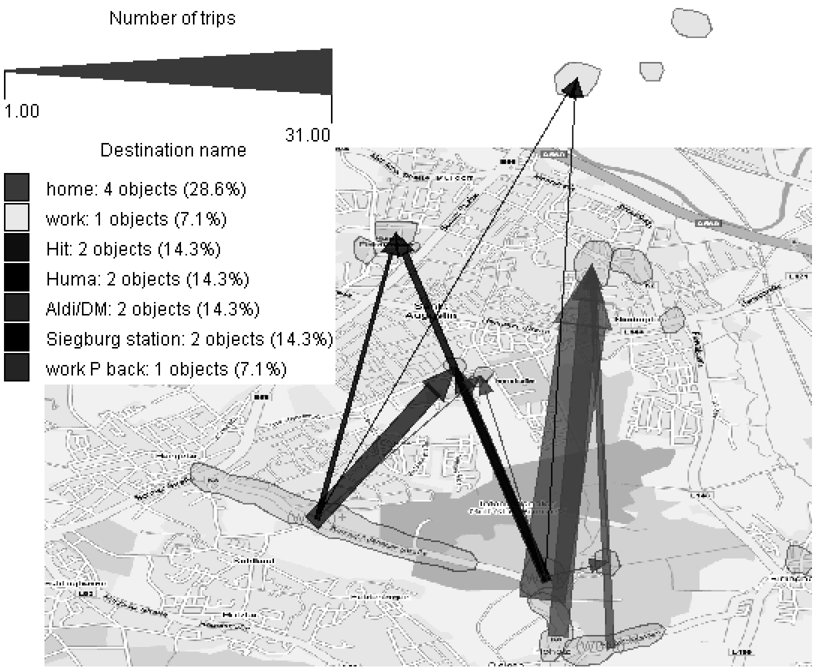


Fig. 7. The same information as in Fig.6 is shown on a map by vectors.

Besides filtering by trip origin, it is possible to set various other filter conditions. For instance, two screenshots of the matrix display presented in Fig.8 show the results of filtering the trips according to the time of the day when a trip begins. On the left, we see the summarised information about the trips starting from 8 to 11 hours, and on the right – from 17 to 20 hours. Most of the morning trips are from home to work, but the evening trips are much more varied. Analogously, it is possible to compare the trips made on working days with the trips made on weekends (Fig.9).

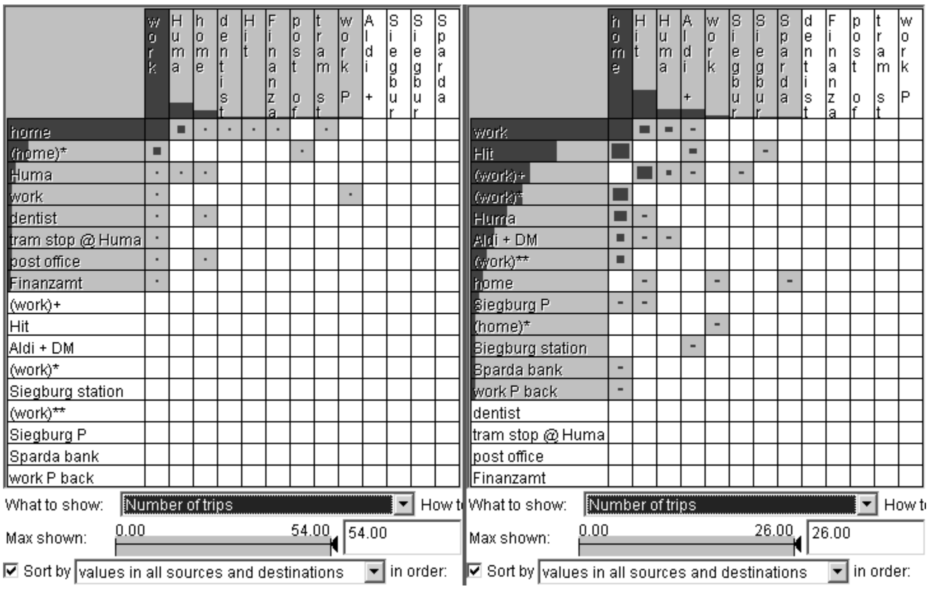


Fig. 8. Filtering of the movement data by the time of the day allows us to compare the major trip directions in the morning (left) and in the evening (right).

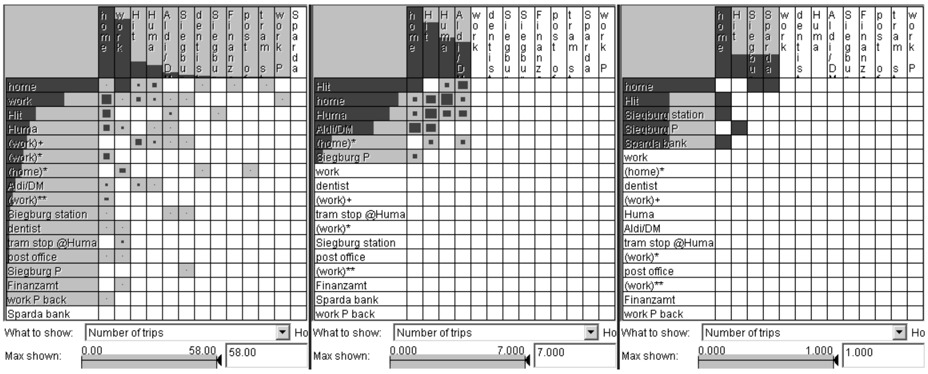


Fig. 9. Left: trips made on the week days from Monday to Friday. Middle: trips made on Saturdays. Right: trips made on Sundays.

5 Analysing trajectories

Both in the matrix display and in the map with vectors (Fig.7) the information about the trips is highly summarised: it is only possible to see the origins and destinations and the numbers of the trips. In studying movement behaviours of individuals, it is also important to investigate their trajectories, or, in other words, the routes they use. For example, in our case we would like to know what routes the person chooses on the way from home to work and back. If the person uses different routes, it is interesting to find out when which route is preferred and to make plausible guesses about the reasons for choosing this or that route.

For such investigations, we need a detailed representation of the person's trajectories in the geographical context, i.e. on a map or in a space-time cube. However, the representation of all trajectories at once results in an unreadable display similar to what can be seen in Fig.1. A reasonable approach is to explore the trajectories by interpretable portions with the use of the tool for interactive filtering. In particular, it is useful to select subsets of trajectories according to the sources and destinations of the trips. Thus, the map in Fig.10A shows only the trajectories starting from the work and ending at home (the trajectories have been defined using a 2 hour temporal threshold for dividing the sequence of positions). The trajectories are represented as polygonal lines, their starting points are marked by hollow squares (Fig.10B) and end points by filled squares (Fig.10C). It is hard to estimate the number of overlapping lines, but the filtering tool informs us that there are 74 trajectories from work to home among 201 trajectories in total.

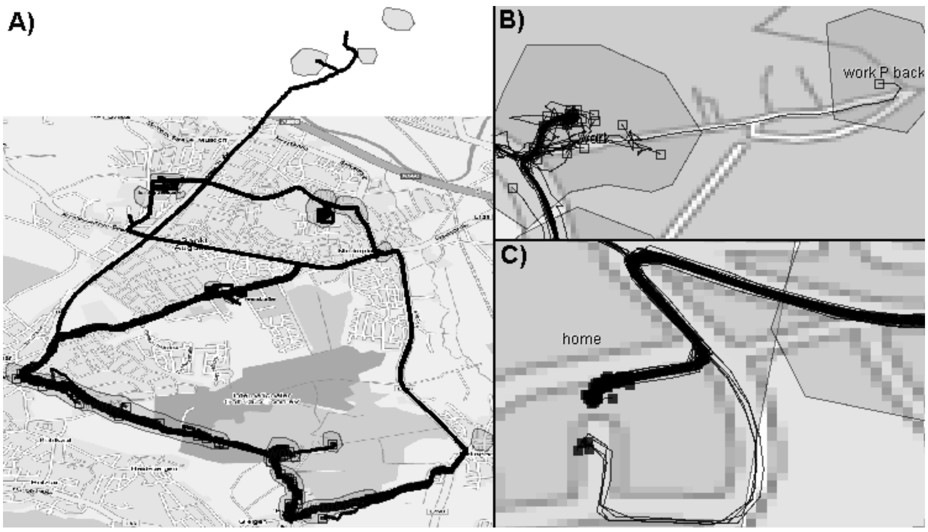


Fig. 10. A) The trajectories from work to home are represented on a map as polygonal lines. B) The start points of the trajectories are marked by hollow squares. C) The end points of the trajectories are marked by filled squares.

As the lines severely overlap, it is hardly possible to understand what routes the person uses for driving from work to home. It is necessary to group trajectories with similar shapes and to look at each group separately. One of the possibilities for groping is by interacting with the map display. Clicking on a position on the map selects all lines passing through this position or close to it. From the lines selected in this way one can make a group, or class. By clicking on different roads, it is possible to make several selections and to form several classes. Thus, in our case, we have defined 6 classes of trajectories from work to home differing in shape (Fig.11), three of which consist of singular trajectories (classes 4, 5, and 6 in the lower row in Fig.11). The most frequently followed route (upper left of Fig.11) is by the road on the east of the territory; the person used it 43 times. The second frequent route (upper middle), which was used 21 times, passes the shopping area in the middle of the territory.

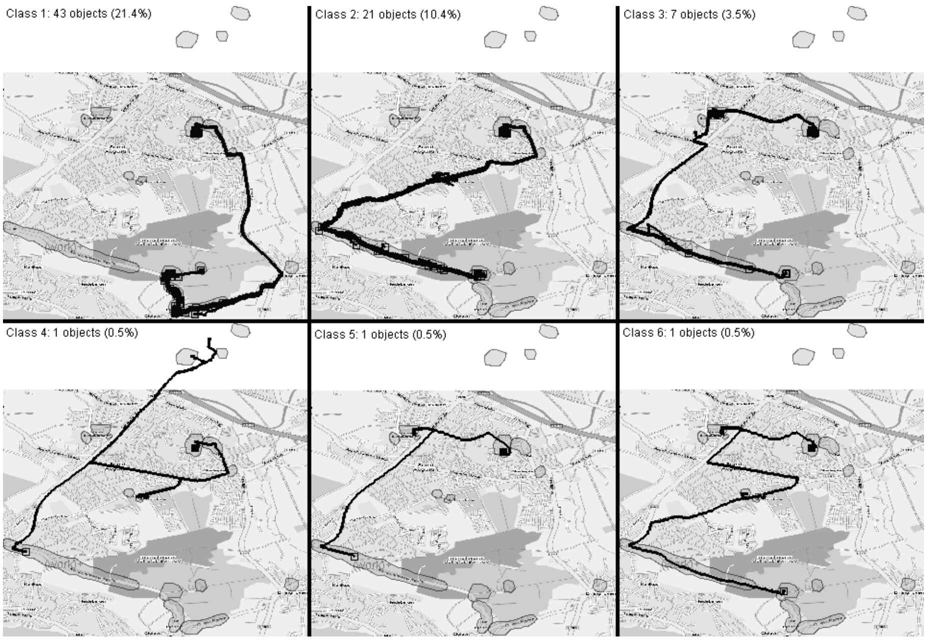


Fig. 11. By interacting with the map display, we have detected 6 different routes from work to home.

	N	N?	min	q1	med	q2	max	ave	stdd
	201	0	390	625	739	2872	14406	2382	2978
■	43	0	435	599	625	662	1028	637	95
■	21	0	641	1658	2745	3192	4501	2607	1163
■	7	0	1234	2018	2832	2971	5765	3239	1256
■	1	0	4306		4306		4306	4306	0
■	1	0	1755		1755		1755	1755	0
■	1	0	4406		4406		4406	4406	0
□	127	0	390	631	828	3358	14406	2863	3513

Fig. 12. Statistics of the trip duration for the different routes shown in Fig.11.

It is useful to look at various statistics about the classes of the trajectories. Thus, Fig.12 shows the statistics of the trip duration, in seconds: minimum, first quartile, median, third quartile, maximum, average, and standard deviation. The upper row of the table corresponds to the whole set of trajectories, the next 6 rows correspond to the groups of the trips from work to home we have defined, and the last row corresponds to the remaining trajectories. We can see that the route of class 1 takes the least time. This can be explained by the absence of POIs that could be visited on this way. It is highly probable that the routes corresponding to classes 2 and 3 are chosen when the person needs to visit one of the shopping areas. In order to check this, we would need a tool computing the time spent in each POI during each trip; however, such a tool is not available at the moment of writing this paper.

The histogram in Fig.13 shows the distribution of the trips by days of week, from Monday to Sunday. The light grey bars correspond to the entire set of trips. The black, dark grey, and medium grey segments show the proportions of the trips from the classes 1, 2, and 3, respectively. It is notable that the route corresponding to class 2 is most often (7times) chosen on Wednesdays but is used also in other working days of the week. The route corresponding to class 3 was used 3 times on Thursdays, 3 times on Fridays, only once on Monday, and never in the other days of the week.

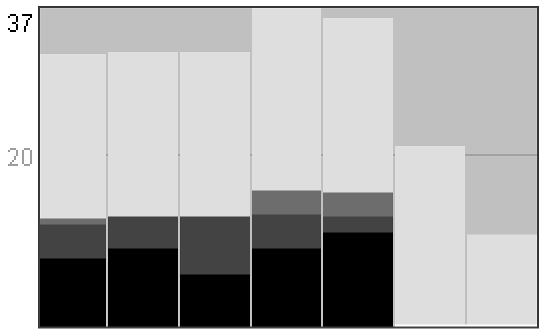


Fig. 13. The histogram shows the distribution of the trips by days of week.

Not only interactive techniques can be used to group trajectories by similarity but also methods for computational clustering, which are preferable in case of great overlaps between trajectories and/or complex shapes with loops and self-crossings. Automatic clustering of any items requires a method that computes the degree of dissimilarity (also called "distance"; this term is used in a wider sense than purely distance in space) between a given pair of items. Such a method is called "distance function". Clustering algorithms and distance functions suitable for trajectories are now under development within the project GeoPKDD. For the car movement data we have, we have implemented a simple distance function that takes into account the incompleteness of the trajectories: it compares trajectories starting from their ends and ignores differences in the lengths. Fig.14 presents a result of automatic clustering of the trajectories from work to home (it should be noted that clustering results may differ depending on the choice of clustering parameters, in our case, the distance threshold – the maximum allowed distance between members of a cluster). The clusters agree very well with the results of our interactive grouping: clusters 1 and 2 are exactly the same as our classes 1 and 2, cluster 3 includes all trajectories from our class 3 plus the single trajectory we have put in class 5, and the remaining two trajectories are treated as “noise”, i.e. as too dissimilar to the other trajectories.

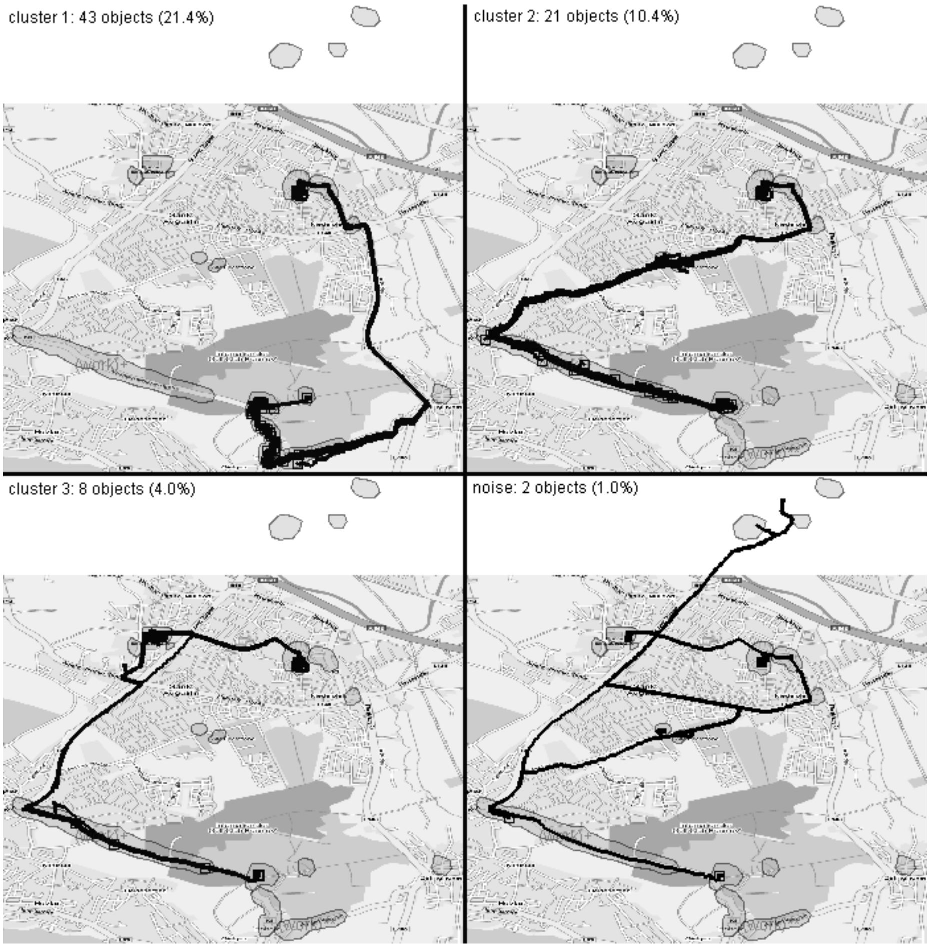


Fig. 14. A result of automatic clustering of the trajectories from work to home. Parameters: the distance threshold is 300 meters and the minimum number of cluster members is 3.

Let us now utilise the clustering tool to investigate how the car user goes from home to work. One of possible results of clustering is presented in Fig.15.

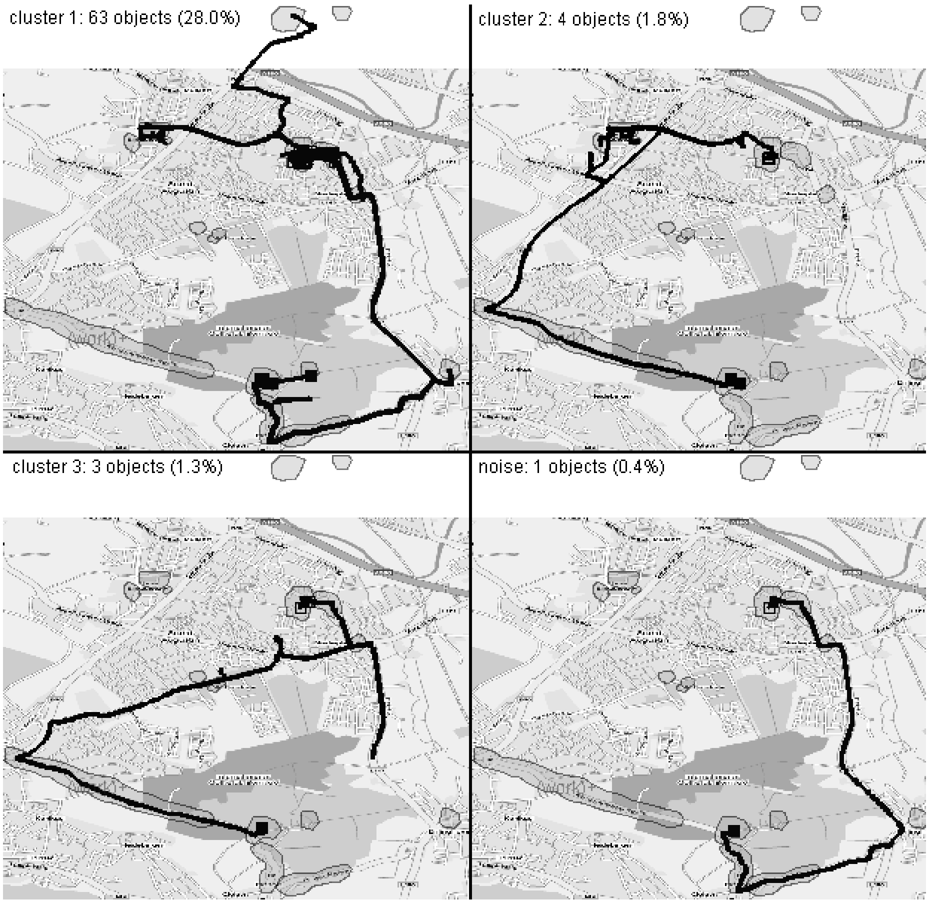


Fig. 15. A result of automatic clustering of the trajectories from home to work. Parameters: the distance threshold is 500 meters and the minimum number of cluster members is 3.

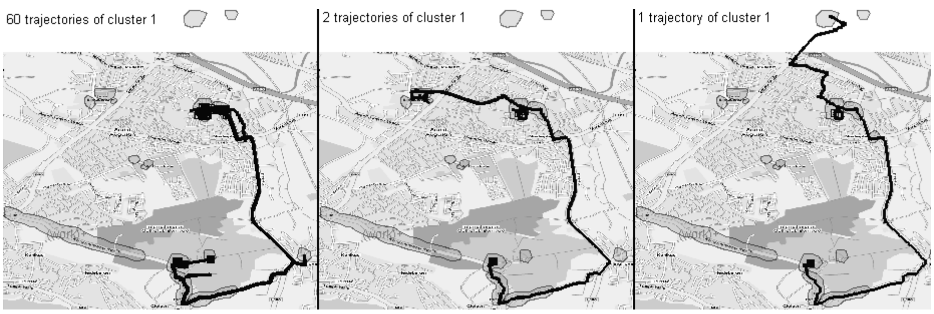


Fig. 16. The composition of cluster 1 from the previous figure.

As can be seen from Fig.15, the person takes almost always the eastern road for driving from home to work. The trajectories following this road are united in cluster 1. Fig.16 shows the composition of this cluster in some more detail. On the left, there are 60 trajectories of very similar shapes. In the middle, there are 2 trajectories where the person first visited the shopping area “Huma”, then returned home, and then drove to work. On the right, there is a single trajectory where the person visited the POI “Siegburg station” before going to work.

From the other trajectories from home to work, 4 go along the western road (cluster 2) and 3 trajectories use the diagonal road (cluster 3). Cluster 3 contains one peculiar trajectory: the person first drove half way along the eastern road and then returned back and took the diagonal road. Perhaps, there was some obstacle on the eastern road that day.

It can be noticed that the single trajectory marked as “noise” (bottom right of Fig.15) has the same shape as the standard trajectories in cluster 1 (Fig.16 left). This exposes a weakness of the distance function we use. For some unclear reason, the trajectory marked as “noise” consists of 525 different positions, which is much more than in the standard trajectories of cluster 1 (from 189 to 300). Our distance function cannot properly cope with such a difference and, evidently, requires improvement. At least two important implications can be derived from this observation. First, peculiarities of data to be analysed must be properly taken into account in designing and/or choosing methods for automated analysis. Second, a careful and critical examination of the results of automated methods is absolutely necessary. Interactive visual interfaces are appropriate instruments for this.

6 Conclusion

The main objective of this article was to demonstrate the use of interactive visual tools combined with database processing and computation for the exploration and analysis of large spatio-temporal datasets, more specifically, data about changes of spatial positions of discrete entities. We have shown how patterns of individual movement behaviour can be extracted from a very large number of position records and semantically interpreted. We could continue this investigation and learn much more about the person’s life style and habits. Such a possibility raises serious concerns about the privacy of individuals. Therefore, one of the main objectives of the project GeoPKDD is to develop mechanisms for preventing the disclosure of sensitive private information. Such mechanisms need to be incorporated both in computational and in visual tools for analysis.

References

1. Andrienko N and Andrienko G, 2007, Designing Visual Analytics Methods for Massive Collections of Movement Data, Cartographica, 42(2):117-138.

2. Dykes JA and Mountain DM, 2003, Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications, Computational Statistics and Data Analysis, 43 (Data Visualization II Special Edition):581-603
3. Laube P, Imfeld S, and Weibel R, 2005, Discovering relative motion patterns in groups of moving point objects, International Journal of Geographical Information Science, 19(6):639–668
4. Thomas JJ and Cook KA (eds), 2005, Illuminating the Path. The Research and development Agenda for Visual Analytics, IEEE Computer Society, USA
5. Tobler W, 1987, Experiments in migration mapping by computer, The American Cartographer, 14 (2): 155-163

Analysing Movement and Behavioural Patterns of Laboratory Mice in a Semi Natural Environment based on Data collected via RFID-Technology

Mareike Kritzler[1], Lars Lewejohann[2], Antonio Krüger[1]

[1]Institute for Geoinformatics, University of Münster, Robert-Koch Str.26-28, 48149 Münster, Germany
http://ifgi.uni-muenster.de
[2]Department of Behavioural Biology, University of Münster, Badstraße 9/13, 48149 Münster, Germay
http://www.ethologie.de

Abstract. In this paper we present a continuous 24 hour data collection and a semi automated data analysis of laboratory mice in a spacious indoor environment. The data is collected via an RFID tracking solution, a scale and an optical tracking system. The visualisation and the premilary analysis of the data provide information about behavioural and movement patterns of laboratory mice under semi-naturalistic conditions.

1 Introduction

In biomedical research mice play a dominant role as an animal model for deciphering gene functions in vivo. Especially in the investigation of hereditary human diseases numerous gene targeted mice were created. A detailed behavioural characterization of these mice aims to find differences between the genetically manipulated (transgenic; TG) mice and their wild-type conspecifics. Most commonly mice are tested in standardized but highly artificial situations that allow to analyze defined behavioural domains in detail but sometimes fail to bring about a thorough, externally valid behavioural phenotype [7]. Here we report on a semi-naturalistic setup where TG mice who carry a genetic predisposition to develop Alzheimer's disease like symptoms are constantly monitored 24h-7d by means of RFID technology.
Aim of this project is to support the direct behavioural observations of the mice that is carried out by humans. The population under surveillance consists of up to 40 TG and wild-type mice that are living in a semi natural environment (SNE). The SNE is realized as a large indoor cage measuring 1.75 x 1.75 x 2.1 m (L x W x H) comprising several floors which are connected by Plexiglas tubes. Mice are individually marked with RFID-chips and positional data is obtained continuously from various antennas placed in the SNE. The automated tracking solution is established to collect behavioural and movement data of the mice 24h-7d. A GIS module is developed to analyze the gathered data. The project focuses on the automated detection of potential differences in behavioural and movement patterns of the TG and wildtype mice.

2 Related work

To obtain behavioural information from different sensors is even used by testing of humans. The combination of different sensors to collect data and subsequently obtain behavioural information was applied in various species including cows [4; 5] and even humans. For humans one well described testing environment was an everyday office and it was demonstrated that simple sensors support models which are able to estimate human interruptibility [1]. Furthermore in a museum environment information about visitors were collected. In use are defined visiting styles to assign the museum visitors to different classes [8]. Also machine learning techniques have been applied also to detect and classify common motion patterns, to support users with dementia in their daily routines [6]

3 Scenario setup

For the collection of behavioural and movement data the SNE had to be structured in a way that only defined passages were accessible. A schematic view of the cage design and RFID based tracking solution is shown in figure 1.

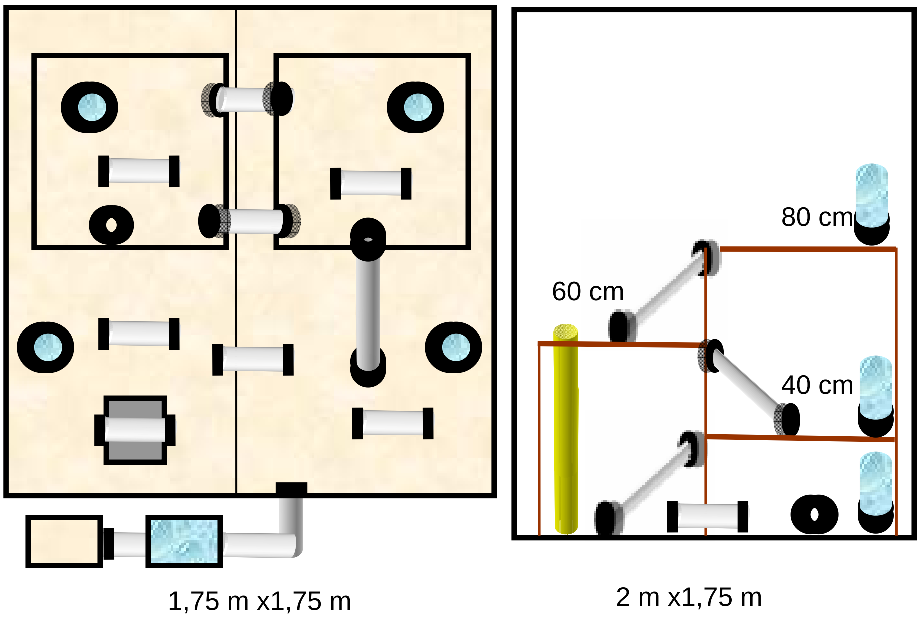


Figure 1: Schematic view of the SNE. Left: top-view, right: side view. Black circles represent coil antennas that are connected to Plexiglas tubes (grey) or water bottles (blue).

Cage design

In the setup following information about the mice is of interest: changing floors, movement on the floors, the direction of movement, the home range of individuals, drinking and emigration behaviour.
The design of the SNE has to warrant that movement on and between different floors can be detected. Therefore several constraints for the freedom of movement must be realized. The SNE comprises five floors, two on the ground and three in different levels above the ground. These floors are connected by Plexiglas tubes and / or rope. Outside the SNE an emigration cage is provided that can be accessed from the ground floor (i.e. to give shelter to low-ranking animals within the group hierarchy) via a tube and crossing a water basin (see figure 1).

Data collection via RFID

The RFID antennas are placed on points where the mice must cross. On every Plexiglas tube two antennas at both ends are attached. This allows to detect when a mouse changes floors, in which direction and at what speed mice cross the tubes. Each floor contains an antenna beneath the drinking bottle to get data about the drinking behaviour and to establish a warning system when a mouse does not drink. Furthermore on every floor is a tube supplied with two antennas which enables to collect data about the movement on the floors. Two antennas are used to identify mice using a scale. In the SNE are at least 29 antennas integrated in the SNE (see table 1).

Table 1: Distribution of the 29 antennas; in the left column: number of antennas, right column: position of the antennas

Number of antennas	Positioning
10 antennas	Floors
6 antennas	Floor connection
2 antennas	Connection floor 0
5 antennas	One drinking bottle per floor
2 antennas	Rope
2 antennas	Exit SNE → Emigration cage
2 antennas	scale

Data collection via Jerry TS

To collect data of the mice we use the RFID technology. The RFID-System (Trovan Electronic Identification Systems) consists of reader (LID 665 Miniature OEM Board), ring antennas (air-core coil antenna for LID 665) and animal glass transponders (ID 100). All mice wear a passive integrated transponder (PIT) that is injected subcutaneously between the scapulas. The transponder ID is read while a mouse traverse the electromagnetic field which is established by the ring antennas, e.g. when passing through tubes or visiting drinking places. The minimum distance between two antennas is 20 cm. The ID of the transponders is read within a distance of 0.5 cm. The readers are able to read several transponders at the same time at a maximum rate of 26 Hertz.

We wrote a Java based software component (JerryTS) to configure the RFID reader and to store the read data in a database. If the transponder ID is read (a mouse gets into the electromagnetic field of a ring antenna), a data set is created which consist of date, time, milliseconds, antenna ID and transponder ID. This data is stored online in a relational data base [2].

Data collection via scale

The described setting was extended by a scale (Kern & Sohn GmbH: Typ 440-33N) that continuously allows to measure the weight of individual animals. The scale was protected against dirt and damage by a plastic body (see figure 2). The transponder ID is read by the RFID antennas placed at the sides of the scale by entering or leaving the scale. The modified scale was integrated in the SNE on the left ground area.

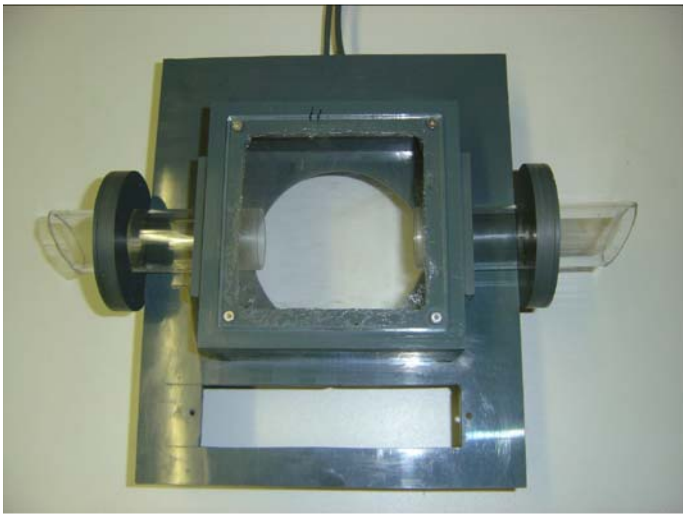


Figure 2: Plastic body of the scale with two RFID antennas

Data collection via optical tracking

Optical camera tracking was considered for tracking a single mouse to get continuous positional data compared to point data derived from the antennas. The camera (Logitech Quickcam Pro 5000) was placed above the highest floor in the SNE (see figure 3). This level can be accessed by just one tube thus a mouse that enters this level is recognized by the antenna. Additionally the antennas on the level allow to reassure the identity of optically tracked mouse in those cases when more than one subject is on the level.

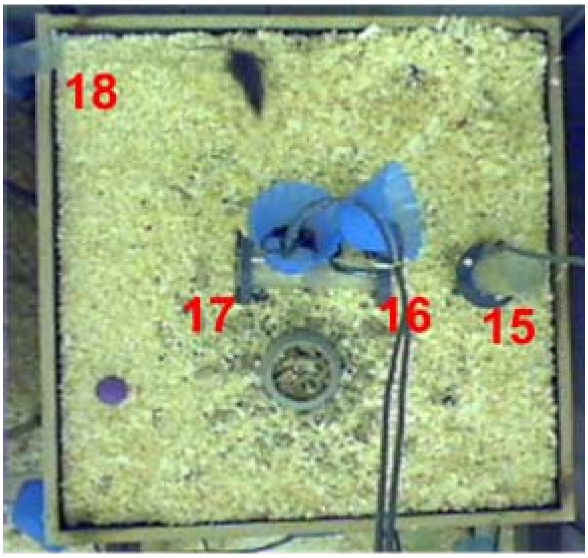


Figure 3: Camera tracking on the highest level, antennas 15 to 18 are labelled in red. Antennas on the level confirm the ID of the animals

4 Data analysis

The different data sources enable different analysis opportunities.

Anlaysis of the RFID data

For the post processing of the data a second software component (TOM) was written in the programming language C#. TOM is an ArcGIS extension and allows a visualization and analysis of the collected data. The attribute data for a mouse can be queried and viewed in a table.
The visualization module of TOM shows the position of the mice in the SNE at a certain point of time. To start the visualization of the movement a time interval has to

be chosen. In this interval the mice move from antenna to antenna. It is possible to select different display speeds and different play back rates. The spatial component is displayed in three dimensions, the temporal component is realized by the clock in the GUI [3]. Through the proposed solution in milliseconds an accurate representation of the data is possible: if during one second, a signal of the same mouse is registered at two antennas—when antennas are directly connected through a tube—a linear movement is displayed.

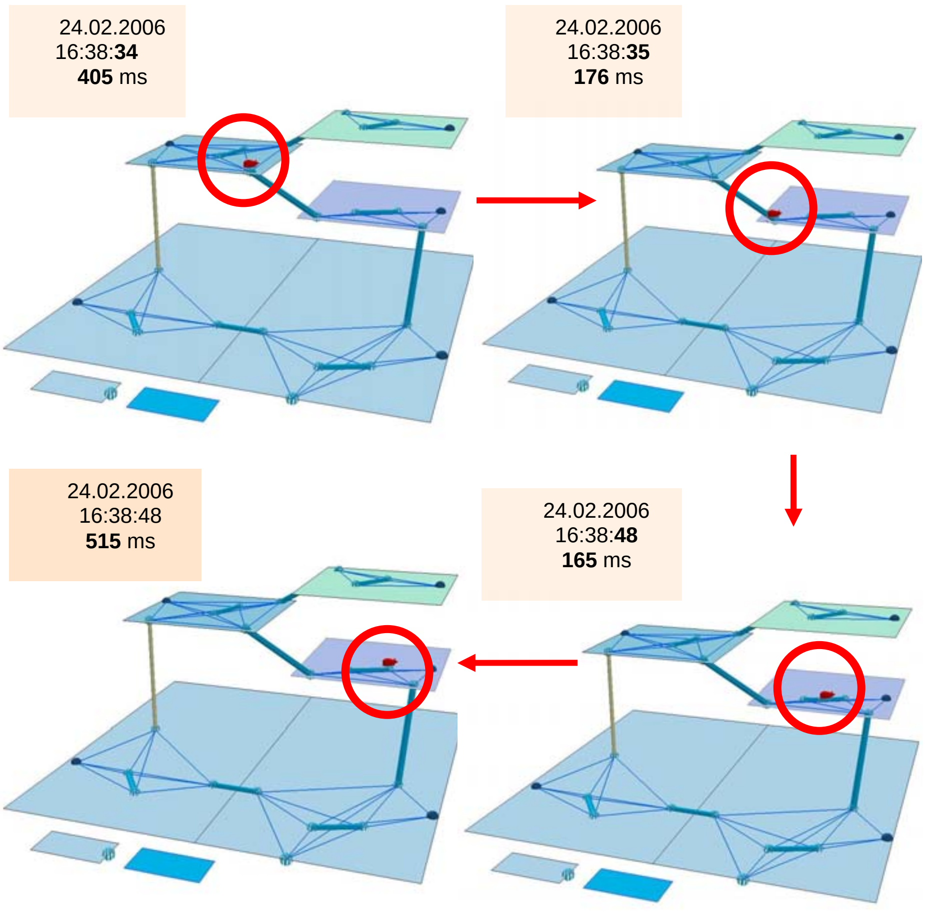


Figure 4 Sequence of mouse movement in time

In figure 4 the movement in time is shown. The advantage of this representation of the data is the movement of the mice can be observed for 24 hours. Especially the movement in the night phase – mice are nocturnal animals – can be observed. By observing the mice different patterns can be figured out. For example: Mice that trigger antennas only on a defined level can be considered as territorial. By analysing the movement patterns of all the individuals on that level dominance hierarchies might be obtained. This can be realized by analysing interaction patterns when two animals enter the tubes from different ends causing the subordinate to leave the tube backwards while the dominant one triggers both antennas. Additionally dominant mice are expected to trigger more antennas by patrolling their territory.

Besides the optical detection of patterns automated analysis of the data is prototypically implemented. On the one hand statistics can be calculated for the analysis per day, one the other hand statistics are available for the analysis per level (see table 2).

Table 2 Analysis functions

Analysis per day	Analysis per level
Time of last drinking	Number of antenna contacts
Time of last weighting	Duration of stay
Number of antenna contacts	Stay with other mice
Number of used levels	

Information about analysis per day and the number of antenna contacts per level are queried as SQL statements out of the database. Below the idea and functionality of the algorithm which calculates the duration of stay of one mouse per level is outlined (see listing 1):

The method "contactToOtherMiceOnLevel", which returns an array with Boolean values has two parameters: the selected mouse and the date for which the information is queried (line 1). In line 2 to 5 necessary attributes are declared and initialized: the array that will be returned as a result is initialized with the Boolean value 'false' for each level (line 2). The clustered data for one day (contacted antennas und timestamps) of the other unselected mice are stored in an array "miceWithoutSelected" (line 3). The number of the antenna where the selected mouse has the first contact on a particular date and the time of this contact are recognized (line 4 and 5). First the changes of levels for the selected mouse must be detected (lines 6-9). Therefore the antennas were associated with the corresponding levels before. Now the list "antennaTimeArray" (antenna and time data) of the selected mouse is scanned until an antenna is found that is not derived from the same level as "antennaStart" (line 8). If a change of level is found, we know the current level (line 9) and the time interval when the mouse was on this level (line 10). In line 13 it is verified whether an antenna entry of an unselected mouse for this time interval exists:

- If an antenna entry of an unselected mouse exists, all entries in the list "miceWithoutSelected" are checked if the antennas are from the same level as antennaStart (line 14):
    - o If the antennas are from the same level, the corresponding Boolean attribute for the level is set to true (line 15). This means the selected mouse had contact with another mouse on this date on this level.

o If the antennas are not from the same level, then the loop starts again: antennaStart gets the first antenna of the next level as new value and the next change of level will be detected (go to line 6).

- If no antenna entry of an unselected mouse within the time interval exists in line 13, the loop starts again, the start antenna gets the first antenna of the next level as new value and the next change of level will be detected (go to line 6).

The algorithm ends when the list "antennaTimeArray" is completely traversed and all possible level changes are detected.

```
// returns an array with Boolean values, which shows
whether a mouse had contact to other mice per level

1 contactToOtherMiceOnLevel bool[] (String mice,
                                String currentDate){
2 bool[]   levelsMeet=
                {false,false,false,false,false,false};
3 String[] miceWithoutSelected;   //array with all
                                     unselected mice
4 String   antennaStart;          //number of start
                                     antenna
5 Date     timeStart              //time of contact
                                     with antennaStart
6     for(i = 1; i < antennaTimeArray; i++){
7           String antenna = antennaTimeArray[i][0];
8           if(antennaStart and antenna not on the
                                          same level){
9                 findLevelWhereMouseIs();
10                DateTime time =
11                antennaTimeArray[i-1][1];
                                //last time on level
12                 if(two mice on one level){
13                      getTimeFromDB&Antennas-
                        ForUnselectedMice;
14                      if(unselectedMiceOnLevel-
                        WithMice){
15                            set levelsMeet
                              corresponding of true;
16                            timeStart = time;
17                            antennaStart = antenna;
18                      }
19                }
20          }
21          return levelsMeet;
22    }
23 }
```

Listing 1: Pseudo code of the method "contactToOtherMiceOnLevel"

Analysis of weight data

The analysis of weight data has to be scaled to the age of the animals in order to compare weight development of two or more individuals. Therefore the weights are ordered by the age of individuals (see figure 5).

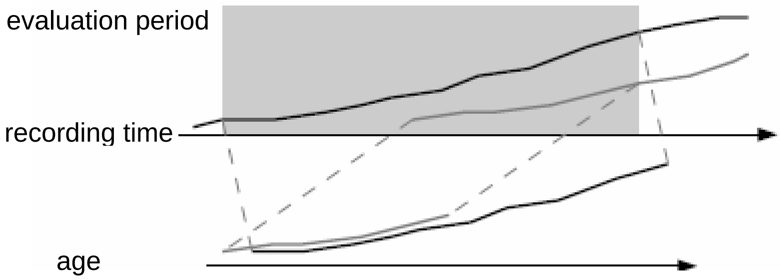


Figure 5: The weight data are independent of the recording time

Additionally movement patterns within the scale can be differentiated by means of the high frequency the scale sends data. The behaviour (see figure 6) of the mice can be summarized as follows:

1. A mouse moves fast by constant speed through the scale.
2. A mouse moves on the scale, remains there for a while and leaves the scale.
3. A mouse moves to the access of the scale and enters it step by step:
    - The mouse enters the scale completely.
    - The mouse returns and leaves the scale without entered it completely.

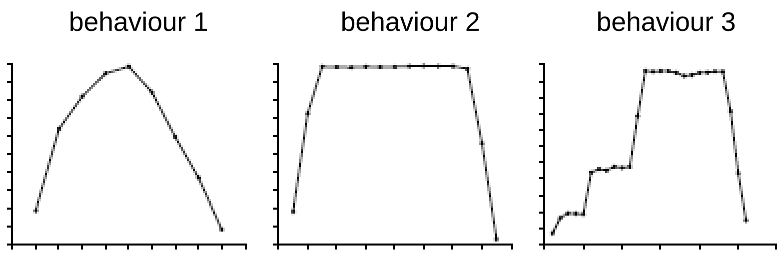


Figure 6: characteristic results of weight data, a mouse crosses the scale fast (behaviour 1), slow (behaviour 2) or step by step (behaviour 3)

Analysis of camera data

With the help of RFID data it is possible to identify each single mouse on the observed level. The positional data and the movement of the mice on the floor can be visualised as shown in figure 7.

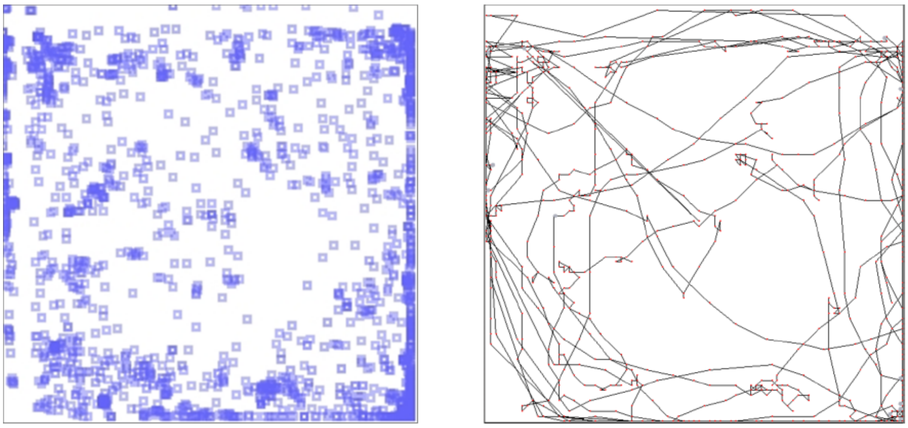


Figure 7: Representation on the positional data of a single mouse. Left: collected positional data as x-y-coordinates, right: tractorial representation

5 Further plans

We hold a huge amount of data and domain specific knowledge in this project. So we plan for our future work the use of data-mining methods for further analysis.
With these methods an automated classification of the mice between wildtype and TG should be derived. The bases for the decision are the different sensor sources in the SNE which provide data about behaviour and movement. To identify a mouse and to obtain behaviour partly a combination of the sensors is necessary (e.g. the RFID data is necessary to know which mice are filmed). We differentiate between simple behaviour and complex behaviour. Whereas complex behavior is combined of the simple behaviour of a mouse and might be related to behaviour of other mice (e.g. to obtain information about dominance behaviour the interaction of two or more mice must be observed). The analysis of the detected behaviour can then guide with a certain likelihood to the decision whether it is a wildtype or TG genotype (see figure 8). Most time will be spend on finding and defining a lot of simple and complex behaviours and movement patterns which guide us to the features. The qualitative premium features must be selected to build a reliable classificatory.
We consider for example the use of unsupervised and supervised classification to group the mice into TG and wildtype animals. For the unsupervised classification the cluster analysis will be used to group objects because of similarity. For a supervised

classification we consider to use naive Bayes classification, Bayses network or a decision tree which represents successive hierarchical decisions.

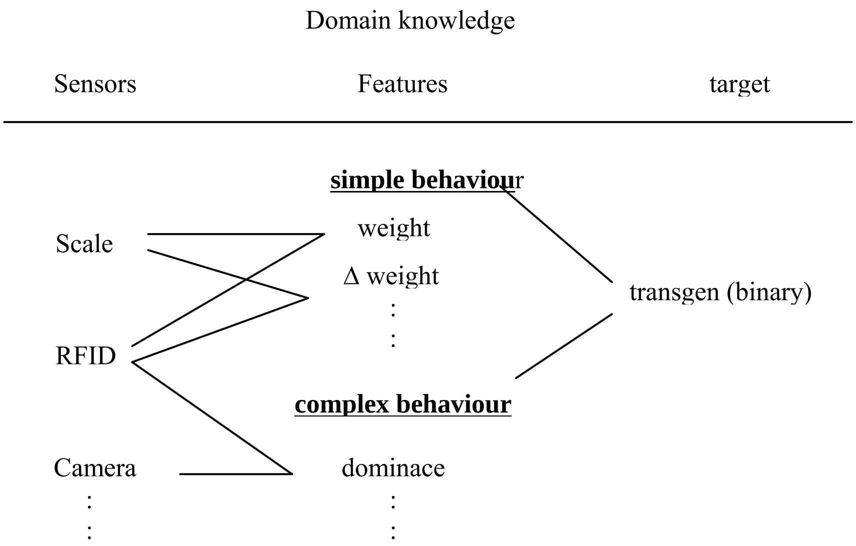


Figure 8: Features must be selected manually which can be measured by the sensors. We differentiate between simple behaviour or states (e.g. weight) and complex behaviour (e.g. dominance, combination of territorial and covered distance) which is combined by two or more simple behaviours

6 Conclusion

In this paper a system for collection and analysis of behavioural and movement data of laboratory mice is presented. The data were collected continuously with an indoor RFID tracking solution for laboratory mice, a scale and an optical tracking system in a SNE. Twenty-four hour per day, seven days a week observation is possible without disturbing the animals. Social interaction and the outward appearance are not influenced by this technology. The data stored in a relational database is analysed by an extended GIS framework. The movement of mice is visualized in a model of the SNE. Furthermore analysis functions, which offer information about the behaviour and movement of the mice per day and per level are implemented. We present initial analysis functions, which show that the collected data can support a continuous observation of the mice in a SNE.

References

1. Forgarty, J., Hudson, S., E., Atkeson, C., G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J., C., Yang, J. : Predictiong Human Interruptibility. ACM Transactions on Computer-Human Interaction, Vol. 12, No. 1 (2005).
2. Kritzler, M., Lewejohann, L., Krüger, A., Raubal, M., Sachser, N.: An RFID-based Tracking System for Laboratory Mice in a Semi Natural Environment. In Strang T, Cahill V, and Quigley A (eds), Pervasive 2006 Workshop Proceedings. Dublin, Ireland (2006)
3. Kritzler, M., Raubal, M, Krüger, A.: A GIS framework for spatio-temporal analysis and visualization of laboratory mice tracking data. Transactions in GIS. In press.
4. Neisen, G., Validation of a local positioning monitoring system (LPM) in relation to dairy cow behaviour. Second Annual Symposium of PhD-Program for Sustainable Agriculture ASPSA (2005)
5. Neisen, G., Wechsler, B., Gygax, L.: Automatic recording of social behaviour in dairy cows. Proceedings of the 40th International Congress of the ISAE 2006, 105 (2006)
6. Patterson, D.J., Liao, L., Gajos, K., Collier, M., Livic, N.: Opportunity Knocks: a System to Provide Cognitive Assistance with Transportation Services. International Conference on Ubiquitous Computing (UbiComp), Springer (2004)
7. Würbel, H. (2000) :Behaviour and the standardization fallacy. Nat Genet 26, 263.
8. Zancanaro, M., Kuflik, T., Boger, Z., Goren-Bar, D., Goldwasser, D.: Analyzing museums visitors' behaviour pattern. In Proceedings User Modeling. Springer (2007)

Monitoring Pedestrian Spatio-Temporal Behaviour

Alexandra Millonig[1] and Georg Gartner[1]

[1] Department of Geoinformation and Cartography, Vienna University of Technology, Erzherzog-Johann-Platz 1, A-1040 Vienna, Austria
millonig@cartography.tuwien.ac.at

Abstract. One of the major issues in the development of mobile pedestrian navigation services concerns the poor understanding of pedestrian spatio-temporal behaviour. Findings reveal that human route choice behaviour relies on a huge variety of influence factors. Therefore, common concepts like those used in car navigation systems will not conform to the requirements of pedestrians, as people on foot do not necessarily prefer the shortest path. This paper introduces an ongoing study focussing upon a multi-method approach towards the observation and interpretation of pedestrian walking patterns and route decision behaviour. The results will serve as a basis for the development of a typology of pedestrian spatio-temporal behaviour, which will allow the provision of customised navigational and environmental information in pedestrian navigation services.

Keywords:, Pedestrian spatio-temporal behaviour, Methodical triangulation, Tracking

1 Introduction

Within the last years, navigation systems providing information about optimal routes and additional location based information have become more and more popular. While on-board navigation systems are already routinely used in vehicle traffic, the development of mobile wayfinding tools providing reliable guiding instructions for pedestrians is now starting to arouse people's interest. Nevertheless, mobile navigation services have not yet the ability to fulfil the pedestrians' expectations. Various reasons are responsible for this fact. Route suggestions usually rely on road networks and do not meet the demand of pedestrians, as walking individuals have more freedom in movement compared to car drivers [1]. Common concepts used in navigation systems usually provide information concerning the "shortest path" or the "fastest path". Studies on human walking behaviour, however, indicate that pedestrians often prefer routes offering different qualities (e.g. "most beautiful", "most convenient") [2,3,4]. Although there are attempts to develop systems providing paths offering other qualities than shortness (e.g. the least risk of getting lost [5]), there are currently no approaches towards the development of pedestrian wayfinding systems providing tailor-made information to different kinds of users. Solely in the field of tourism research there are some efforts to offer information based on interest profiles [6].

The rapid development in the field of mobile information and communication technologies as well as the increasing amount of ubiquitously available information offer a wide range of possibilities to supply mobile users with location based information. Mobile tools for wayfinding combined with Location Based Services (LBS) can provide pedestrians with practical information concerning optimal routes and useful facilities in their vicinity. However, what is considered as "optimal" and "useful" largely varies between different kinds of individuals. Inappropriate information may hinder effective information extraction for a person seeking specific navigational and environmental information. A successful mobile spatial information service should therefore be based on a profound understanding of pedestrian spatio-temporal behaviour.

It can be assumed that the choice of a specific route and the actual walking behaviour depends on a variety of influence factors, like the task a user wants to perform, the present environment, or the individual preferences associated with personal attitudes and lifestyles. Generally, People are not aware of the factors underlying their spatio-temporal activities, and motion behaviour appears to occur in a somewhat automatic way. Methods used in monitoring pedestrian spatial behaviour are therefore facing several issues concerning the interpretation of pedestrian walking behaviour, as observations of the visible behaviour often fail to explain certain phenomena and inquiries may not be able to reveal reliable data. Thus, in an ongoing study we are combining several methods to thoroughly comprehend pedestrian motion behaviour. The results of the empirical study will serve as a platform for the development of a typology of lifestyle-based pedestrian mobility styles, which can be implemented in a wayfinding system in order to deliver customised information.

In this contribution, firstly, an overview about previous studies and commonly used methods in human spatial behaviour research is given, pointing out major advantages and drawbacks of each method. Secondly, the design of our approach towards the development of pedestrian mobility styles is introduced. Thirdly, the currently ongoing heuristic phase of the study is described and related preliminary results are presented.

2 Related Studies and Methods

Researchers focussing on human spatial behaviour have used a variety of different methods to register and assess the motion behaviour of pedestrians. Related studies are aiming at the investigation of different problems, such as tourism research, monitoring evacuation behaviour, tracking people for security reasons, planning guidelines, or the development of navigation and guiding systems [2,7,8].

First attempts to analyse pedestrian spatial behaviour in the 1960s mainly employed direct observations and questionnaires as usual methods of data collection [9]. Direct observations, also known as behavioural mapping or "tracking", have first been employed for studies concerning the movement behaviour of visitors of museums and exhibitions. Questionnaire survey techniques have primarily been used to collect data concerning pedestrian route choices, modal split and other transportation issues. In recent years, several technology-based methods have been developed to either track individual routes within a large (e.g. citywide) environment using digitally based localisation techniques [10,11,12], or to investigate microscopic walking patterns using video analysis [13,14].

All empirical techniques used in spatio-temporal behaviour research posses their advantages and drawbacks. Methods focussing upon the investigation and interpretation of visible behaviour fail to reveal motivations and intentions underlying pedestrian activities. Other techniques such as inquiries aim at the collection of data concerning route decisions and individual habits, motives, and intentions. However, as human behaviour is never fully determined by verbalised structures [15], the accuracy and validity of information gathered from questionnaires may suffer. Therefore, a combination of several complementary empirical techniques appears to be appropriate. In our current project an across-method triangulation of several qualitative and quantitative methods is applied. Before describing the details of our study, commonly used empirical techniques in pedestrian spatial behaviour research are briefly reviewed.

Questionnaire Surveys. Inquiries represent one of the most important data collecting techniques in transportation studies. They are relatively cheap and allow the collection and analysis of data taken from comparatively large samples. Inquiries are commonly used to gather information concerning route decision processes, individual habits, motives, and intentions. However, as spatio-temporal behaviour is mainly based on subliminal decisions, responses may be incorrect and constructed ex post. Moreover, it is known that people tend to adapt their answers – consciously or subconsciously – to what they expect to be socially desired behaviour [16]. Consequently, studies relying solely on results based on questionnaire data will have to accept a certain degree of inaccuracy [9].

Trip Diaries. Another frequently used method is the time-space budgets technique, including recall diaries, face-to-face interviews, and self-administered diaries [10,17]. Recall diaries and interviews are strongly dependant on the participant’s memory, which will result in a lesser degree of accuracy. Self-administered diaries are written in real-time and can therefore provide very detailed information. However, they demand considerable effort on the part of the subjects; consequently, only few people are willing to participate in these kinds of studies, and significant variation in the quality of the information must be expected.

Direct Observation (Tracking). Observations focus upon the investigation and interpretation of visible motion behaviour. Participatory observation techniques involve the observer taking part in the participant's activities, in order to identify the main purposes influencing the subject's decisions. Similar to inquiry methods, participants are aware of the fact that they are being under observation, and may tailor their behaviour to the researcher's expectations.
In non-participatory, unobtrusive observations the researcher follows the subject at a distance, recording her movements by drawing a line corresponding to the subject's activities on a map of the investigation field. Resolving the problem of "observer effects", this method provides detailed information about the "natural" behaviour of pedestrians [9,18]. Yet, this technique is very time-consuming and labour intensive, and findings are limited to the visible activities of pedestrians.

Video-based Analysis. Especially the development of agent-based simulation models uses video captured data for calibration and validation, in order to confirm the accuracy of simulated human behaviour [2,8,13]. Many studies using video-based techniques are conducted in laboratories, and are therefore limited to a very small observation field. There are also approaches observing a larger area by a network of several surveillance cameras [7]. Yet solely still visible behaviour can be investigated, leaving the subjects' intentions and motives as well as most other personal characteristics in the dark.

Localisation Technologies. In recent years, digitally based localisation technologies have been applied to track individuals in large environments. These include satellite-based technologies (Global Positioning System, GPS), land-based technologies (cell identification), or hybrid solutions [10]. Collecting localisation data with the help of tracking technologies can be of a rather invasive nature and quite cost-intensive, if the participants have to be equipped with tracking devices; therefore, observer effects may be suspected. The use of data gathered from private mobile phones without knowledge of their owners, on the other hand, may pose various ethical questions. Apart from that, the application of localisation techniques only allows to describe observable motion behaviour.

Several approaches try to minimise the limitations each method implies by combining two or more empirical techniques, for instance in the development of activity-based transportation models by collecting data with the help of GPS enhanced self-administered diaries recorded on PDAs [19], the combination of unobtrusive tracking methods and inquiries to analyse urban tourism [18], or the study of tourist behaviour using video and behavioural mapping techniques [20]. In our current study, we are combining tracking technologies, interviews, and localisation techniques to obtain a comprehensive insight into human spatio-temporal behaviour.

3 Multi-Method Approach to the Interpretation of Pedestrian Behaviour

According to the suggestions of several scientists in empirical research, we decided to combine qualitative and quantitative methods following the concept of "*across method*" triangulation [21,22]. The methods being used refer to different aspects of human spatial behaviour (e.g. observable patterns and interpretative investigation of motives and habits) and are to complement one another. Following the assumption that to a certain extent the individual behaviour of a person is influenced by the context a subject is acting within, we decided to observe pedestrians in a shopping environment in order to avoid the risk of investigating behaviour which is largely influenced by different contexts. The theory of *behaviour settings* [23] states that individual behaviour can be better explained by the current environment than by individual characteristics. When regarding the spatio-temporal behaviour of pedestrians, it may therefore be possible that behavioural differences are caused by the context a person is acting within (e.g. a tourist may behave different from a person on the way to her workplace). Hence, we decided to observe pedestrians in an environment where it can be assumed that the majority of people are acting within the same context – in this case a shopping environment.

The study includes two phases of empirical data collection combining observation and inquiry methods. The systematic integration of both qualitative-interpretative and quantitative-statistical methods is expected to result in a reciprocal fortification of the techniques and in a deeper understanding of pedestrian spatial behaviour. We aim at the identification of typical classes of spatio-temporal behaviour based on observed motion behaviour as well as lifestyle related attributes. Possible mobility types may for example include the "broadly interested flaneur" (low velocity, frequent turns, many stops at different kinds of facilities, various interests), or the "goal-oriented, efficient go-getter" (high velocity, shortest routes between stops, specific interests). The classes of spatio-temporal behaviour will be determined by extracted discriminative features from the qualitative and quantitative datasets. Those features can subsequently be used to assign a user to a mobility profile and provide customised information by an implemented wayfinding system.

The first phase of our study consists in a heuristic approach, aiming at the identification of a provisional pedestrian typology, which will be tested in the second, deductive phase of the study. Results of both empirical phases will then be consolidated and compared in order to delineate a model of pedestrian mobility styles, which will be used as basis for the description of mobility-style-based pedestrian profiles to be integrated in pedestrian navigation systems.

Fig. 1 illustrates the different steps leading to the development of a model of pedestrian mobility styles.

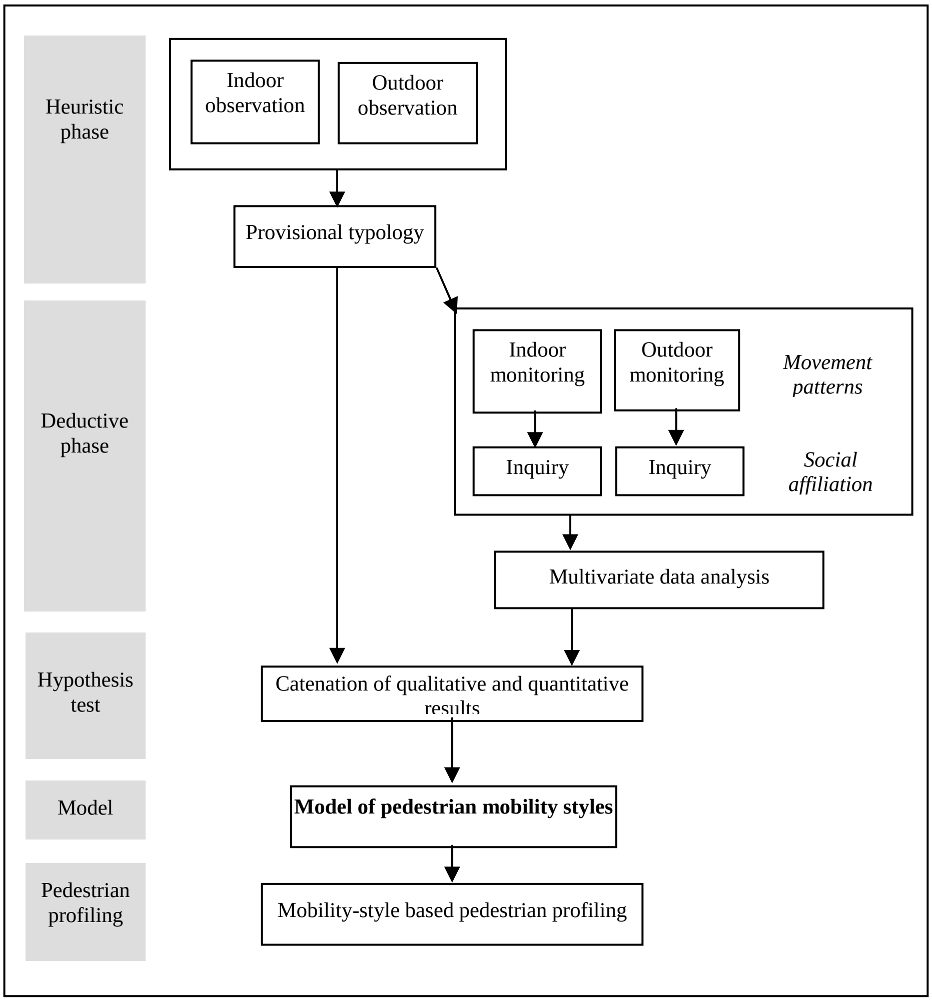


Fig. 1. Study design.

In our current study, we use the following empirical techniques to benefit from their specific strengths and minimise their disadvantages:

Unobtrusive Observation (non-participatory, unobtrusive, structured observation). This method allows the observation of the "natural", unswayed spatio-temporal behaviour of pedestrians. However, solely the visible behaviour can be recorded; intentions and motives cannot be unveiled.
Non-disguised Observation (non-participatory, non-disguised, structured observation). This allows continuous observation over a long period and can be combined with standardised interviews to obtain data from both the structural and the agent-centred perspective. Though, as the participants are aware of the observation, their

behaviour may be influenced (consciously or subconsciously) and differ from normal behaviour.
Inquiry (standardised and partially standardised interviews). Motivations underlying the activities can be revealed and self-assessments of individual motion patterns can be surveyed. Nevertheless, as individuals usually are not able to directly observe the cognitive processes concerning their walking patterns, they are oblivious to their spatio-temporal behaviour; responses may therefore be incorrect and constructed ex post.

The heuristic phase of the study contains observations of a non-participatory, unobtrusive type and standardised interviews. The motion behaviour of randomly selected pedestrians in an indoor as well as in an outdoor environment is mapped regarding route selection, turnarounds, velocity, stops, and duration of stops. A standardised inquiry following the observation part will provide data concerning socio-demographic factors, individual intentions and habits, and a self-assessment of the participants regarding their walking patterns. A detailed description of the currently ongoing tracking part can be found in section 4. The collected data will subsequently be analysed in order to inductively derive analytical classes by a coherent and systematic approach (constant comparison, cluster analysis). This leads to the development of provisional types of walking and route choice behaviour.

In the following deductive phase of the study, a non-participatory, non-disguised observation technique is employed. Pedestrians in indoor and outdoor environments are tracked by using technological localisation methods (indoor: Bluetooth; outdoor: GPS). The research conditions are diversified according to weekday, daytime, weather conditions, and time pressure. Participants are equipped with devices and their routes are tracked to continuously record the actual position, velocity, and moving direction. After the tracking process, detailed standardised Interviews are conducted to obtain information about their actual intentions, their attitudes, and lifestyle and socio-structural attributes. Results are used to verify the provisional types defined in the heuristic phase. The obtained data are related to specific mobility types, allowing their validation with regard to internal homogeneity and external heterogeneity.

Results of both empirical phases related to each other in order to identify a specific behavioural style for each provisional category. Finally, a model of pedestrian mobility styles will be developed, including descriptions of each type with respect to multiple aspects (basic parameters, behavioural characteristics, preferences, requirements, and main socio-demographic characteristics within the sample).

As the results of the survey are based on data collected of pedestrians acting within a specific context (shopping), the outcomes will be tested with regard to their validity in other context situations. Based on the final model of mobility styles group-specific routes and information in mobile navigation services can be offered to homogenously behaving target groups. The allocation of a user to a specific ideal type by inquiring the previously defined key attributes allows the consideration of specific preferences concerning route choice and navigational information.

4 Current Tracking Study

This section introduces the observation part of the primarily described heuristic phase. The aim of this empirical study is to observe, analyse, and interpret visible walking patterns of pedestrians in a shopping environment – a shopping street and a major shopping centre.

Recently the unobtrusive observations in the outdoor area and the indoor environment have been undertaken. The outdoor investigation field consist of two popular shopping streets in Vienna including the adjacent area. The total length of the two regarded streets amounts to approximately 2.5 km. Indoor observations have been made in a shopping centre in Vienna containing 180 retail shops and restaurants on a total area of 178 000 m² on two levels. The observations were of a direct, non-participatory, unobtrusive type, which means that the observer follows the target persons at a certain distance, recording the pattern of their activities over time and space.

Although this method is extremely time-consuming and labour-intensive, it is the only technique offering the possibility to yield a great amount of accurate information concerning the "natural", i.e. unaffected spatial behaviour of pedestrians in a large area. Other than in empirical methods using video or localisation techniques (e.g. cell-IDs from mobile phones), where individual-related data could be extracted from the stored datasets without knowledge of the observed people; this method arouses less ethical concerns. Most researchers agree that the observation of anonymous individuals in public areas will not cause major ethical problems [24].

4.1 Empirical Set-up

Mapping a participant's trajectory with conventional paper maps or detailed floor plans poses some difficulties, as the investigation field covers a rather large area. A map showing enough details to locate the target persons precisely would be difficult to handle, whereas a map of a smaller scale would diminish the accuracy of the recorded trajectories. Hence, a Java application has been developed in order to plot the individual routes on a digital map of the outdoor area (Source: Stadt Wien – ViennaGIS) and on a digital floor plan of the shopping centre.

Research instruments. The tracking tool was used on a tablet PC and provided data concerning the position and time of the trajectories drawn in the map during the observation process. Additionally, notes were taken concerning the visual attributes of the target person (gender, age, visual appearance), the observed stops, and the reason of termination for each observation. Additionally, a camera mobile phone was used to take pictures of the selected individuals, in order to form a rough estimate about the subject's socio-economic and lifestyle status.

Participants. Subjects were randomly selected unaccompanied individuals with a balanced gender ratio. Persons walking in company were exempted from observation to avoid influences on the individual behaviour. Other reasons for exclusion of specific persons occurred if a pedestrian was apparently following intentions other than shopping (e.g. police officers, mail carriers), if a person had been previously observed, or if the person was personally known to the observer.

Procedure. Observations were carried out under varying conditions (daytimes, weekdays, weather conditions). The observer placed herself at different randomly distributed points within the study site (intersections, underground-exits, bus stops in the outdoor area; entrances to the shopping centre). After a "clearing-period" of two minutes, a picture of the scene was taken and an unaccompanied individual was selected. The researcher then followed the subjects at a distance as long as possible and recorded the route in the map. Each point drawn in the map was recorded with respect to its specific point in time and its coordinates within the map. Stops and cases where subjects enter a shop or similar were marked in the map. Fig. 2 shows an example of a typical trajectory in the outdoor environment.

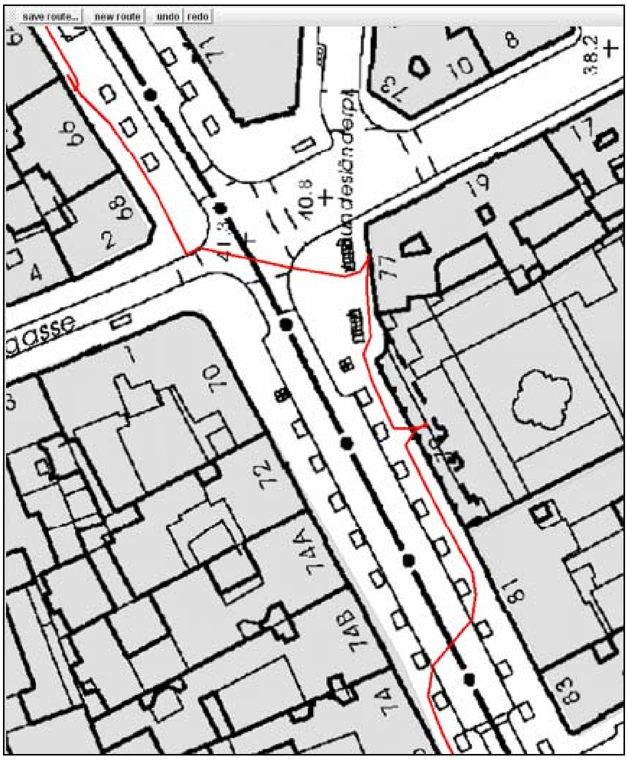


Fig. 2. Part of the field map containing a typical trajectory (map: Stadt Wien – ViennaGIS).

The observations had no predefined duration. To complete an observation, one of the following termination criteria had to occur: *(a)* The subject apparently notices the observation; *(b)* the subject leaves the study site; *(c)* the observer loses sight of the subject; *(d)* the subject enters a building (shop, café, etc.) and remains inside for more than 20 minutes; *(e)* the subject meets another person and they continue to walk together; or *(f)* the tablet PC battery is running low. If one of the termination criteria was met within two minutes after the observation had been initiated, the observer would forbear from saving the data due to the minor information content.

After completing an observation, additional notes are being taken concerning visible attributes of the target person (gender, age, appearance), detected stops (stops inside or outside a shop, category of visited locations), and cause of termination.

4.2 Analysis and Preliminary Results

In total, there have been 111 trajectories completed (57 outdoor, 53 indoor). About 60 further observations have been initiated, but had to be terminated within less than two minutes without saving utilisable data, as the observed subject turned out to be in company, the subject left the investigation area, or the researcher lost sight of the person.

Among the termination criteria, *(c)* and *(d)* turned out to be the most frequent causes of completing an observation. The regarded streets of the outdoor setting belong to the most popular shopping areas in Vienna; therefore the site is usually very crowded, which increases the risk of losing an observed individual. And, as the researcher usually does not follow a target person into a building, it may occur that the observer misses the moment when the person is leaving the building again. Hence, in some of the cases the observer terminated the observation after 20 minutes of waiting time, but had actually lost the target person. In the indoor environment it turned out to be easier not to lose sight of the target person. However, due to the cramped conditions within the shopping centre, it was more difficult to observe individuals without potentially being noticed. In the outdoor investigation field, the observed individuals have been tracked for an average of approximately 12 minutes, the longest tracking period lasting for about 62 minutes. Indoor observations had an average length of 16.5 minutes (maximum: 57 minutes).

The empirical set-up originally intended to combine the collected trajectories directly with inquiries following the observations. During the tracking procedure, however, it turned out to be difficult to realise this purpose. The main intention of the observation is to follow the subject as long as possible; therefore the observer tried to prolong the observation rather than terminating and interviewing the observed person. Hence, it was decided to carry out interviews after the observation period.

After collecting the trajectories and inquiry data, the data will be analysed according to the following attributes:

Motion data. Each point drawn in the map or the floor plan is recorded according to its exact date, time, and coordinates. A number of statistical computations are now being performed with respect to the following features:

Stops: Stops are detected according to different definitions of a stop (varying durations and radii; e.g. moving no more than 1, 3, or 5 metres during 10, 30, or 60 seconds). Fig. 3 shows an example of stops detected in two different trajectories. Shops and facilities, where stops had occurred, are categorised.

Velocity: Velocities between marked points are computed for each trajectory and categorised according to specific classes of velocity. Based on these results, a velocity profile will be created for each observed individual.

Turns: The frequency and characteristics of changes in direction will be analysed for each trajectory.

Visual appearance: Pictures that have been taken from the observed individuals are analysed with respect to lifestyle related feature classes.

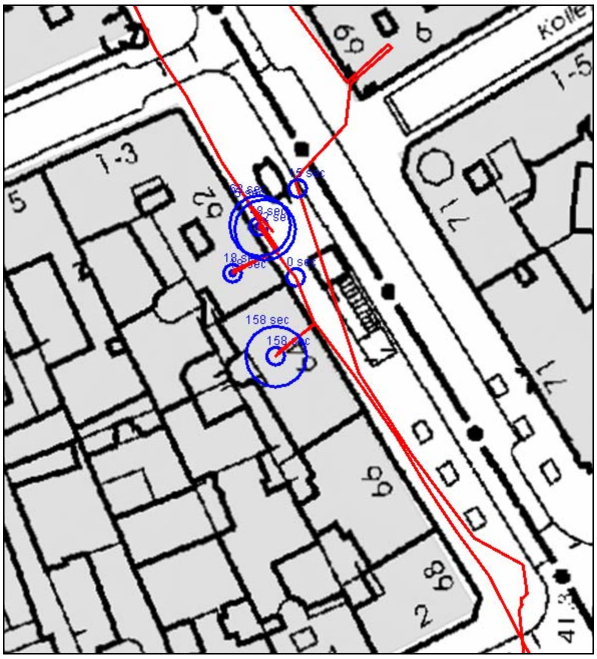


Fig. 3. Example of detected stops (map: Stadt Wien – ViennaGIS).

Inquiry. The inquiry contains of standardised questions concerning individual socio-demographic attributes, information concerning the current stay in the study site, frequency of visits, and questions referring to individual walking habits. The dataset will be analysed with statistical methods according to the following features:

Socio-demographic background: Gender, age, education and similar attributes will be evaluated statistically.

Familiarity with location: The frequency of visits, the vicinity to the place ot residence or workplace, and the reachability with different modes of transport will provide information concerning the individual's familiarity with the investigation area and basic mobility and shopping habits.

Mobility profile: Individuals are asked to provide a self-assessment with regard to a set of specific motion attributes (e.g. slow – fast, exploring – goal-oriented).

The collected data will then be analysed in order to inductively derive analytical classes by a coherent and systematic approach (constant comparison, cluster analysis). Pivotal attributes will be identified during the analysis. They will form the basis for a provisional typology of walking and route choice behaviour and will influence the research focus in the second phase of the project.

4 Summary

Research on pedestrian spatio-temporal behaviour has revealed that the complexity of pedestrian walking behaviour requires the combination of multiple methods to investigate and interpret the motion behaviour as well as the purposes underlying an individual's decisions and activities. Therefore, we are currently combining different observation methods, inquiries, and localisation technologies in order to obtain a comprehensive insight into pedestrian spatio-temporal behaviour. The combination of a number of complementary empirical techniques leads to the minimisation of method-related limitations and takes both internal and external factors influencing pedestrian behaviour into account.

The analytical process of the collected data aims at the identification and description of typical classes of pedestrian spatial behaviour. The determination of characteristic attributes for each class is to serve as a basis for the definition of pedestrian mobility and interest profiles in navigation systems. Based on the results of the project, future navigation applications for pedestrians will be able to classify a user according to pivotal characteristics identified in this study. Subsequently, a pedestrian can be provided with customised route information and location based services in ubiquitous environments.

Acknowledgments. This work is supported by the Austrian Funds for Scientific Research (FWF). The authors would like to thank M. Ray (arsenal research) for providing the tracking tool used in this survey. The digital map used in Fig.2 and 3 has been provided by Stadt Wien – ViennaGIS (www.wien.gv.at/viennagis/).

References

1. Corona, B., Winter, S.: Datasets for Pedestrian Navigation Services. In: Angewandte Geographische Informationsverarbeitung. In: Strobl, J., Blaschke, T., Griesebner, G. (eds.): Proc. of the AGIT Symposium, 2001, Salzburg, Austria (2001) 84-89

2. Helbing, D., Molnár, P., Farkas, I.J., Bolay, K.: Self-organizing pedestrian movement. Environment and Planning B: Planning and Design 2001 28 (2001) 361-383
3. Millonig A., Schechtner K.: Decision Loads and Route Qualities for Pedestrians – Key Requirements for the Design of Pedestrian Navigation Services. In: Waldau, N., Gattermann, P., Knoflacher, H., Schreckenberg, M. (eds.): Pedestrian and Evacuation Dynamics 2005. Springer Berlin Heidelberg (2007) 109-118
4. Thomas, C.: Zu Fuss einkaufen. Project report (2003) http://www.fussverkehr.ch/presse/zufuss_schlussbericht.pdf (accessed June 2007)
5. Grum, E.: Danger of getting lost: Optimize a path to minimize risk. Proceedings, CORP 2005, Vienna (2005)
6. Wiesenhofer, S., Feiertag, H., Ray, M., Paletta, L., Luley, P., Almer, A., Schardt, M., Ringert, J., Beyer, P.: Mobile City Explorer: An innovative GPS and Camera Phone Based Travel Assistant for City Tourists. Lecture Notes in Geoinformation and Cartography: Location Based Services and TeleCartography, Springer Berlin Heidelberg (2007) 557-573
7. Millonig, A., Schechtner, K.: Developing Landmark-Based Pedestrian Navigation Systems. IEEE Transactions on Intelligent Transportation Systems 8 (1) (2007) 43-49
8. O'Connor, A., Zerger, A., Itami, R.: Geo-Temporal Tracking and Analysis of Tourist Movement. Mathematics and Computers in Simulation 69 (2005) 135-150
9. Hill, M.: Stalking the Urban Pedestrian: A Comparison of Questionnaire and Tracking Methodologies for Behavioral Mapping in Large-Scale Environments. Environment and Behavior 16 (1984) 539-550
10. Shoval, N., Isaacson, M.: Tracking Tourists in the Digital Age. Annals of Tourism Research, 34 (1) (2007) 141–159
11. van der Spek, S.C.: Legible City – Walkable City – Liveable City: Observation of Walking Patterns in City Centres. Introductory paper, Urbanism On Track – Expert meeting on the application in urban design and planning of GPS-based and other tracking-based research, Delft, The Netherlands (2007)
12. Svetsuk, A.: Experiments in urban mobility analysis in Rome using mobile phone data. Position paper, Urbanism On Track – Expert meeting on the application in urban design and planning of GPS-based and other tracking-based research, Delft, The Netherlands (2007)
13. Daamen, W., Hoogendoorn, S.P.: Research on pedestrian traffic flows in the Netherlands, Proceedings Walk 21 IV. Portland, Oregon, United States: Walk 21 conference (2003) 101-117
14. O'Connor, A., Zerger, A., Itami, R.: Geo-Temporal Tracking and Analysis of Tourist Movement. Mathematics and Computers in Simulation 69 (2005) 135-150
15. Nisbett, R.E., Wilson, T.D.: Telling more than We can Know: Verbal Reports on Mental Processes. Psychological Review 84 (1977) 231-259
16. Esser, H.: Befragtenverhalten als „rationales Handeln" – Zur Erklärung von Antwortverzerrungen in Interviews. ZUMA-Arbeitsbericht Nr. 85/01 (1985)
17. Thornton, P., Williams, A., Shaw, W.G.: Revisiting Time-Space Diaries: An Exploratory Case Study of Tourist Behavior in Cornwall, England. Environment and Planning A 29 (1997) 1847–1867
18. Keul, A., Kühberger, A.: Tracking the Salzburg Tourist. Annals of Tourism Research 24 (1997) 1008-1012
19. Janssens, D., Hannes, E., Wets, G.: Planning interventions in the interactions between individual activity patterns, patterns of functions and infrastructure. Position paper, Urbanism On Track – Expert meeting on the application in urban design and planning of GPS-based and other tracking-based research, Delft, The Netherlands (2007)
20. Hartmann, R.: Combining Field Methods in Tourism Research. Annals of Tourism Research 15 (1988) 88-105

21. Fielding, N., Schreier, M.: Introduction: on the compatibility between qualitative and quantitative research methods. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research (On-line Journal), 2(1) (2001) http://www.qualitative-research.net/fqs-texte/1-01/1-01hrsg-e.htm (accessed June 2007).
22. Jakob, A.: Möglichkeiten und Grenzen der Triangulation quantitativer und qualitativer Daten am Beispiel der (Re-) Konstruktion einer Typologie erwerbsbiographischer Sicherheitskonzepte. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research (On-line Journal), 2(1) (2001) http://www.qualitative-research.net/fqs-texte/1-01/1-01jakob-d.htm (accessed June 2007)
23. Barker, R.G., Wright H.F.: Midwest and its Children: The Psychological Ecology of an American Town. Evanston, Illinois (1955)
24. Keul, A., Kühberger, A.: Die Strasse der Ameisen: Beobachtungen und Interviews zum Salzburger Städtetourismus. Profil-Verlag, München/Wien (1996)

Measurability, Representation and Interpretation of Spatial Usage in Knowledge-Sharing Environments – A Descriptive Model Based on WiFi Technologies

Teresa Heitor [1], Ana Tomé [1], Paulo Dimas [2], and João Pedro Silva [3]

[1] Instituto Superior Técnico – ICIST[1], Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
[2] IN+[2], Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
[3] YDreams PORTUGAL, Madan Parque - Quinta da Torre P-2829-516 Caparica
{teresa@civil.ist.utl.pt, anatome@civil.ist.utl.pt, pdimas@dem.ist.utl.pt, joao.silva@ydreams.com}

Abstract. The paper explores the potential of using available WiFi networks as an input for a space-use analysis model capable of describing - observing, recording and quantifying - and visualizing spatial usage and users' spatial behaviours in knowledge-sharing scenarios and correlating this information to spatial structure. Knowledge-sharing scenarios are defined as physical locations where people go for acquiring, transmitting and producing knowledge carrying mobile devices functioning as location probes. The proposed model is based on a crossover of Space Syntax and Spatial Information Visualization. Emerging spatial patterns of knowledge-sharing are identified by combining spatial description with spatial information visualization. The paper considers the representation of inputs acquired by WiFi network of Instituto Superior Técnico campus using FLUX* visualization platform; the comparison between patterns of spatial configuration and user mobility and the discussion of the proposed space-use analysis model potential.

Keywords: wireless sensors, WiFi, FLUX*, information visualization, space-use analysis model, knowledge-sharing patterns.

1 Introduction

In the paper we developed a space-use analysis model capable of describing – observing, recording and quantifying – spatial usage and users' spatial behaviours in knowledge-sharing scenarios and correlating this information to spatial structure. Knowledge-sharing scenarios are defined as physical locations where people go for acquiring, transmitting and producing knowledge carrying mobile devices functioning as location probes.

Space-use analysis is about techniques that objectively describe environments and relate this description to specific problems of use. It is about mapping environments

[1] Instituto de Engenharia de Estruturas, Território e Construção.
[2] Center for Innovation, Technology and Policy Reseach.

and users' spatial behaviours and exploring their relationship. In recent decades various much research effort has gone into defining spatial descriptive models, which explore patterns of co-visibility and co-accessibility and can be quantified and correlated with functional/behavioural data. These included Isovists [1] and Space Syntax [2], further complemented by space partitioning schemes [3] and visibility graph analysis [4].

Space Syntax provides a valuable framework for space-use analysis. It has been applied to different types of knowledge-sharing environments such as university campuses [5], research labs [6,7], design studios [8], working spaces [9,10] or exhibition settings [11,12,13]. These studies above all focus on configurational properties and suggest that social and informational interaction are influenced by how a space is defined as well as by how that space relates to and is integrated with other spaces. However, users' spatial behaviours are established and assessed through direct observations or questionnaires. These procedures, besides being intrusive, also depend on the human factor. Hence they do not allow for the collection of unbiased data, regardless of observer accuracy.

As far as knowledge-sharing scenarios are concerned, we are currently witnessing a great increase in the use of mobile technologies within the university community, including laptops, PDA's and other WiFi devices. Due to the existing infrastructure and its increasing widespread use, the use of wireless antennas behaving like ad-hoc sensor networks as an input for space-use analysis seems to be a plausible one since it allows for recognition of user location. The research question refers to such technology's potential to identify spatial patterns of knowledge-sharing and to further explore how knowledge-sharing scenarios may become "interactive learning devices" [14].

In a recent research project carried out at the MIT, log files from MIT wireless network (2300 Access Points) were used [15]. The aim was to monitor and collect extensive data on on-campus WiFi usage, in order to understand the emerging daily working patterns of the academic community in real time and re-evaluate the qualities of physical space supporting them. This study provides new insight into space-use analysis at the urban level. It works as a real-time mapping exercise, permitting a dynamic view capable of acquiring different layers of real-time information through simple cartographic evidence.

The paper investigates the capacity of WiFi networks to give space-use inputs at the building level. Large-scale applications are more compatible with the identification of Access Points` ranges. At the building level that correlation is not so obvious since built elements introduce great irregularities in wireless coverage of the signal distribution. Subsequently, it was necessary to develop a methodology to determine that coverage and correlate the extension of Access Points` signals with the physical space.

The proposed space-use analysis model explores relationships between the virtual web space (user communication in a more or less ubiquitous field) and the physical space (users movements in a more or less permeable system) in order to identify patterns of knowledge-sharing. It is based on a crossover of Space Syntax and Spatial Information Visualization.

Spatial Information Visualization is suitable for producing emerging social and mobility patterns and correlating this information to spatial properties. Geographic

Information Systems have been extensively used to input, store, query, transform and visualize spatial data. Recent approaches to Spatial Information Visualization have reduced the data operations available to users by separating visualization from data transformation and storage. Examples of this are Google Earth and Microsoft Virtual Earth, the former being closer to the definition of a spatial Virtual Environment [16] due to the 3D nature of the background information. Such environments have an advantage over traditional cartographic maps in exploration and analysis tasks because they do not require thorough understanding of symbols, conventions and formalisms. By relying on an open specification and being relatively user-friendly, Google Earth is now becoming de facto a geographic visualization platform.

The Civil Engineering and Architecture department building on the Instituto Superior Técnico (IST) campus in Lisbon was selected as a case study to test the space-use analysis model. The WiFi network allowing Internet access covers the whole IST campus. All department buildings including learning and social spaces are now fully covered. Information Technology (IT) is used extensively for information, communication, collaboration and socializing by the IST community. In preliminary analysis it was identified that the Civil Engineering and Architecture department building presents a considerable activity in terms of WiFi network use. Moreover it receives many students from others departments demanding spaces for individual and collective work and socializing. The resident population that uses WiFi is, essentially, students. Staff use, predominantly but not exclusively, the fixed network installed in their own work spaces despite the increasing use of the wireless network. For the purpose of this paper, the reference population is the student population, since it represents a larger group with greater impact in the use of wireless network.

The IST campus has 158 Access Points, with the Architecture department building having 18. The visualization platform FLUX* [17,18] allows one to register device traces in Access Points. Traces can be used to obtain two kinds of information: locations (how many devices are in a specific antenna, in a specific moment) and flows (when a device is subsequently detected by two, or more, antennas in a given period).

FLUX* started registering traces in January 2006 but not in a continuous way. This work considers a period of intensive student usage (classes and evaluation period). During this period, FLUX* registered a total of 553,759 traces in the Architecture department building and a total of 2,370,283 traces in the whole IST campus. The total number of devices registered in the building was 2,683, while in the campus it was 5,879. The time span granularity between each registration in an Access Point is 5 minutes.

The Architecture department building population is about five thousand. This means that the number of devices registered by FLUX* inside the building is a sample superior to 50%. In relation to the entire campus (about eight thousand) the sample decreases to 32%.

The paper considers four parts. The first one describes the methodological procedures applied to develop the model. The second refers the main tasks carried out. In the third part, inputs acquired by the IST campus WiFi network (user locations) are represented using the FLUX* visualization platform. The mobility patterns, emerging from the ubiquitous network access, are analysed. Patterns of spatial configuration and patterns of user mobility are compared. Correlations between space and mobility are established. Reflections about the potential of FLUX* visualizations in what concerns to the representation of spatial behavior are made. Finally the potential of the proposed space-use analysis model is discussed.

2 Methodological Procedures

Space-use analysis model development considers several stages. Spatial Information Visualization and Space Syntax procedures are carried out separately (Fig. 1). Spatial Information Visualization entails six stages: 1) definition of an Access Points taxonomy; 2) production of an exploratory trace; 3) identification of the spaces covered by each Access Point; 4) identification of overlapping positions between each Access Point spatial range; 5) production of an ad-hoc mobility matrix; 6) definition of queries. Queries are filters applicable to FLUX* database. Through these filters, it was possible to extract useful information concerning space-use analysis model goals. The information is mapped using different visualization models: force-directed graphs [19], treemaps [20] and 3D representations based on Google Earth application.

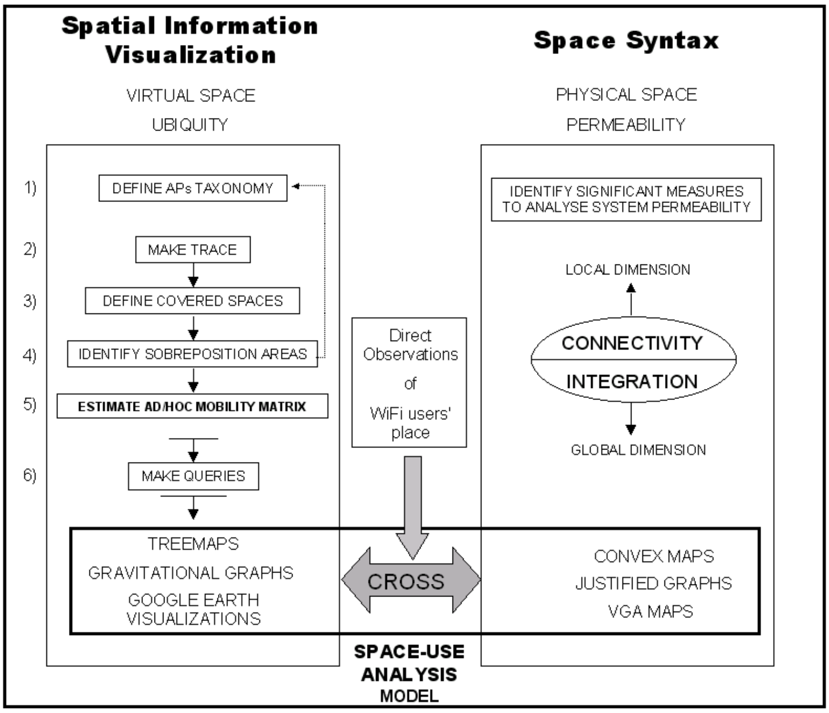


Fig. 1. Methodology diagram

In the proposed space-use analysis context, force-directed graphs will help to identify, syntactically, central nodes in the space. Treemaps will be used to visualize variables like people's occupation of space and to compare patterns between different types of knowledge-sharing of the same place. The Google Earth geographic visualization platform will be used to produce 3D visualizations of mobility flows in Architecture department building and IST campus.

Procedures 2), 3), 4) are sequential. The goal is to develop the ad-hoc mobility matrix, an important tool for FLUX* detection of flows between Access Points. Procedure 4) allows readjusting the Access Point taxonomy. Procedures 1) and 5) are the basic ones for making queries - the interfaces between FLUX* database and FLUX* visualizations.

Space Syntax procedures are based on the identification of the basic configuration properties allowing the analysis of spatial system permeability. In the virtual space of the Internet, ubiquity is the corresponding property of permeability. Encounters in space and wireless communications through net interact to make possible the production of knowledge. Hence, the main task is to analyse spatial system connectivity (permeability at a local level) and integration (permeability at a global level).

Space Syntax methodology and analytical tools were applied to investigate the configurational properties of architectural space for correlation with socio-functional implications and the emergent communication patterns (Spatial Information Visualization). As such, visibility graph analysis (VGA) was used, through "Depth Map"[3], convex maps and justified graphs. The measures applied consider the visual integration and connectivity. Depth Map enables analysis of the spatial structure of the learning setting and correlation to spatial usage and user movements by exploring patterns of co-visibility and co-accessibility. It was also attempted to interpret the identifiable relations established with the Access Points` locations and their influence over the system space. Convex maps and justified graphs [2] were adopted to understand, at a local level, the direct permeability of the structuring of the spatial layout and its implications.

Space-use analysis model crosses Architecture department building space patterns with WiFi network wireless communication patterns. Through that intersection, it evaluates relationships between social dynamics and the built space and identifies spatial patterns of knowledge-sharing.

Parallel to this, direct observations were made to WiFi users inside the Architecture department building. The aim was to map their distribution in space and time. This procedure has allowed to validate FLUX* results.

---

[3] "Depth Map" was developed at University College London. It consists on a class of tools for spatial description – analysis, interpretation and evaluation – of the spatial configuration of built environments, incorporating Benedict's pioneering work on Isovists (1979) and other models of the description of built space developed by researchers on space syntax. The visibility graphs comprises the breaking up of space into a grid of points which is then analysed on the basis of how many points can see how many other points providing spatial measures capable of explore patterns of co-visibility and co-accessibility.

3 Development

3.1 Labeling the Access Points

In FLUX*, to extract information from the recorded database, one needs to define categories for the Access Points. Categories are tags that can be applied to wireless antennas. Each antenna can have multiple tags.

In this case, seven tags were defined. One tag considered antennas identification through its Internet Protocol number: IP. Three tags concerned building space: “floor” (vertical position of Access Point); “location” (space where Access Point is installed); and “covered spaces” (spaces where the Access Point signal is detected). The three other tags concerned building use: “occupation” (department that occupies a space); “sectors” (departmental group that occupies a space) and “knowledge-sharing types” (Fig. 2). This last category is based on the typology defined by Scott-Webber [21] for knowledge-sharing: a) delivering knowledge; b) applying knowledge; c) creating knowledge; d) communicating knowledge; e) using knowledge for decision making.

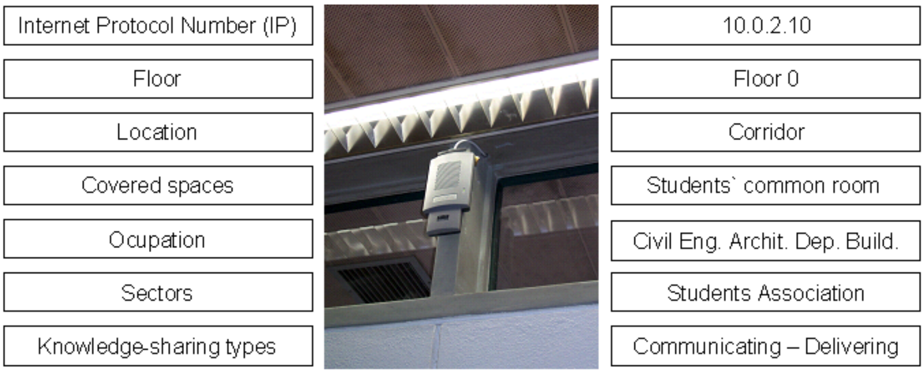


Fig. 2. Access Point 1 in ground floor – tags

These scenarios where adapted to the Architecture department building in order to understand the kind of knowledge flows supported by the physical space. Each typology describes an environment where knowledge is sharing in a specific way. Those knowledge environments implicate particular behavioural premises, layouts and protocol attributes.

Synthetically, we may describe each one in the following way: “delivering” describes places where information is transmitted in a formal method so that others can learn (classrooms, auditoriums); “applying” describes places where organizations puts knowledge into practice (labs); “creating” describes places where organizations produce and implement new ideas (researchers` and teachers` rooms); “communicating” addresses places where people go to exchanging information in an informal way, verbally and non verbally (atriums and others circulations spaces, cafeteria, students` rooms); places where knowledge is used for “decision making” refers to environments where information is distilled and judgments are made and acted upon (teachers` and administration rooms). The correspondence between some current spaces uses and archetypal layouts is indicated in Fig. 3. Some spaces, like teachers' working rooms belong to hybrid categories (Decision making/Creating).

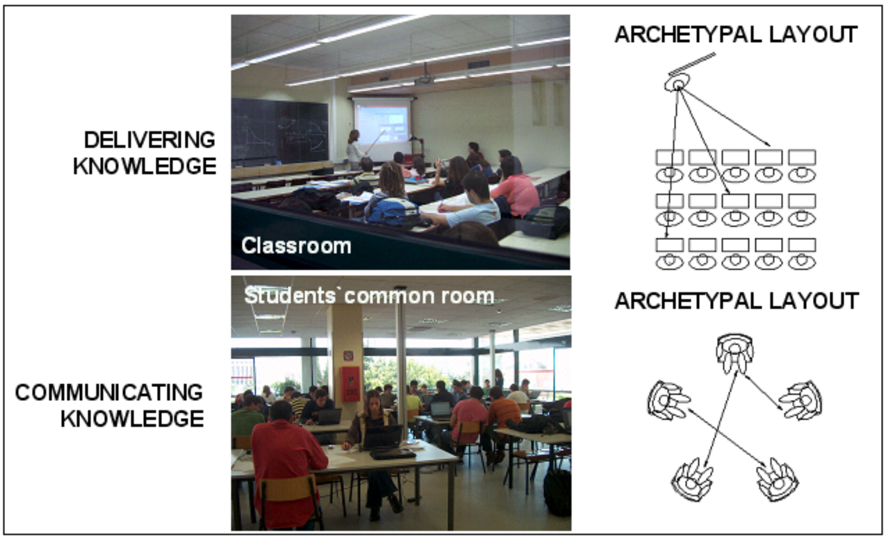


Fig. 3. Correspondence between some current spaces uses and knowledge-sharing types` layouts

The vertical distribution of the knowledge-sharing types is the following one: at ground floor level, the main type is "Communicating" which take place in atriums, cafeteria, students` common room and several classrooms open twenty-four hours for students learning. At floors 1 and 01, locate "Delivering" knowledge-sharing type. Associated spaces with this category are classrooms and auditoriums. "Decision making", "Applying" and "Creating" knowledge-sharing types are located in deeper levels. "Decision making" is related with teachers` rooms and administration rooms (floors 2 and 3). "Applying" is related with Labs (floor 02). Spaces identified with this "Creating" are: researchers and teachers rooms (upper floors) and experimental labs (lower floors).

3.2 Tracing the signal

The signal propagation model looks like a donut one hundred meters in diameter in open space. This concept, in real situations, is distorted by the built barriers in buildings (walls, ceilings, floors) and by equipment in rooms. Hence, to establish relationships between the spatial system and the wireless communication system it was necessary to determine the range of each Access Point in space.

To this end, exploratory traces were generated by walking around the building with a WiFi device. On this walkthrough, the device registered the Basic Service Set Identifier (BSSID) of each Access Point mapping the extension of each Access Point coverage area, frontiers between Access Points, variation of signal quality and “dead zones”. The BSSID is a unique identifier that acts like a name for a particular network adapter. This unique number allowed us to identify, in every moment, which antenna was captured by the exploratory device.

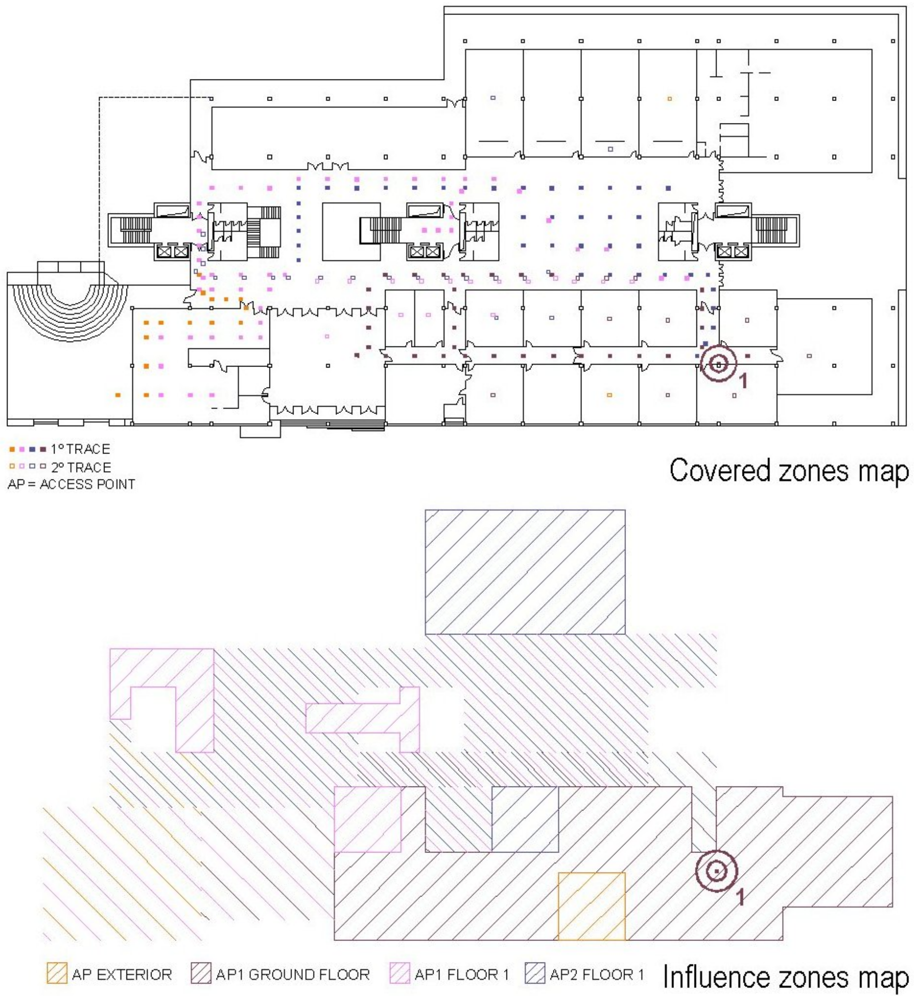


Fig. 4. Civil Engineering and Architecture Building (ground floor): Access Points` maps

After these exploratory traces, the zones covered by each Access Point were identified. In open spaces (e.g. central atrium on the ground floor) several Access

Points were detected, making it difficult to establish clearly defined frontiers between Access Points. In many cases, as expected, their coverage areas go beyond the built barriers and don't match with space geometry.

The maps of spaces covered by each Access Point resulted in diagrams showing their coverage area extensions (Fig. 4). Sometimes the coverage areas have overlapped zones (zones simultaneously covered by more then one Access Point) in the same floor and between floors.

Analysis of the diagrams revealed great irregularity concerning the spatial distribution of the Access Points` coverage areas. This was particularly evident in floors 0 and 2 because of the interference of Access Points 1 and 2 (floor 1) through main atrium. This irregularity was accentuated by the action of external Access Points. Indeed, rooms near the west, east and south façades (especially on floors 1 and 2) are covered by such Access Points.

3.3 Defining mobility

In terms of FLUX*, mobility refers to users' movements between Access Points` coverage zones during their daily activities. Hence, mobility is given when people take their WiFi devices and carry them to another space/Access Point in reasonable time intervals. When a device is registered in an Access Point on a certain day, and is registered, in another Access Point, in the next day or the following week, this is not considered mobility. For defining mobility patterns two factors must be considered: movement in space and the time span.

The ad-hoc mobility matrix was an important tool for generating mobility visualizations between Access Points (Fig. 5). The matrix sets, in an ad-hoc way, the mobility probability between Access Points. If a device is captured inside an overlapped zone it means that mobility between Access Points might not have taken place. Zones with overlapping Access Points` ranges introduce a degree of uncertainty as to mobility that must be estimated.

The probability of mobility among Access Points was estimated by considering the coverage area of each Access Point. The dimensions of these areas are always different. This means that overlapped areas don‘t have unique mobility values. Also, the certainty that mobility took place depends on the movement direction: mobility from Access Point 1 to Access Point 2 is different from Access Point 2 to Access Point 1.

The asymmetry of the matrix is a property resulting from the irregular distribution on space of Access Points` coverage areas. The lower probability (<1) is concentrated along the table diagonal (on each floor). That is particularly evident on floor 1 (all Access Points fields cross with each other). Values <1 expand from floor 1 (Access Points 1 and 2) in all table directions. It seems to be related to configuration of main atrium as a nine meters height open space.

Time threshold was defined, too. The maximum value considered was 24 hours. Beyond a day, mobility is not considered. The minor is the time interval, the greater is the probability that mobility happened. So, the smallest interval considered was 5 minutes.

FLOOR		f3			f2		f1				f0	f01				f02			
	AP	1	2	3	1	2	1	2	3	4	1	1	2	3	4	1	2	3	4
f3	1		1	0,95	1	1	1	0,99	1	1	1	1	1	1	1	1	1	1	1
	2	1		0,87	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	3	0,93	0,86		1	1	0,98	1	1	1	1	1	1	1	1	1	1	1	1
f2	1	1	1	1		1	1	0,89	1	1	1	1	1	1	1	1	1	1	1
	2	1	1	1	1		1	0,97	1	1	1	1	1	1	1	1	1	1	1
f1	1	1	1	0,99	1	1		0,59	0,97	0,97	0,9	1	1	1	1	1	1	1	1
	2	0,99	1	1	0,95	0,98	0,43		0,93	0,97	0,92	1	1	1	1	1	1	1	1
	3	1	1	1	1	1	0,92	0,92		0,92	1	1	1	1	1	1	1	1	1
	4	1	1	1	1	1	0,9	0,93	0,89		1	1	1	1	1	1	1	1	1
f0	1	1	1	1	1	1	0,73	0,85	1	1		1	1	1	1	1	1	1	1
f01	1	1	1	1	1	1	1	1	1	1	1		0,78	0,86	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1	0,8		0,95	1	1	1	1	1
	3	1	1	1	1	1	1	1	1	1	1	0,95	0,98		0,92	1	1	0,96	0,99
	4	1	1	1	1	1	1	1	1	1	1	1	1	0,9		1	1	1	1
f02	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		0	1	1
	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0		1	1
	3	1	1	1	1	1	1	1	1	1	1	1	1	0,94	1	1	1		1
	4	1	1	1	1	1	1	1	1	1	1	1	1	0,97	1	1	1	1	

AP = Access Point

Ad-hoc mobility matrix.

Mobility traces scatter map.

Each square represents a flow of a device moving from one AP to another. Upper left flows have a greater probability of representing mobility in the same user context (small time span).

The majority of flows have a high probability of mobility because the coefficient is close or equal to 1. Visualizations used flows from the upper left corner.

Fig. 5. Ad-hoc mobility matrix and Mobility traces scatter map

3.4 Inquiring FLUX*

Queries are means of extracting meaningful information from the FLUX* database. The answers provide input for FLUX* visualization tools. Two main groups of queries were made: location queries (traces) and mobility queries (flows). Location queries were divided into three subgroups: general queries, category queries and time queries. Mobility queries consider the number of devices moving between inbound and outbound, between each knowledge-sharing type and between each department sector. Queries supply quantitative information complementary to the visualizations models used (GE representations, force-directed graphs and treemaps).

3.5 Detecting relationships between permeability and (conditioned) ubiquity

Wireless antennas are mainly located in circulation spaces: atriums and corridors with few exceptions on floor 02, Access Points 1 and 2 (auditorium) and Access Point 4

(Geotechnics Lab). Those distribution zones are, naturally, closely connected to the others spaces in the system.

Analysing relationships between Access Points` location and configurational properties, it is possible to conclude that on floors 0, 1 and 01, Access Points are positioned in more narrow spaces with a high number of direct adjacencies (Fig. 6). On other levels, particularly floor 2, Access Points are located in deeper zones seeking to cover specific sub-areas of the spatial system.

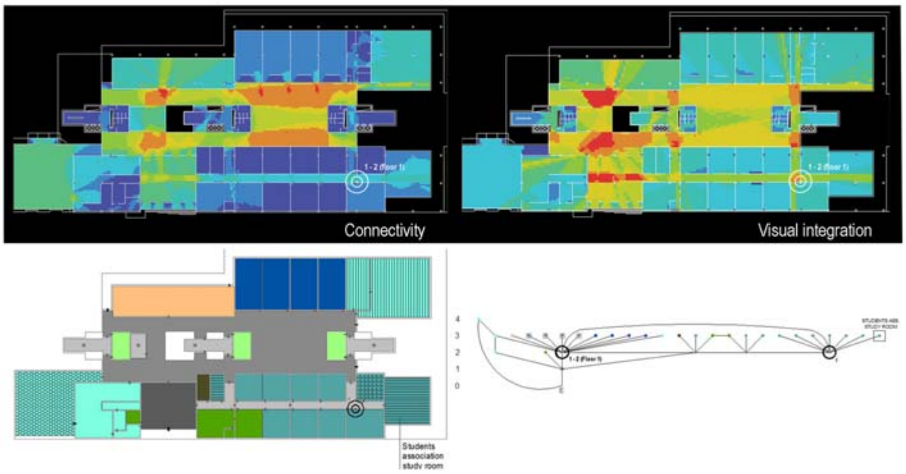


Fig. 6. Civil Engineering and Architecture Building (ground floor): syntactic models

Another strategic Access Points` localization detected was in nodes that belong to cycles or nodes that gather cycles. Access Point 1, on the ground floor, belongs to a cycle that also includes main atrium. This node is within the range of Access Points 1 and 2 (floor 1). This cycle is linked to another one where the entrance node is located. Analysis of VGA maps confirms, generally, Access Points` location in more connected/integrated spaces in spatial system: atriums and corridors. In fact, overlapping zones are given, particularly, in those spaces.

Main atrium - the great circulation space - stands out for its high spatial connectivity/integration. A comparison with the Access Points` coverage zones shows that this space supports the most mixed and extensive of all overlapped zones. A comparison of Figs. 4 and 6 illustrates this.

A direct relationship was detected between spatial permeability and the concern with maximizing the wireless signal propagation. Where ubiquity is highly conditioned by the built barriers, the logic for Access Points` installation was to seek spaces that were highly connected or integrated in the spatial system. Coverage flaws detected in some rooms (floors 1, 2 and 3) are due to scarcity of Access Points, and not mistakes in locating the existing ones.

4 Results

Visualizations (V) allowed to understand users' spatial behaviours and to identify spatial patterns of knowledge-sharing. Some data filters (indicated in figures) were applied for obtaining emerging visual patterns.

Analysis of the global wireless relationships inside IST campus shows that the Architecture department building has the most wireless activity of all campus buildings (Fig. 7 – V1). The greater number of mobility flows (inbound+outbound) occur between the Architecture department building, the main building and Mechanics and Computer Science department buildings (Fig. 7 – V2). Inside the Architecture department building, most flows are linked to Computer Science Lab, students` common room and main atrium (Fig. 7 – V3). Students from other departments intensively use these spaces to work, study and socialize. These spaces correspond to articulation nuclei (external/internal) in terms of "Communicating" knowledge type. They are located in the shallowest places and relate to each other through highly integrated spaces.

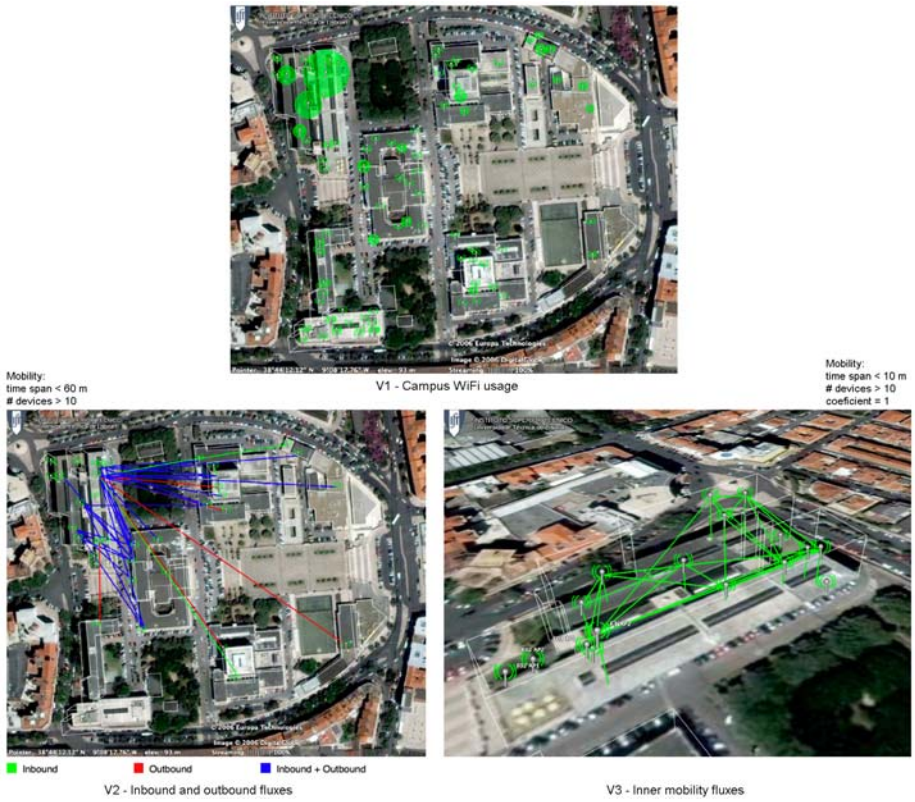


Fig. 7. FLUX* mobility visualizations (V1; V2; V3)

A look at the campus syntactic model shows that the Architecture department building is close to the integrated nucleus, where most of the movement and campus life can be observed (Fig. 8). It belongs to a cycle linked to one of the entrances and an axis that passes through the main entrance campus. Moreover, the most globally segregated lines show a high degree of local integration. Therefore, no area is left without a natural flow of people. This spatial condition allows for generation of movement through the Architecture department building and supports wireless mobility flows. Within the campus, the most frequent paths surround the Architecture department building, making the building recognizable to users. Nevertheless, as the Architecture department building is not on the main axial line, it maintains a certain degree of reserve with respect to campus.

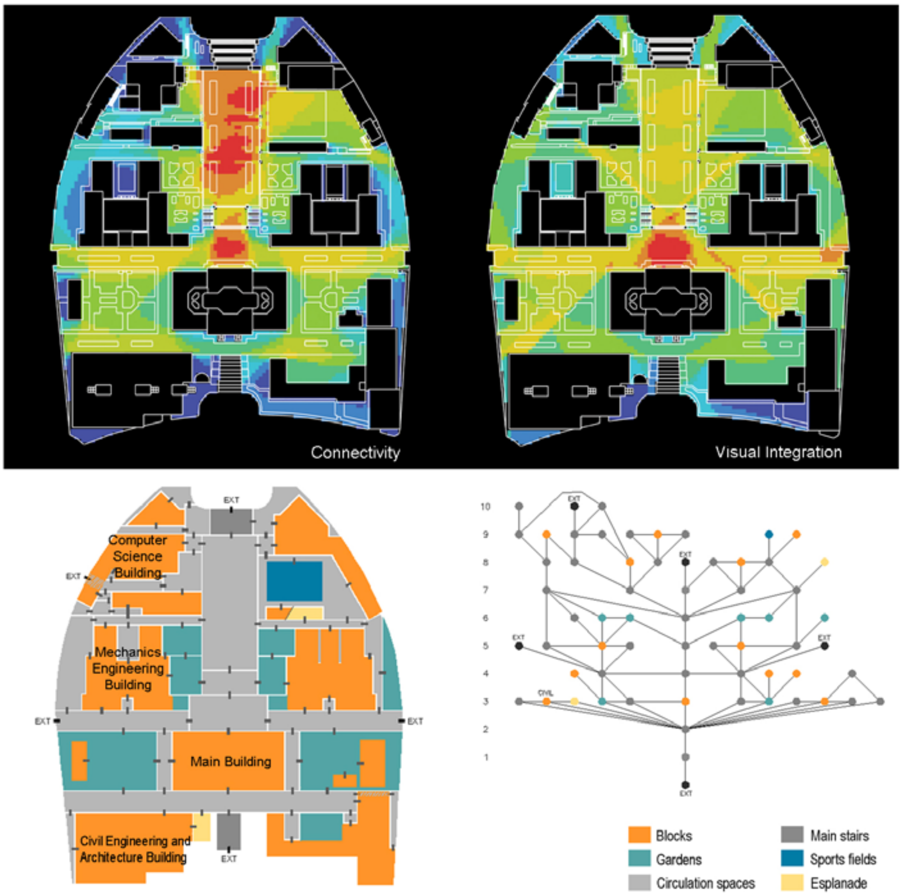


Fig. 8. Instituto Superior Técnico campus: syntactic models

This condition stimulates, simultaneously, movement through and permanence in the Architecture department building. Some flows (inbound+outbound) link main atrium, classrooms (architecture classrooms, specifically), library and students` common room to the cafeteria (Access Points located on Central Building terrace), revealing intense daily exterior-interior wireless mobility.

Main atrium, the most integrated space in the system, provides everyday access to the Architecture department building. The generalised and relatively homogenous high levels of connectivity and visual integration of main atrium play a determinant role in influencing the generation of movement and co-presence. Observations suggest that there is no deterministic circulation pattern in that movement and the potential level of co-presence and encounters are generated by randomness in this space.

Syntactic models show that exterior-interior flows are induced in the Architecture department building through two cycles that link main atrium to the exterior through the outdoor café, on the one hand, and more segregated corridors, on the other. Main atrium is the kneecap of this double-cycle subsystem that promotes movement and random encounter in the Architecture department building.

Main atrium is linked (horizontally and vertically) to a sub-system of corridor-spaces that also responds to special needs for management of visibility and permeability relations. The visual articulation of main atrium, with these corridor-spaces (open galleries), allows an immediate local correspondence between permeability (where you can go) and visibility (what you can see) defining a natural compensation between visual integration and axiality. Open galleries, and corridor-spaces, in general, allow visibility to a series of near and distant convex spaces like Computer Science Lab and students` common room.

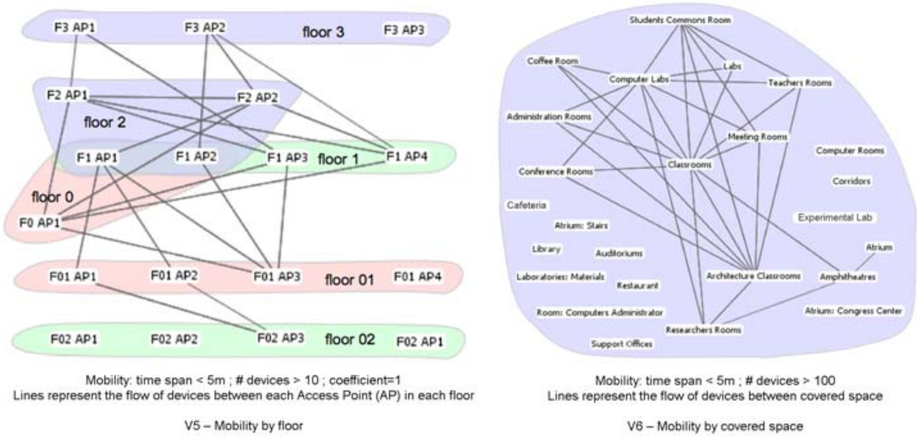


Fig. 9. FLUX* mobility visualizations (V5; V6)

Floors 0, 1 and 2, have the highest wireless activity (more traces/devices) (Fig. 9 – V5). Some Access Points are central to the mobility flows: Floor 0 – Access Point 1 (students` common room) presents the largest number of flows; Floor 1-Access Point 1 (classrooms) and Floor 1-Access Point 2 (main atrium) presents the more extensive flows, directly linked to the lower floors (3-01). Main atrium, during evaluation time, was equipped with tables and chairs. Students spontaneously appropriated it for individual and group activities. In the force-directed graph, Floor 1-Access Point 1 and Floor 1-Access Point 2 polarize other floors because of their suspension over the main atrium open space, which is, visually, highly integrated in the spatial system.

Mobility by covered space shows central spaces in mobility flows: classrooms, architecture classrooms and students` common room present the largest flow figures (and the highest number of traces/devices) (Fig. 9 – V6). Those spaces, located in the more confined zones of the spatial system, with high accessibility, support many WiFi connections with other spaces. Particularly, flows with classrooms as source, or target, always present the highest number of devices moving between covered spaces.

Mobility by knowledge-sharing types shows great mobility between all types of knowledge-sharing scenarios (Fig. 10 – V8). Particularly, "Communicating" and "Delivering" are the knowledge-sharing scenarios most linked to others. Creating is the more isolated one and its WiFi connections are, essentially, with "Communicating/Delivering" types. "More isolated" means that this knowledge-sharing type registered less number of dislocations then the others. It seems a more "static" knowledge-sharing type. This result is coherent with the current space usage because of some reasons: teachers and researchers group represent a minor group comparatively with the student population. So, it is reasonable that their wireless connections are much lesser then those made by students` population (in spite of staff use mostly the fixed network). Their activity is much more located in space then students activities. During a day of classes, students have to move between several rooms and floors. The results of the query about "How much time is spent by knowledge?" shows that "Creating" has the higher time average (58 minutes) of all knowledge-sharing types. "Communicating" and "Delivering" have the lower time average (30 minutes). These results are compatible amongst themselves and with the "static" nature of "Creating".

"Delivering" and "Communicating" are the knowledge-sharing types with more traces/devices (Fig. 10 – V9). Flows related with them, as source or target, support the higher numbers of devices moving between knowledge-sharing types. Spatially, those categories are located on floors 0, 1 and 01, corresponding to highly permeable spaces (visually and physically), vertically linked by the main atrium. "Decision Making" and "Creating" types are located in lower areas where students only go occasionally. Labs are essentially located in floors 01 and 02. Those are spaces of more punctual use than current classrooms. The more intense wireless activity is concentrated on narrow levels.

From these results (confirmed by WiFi users` observations) a spatial pattern of knowledge-sharing use emerged revealing a more dynamic and permanent wireless activity located at more permeable levels and confined spaces (floors 0 and 1), specially related with knowledge-sharing types of "Communicating" and "Delivering". The main atrium spatial structure – the central distribution space of the Architecture department building – promotes random users encounters and co-presence. The greatest proportion of interaction was found in the open space of main atrium and common spaces (Computer Science Lab and students` common room), where people meet to talk as well as study. The integrating nucleus of main atrium links the more relaxed and informal area used by students on the ground floor (Computer Science Lab, students` common room and studying classrooms) to the more segregated spaces on the upper and lower floors (teachers' rooms, researchers' rooms, labs, amphitheatres and auditorium). More restricted wireless uses, related with knowledge-sharing types of "Creating" and "Decision making", were located in deeper and segregated spaces (floors 2, 3 and 02) assuring the necessary privacy to those activities.

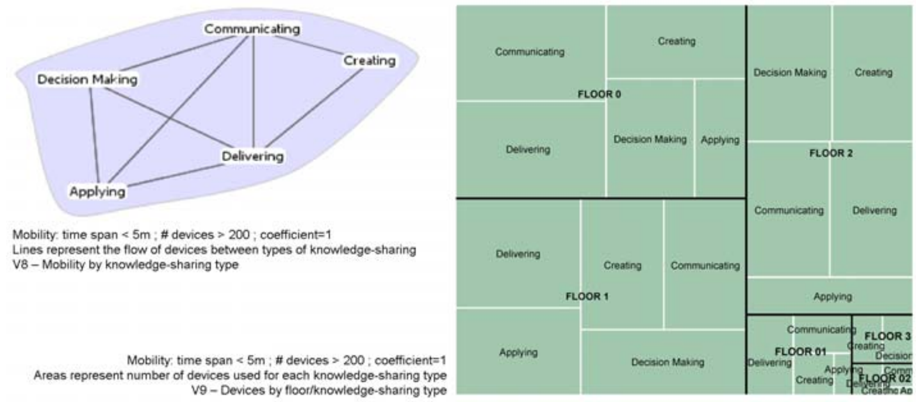


Fig. 10. FLUX* mobility visualizations (V8; V9)

These results also suggest that FLUX* visualizations may became a vehicle for the interpretation of spatial behavior observed. The interpretation of people spatial behavior implies to known how people and space interact with each other. Space is an active agent in the activities processing, influence people mobility. The syntactic models allowed us to know the configurational space properties. Spatial Information Visualization was applied to the representation of users` spatial behaviors. The two representation types are associated. This association is essential to study interactions between space and users.

It is important to notice some intrinsic qualities of FLUX* representations. Besides the direct information that these representations express, some of them support other levels of non-direct information. Direct information is information that the representation is supposed to transmit in the first place. Non-direct information is information that is not included on the FLUX* numeric data tables and only emerge from the representations. Corresponds to an increment of information besides the principal data that is supposed the representation communicate. Google maps allow other levels of non-direct information like wireless flows extension – cases of "Inbound+Outbound Flows" and "Inner mobility fluxes". These one, shows flows extensions in vertical dimension too. All Google maps allow the direct representation of the use spatialization. The consideration of space representation makes the interpretation of the relationships space-users easier.

"Devices by floor" is a more abstract representation. The spatial dimension is not so obvious but, nevertheless, the essential information - vertical organization of space - is expressed. It is possible to identify the more intense wireless communication concentrated in specific floors, the antennas that polarize those communications, and the more extension flows established between floors. Another non explicit information that arises exclusively from the representation itself.

The more abstract representations are related with knowledge-sharing types. Abstraction is a consequence of the absence of space representation. In spite spatial connotations are underlying to the concept of each knowledge-sharing type, the location of those spatial behaviors are not directly established through representations. On those cases, the interpretation process only can deal with direct information. Those are narrow representations: they don't allow adding the interpretation process any other information. The absence of other levels of information (non-direct ones), don't contribute with more profound insights about the knowledge of the interaction space-use like happens with the more deep representations like Google maps.

5 Conclusions and Future Work

This paper explores the capacity of WiFi networks as a space-use analysis input tool applied to describing emerging spatial patterns of knowledge-sharing. It is based on a crossover of Space Syntax and Spatial Information Visualization analysis applied to the building scale. For this purpose, a department building within a university campus was used as a case study.

One can conclude that the WiFi network shows capacity to function as an ad-hoc WiFi user trace system since it enables detection of user locations as well as movement flows in space and time. This data is non arbitrary and free of human factor.

Through the application of queries to the database, the FLUX* visualization platform allows one to extract relevant information for understanding the spatial knowledge-sharing patterns. The taxonomic definition of the wireless antennas is fundamental in this process. The models of visualization of the spatial use dynamics, and their quantification, have become adequate complements to the Space Syntax methodologies. The space-use analysis model revealed analytical capacities for spatial contexts for the sharing of knowledge.

However, the application of the space-use analysis model to an architectural context revealed the need for refinement in the methodological processes used. The main difficulty identified had to do with matching the wireless antennas' coverage area with the physical space.

The results seem to suggest that this analysis could have many further developments: the use of autonomous probes to track the space and registering Access Points` signal quality; the development of specific software for mapping probe information and automatically drawing, on the building plans, the course of the probes and the antenna coverage zones and calculation of overlapping areas. The definition of the Coverage Matrixes for the FLUX* programme would allow one to establish more direct space/signal correlations.

Acknowledgments. This work is part of the research activity carried out at ICIST, Instituto de Engenharia de Estruturas, Território e Construção and IN+, and has been partially supported by the Science and Technology Foundation through the funding of the research units. This work was developed with the collaboration of Diana Vaz, Sónia Ildefonso, Inês Guerreiro and Lúcia Pato. The authors would like to thank the indispensable support of Prof. Fernando Mira da Silva (CIIST, Centro de Informática do Instituto Superior Técnico) and Prof. Eduardo Pereira.

References

1. Benedikt, M. L.: To take hold of space: isovist and isovist field. Environment and Planning B: Planning and Design 6, 47–65 (1979)
2. Hillier, B., Hanson, J.: The social logic of space. Cambridge University Press, Cambridge (1984)
3. Peponis, J., Wineman, J., Rashid, M., Kim, S.-H., Bafna, S.: On the description of shape and spatial configuration inside buildings: three convex partitions and their local properties. Environment and Planning B: Planning and Design 24, 761–781 (1997)
4. Turner, A.: Analysing the visual dynamics of spatial morphology. Environment and Planning B: Planning and Design 30:5, 657–676 (2003)
5. Greene, M., Penn, A.: Socio-spatial analysis of four University Campuses: the implications of spatial configuration on creation and transmission of knowledge. In: Space Syntax 1st International Symposium, pp. 13–15. University College London, London (1997)
6. Hillier, B., Penn, A.: Visible colleges: structure and randomness in the place of discovery. Science in Context 4:1, 23–49 (1991)
7. Serrato, M., Wineman, J.: Spatial and communication patterns in research and development facilities. In: Space Syntax 2nd International Symposium. Universidade de Brasília, Brasília (1999)
8. Peatross, D., Peponis J.: Space, Education and Socialization. Journal of Architectural and Planning Research 12:4, 266–385 (1995)
9. Grajewski, T.: The SAS Head Office – Spatial Configuration and Interaction Patterns. Nordisk arkitekturforskning 2 (1993)
10. Penn, A., Desyllas, J., Vaughan, L.: The of Innovation: Interaction and Communication in the Work Environment. Environment and Planning B: Planning and Design 6 (1999)
11. Peponis, J., Conroy Dalton, R., Wineman, J., Dalton, N.: Path, theme and narrative in open plan exhibition settings. In: Space Syntax 4th International Symposium, pp. 29.1–29.20. University College London, London (2003)
12. Psarra, S., Grajewski, T.: Architecture, narrative and promenade in Benson+Forsyth`s Museum of Scotland. ARQ 4(2), 122–136 (2000)
13. Heitor, T., Tostões, A., Barranha, H., Pessoa, A.: Inside-out and outside-in: when setting and content interact. In: Space Syntax 5th International Symposium, pp. 233–249. Delft University of Technology, Delft (2005)
14. Mitchell, W.: Rethinking campus and classroom design. NLII [on line], 9, http://www.educause.edu
15. Sevtsuk, A., Ratti, C.: iSPOTS how wireless technology is changing life on the MIT Campus. In: 9th International Conference on Computers in Urban Management and Urban Planning. University College London, London (2005)

16. Neves, N., Silva, J.P., Gonçalves, P., Muchaxo, J., Silva, J.M., Câmara, A.: Cognitive spaces and metaphors: a solution for interacting with spatial data. Computers & Geosciences 23(4), 483–488 (1997)
17. Dimas, P. Silva, J.P.: Mapping the information society: the FLUX* project and Spirographis platform [on line], http://envc2005.multimeios.pt/Artigos/Paulo%20Dimas%20Artigo.pdf
18. Dimas, P. Silva, J.P., Silva, J.M.: When networks become people: using FLUX* to monitor wireless networks. In: TERENA Networking Conference, Amsterdam (2006)
19. Heer, J., Card, S.K., Landay, J.A.: Prefuse: a toolkit for interactive information visualization. In: CHI 2005 – Conference on human factors in computing systems. Portland, Oregon (2005)
20. Shneiderman, B.: Tree visualization with tree-maps: 2-d space-filling approach. ACM Transactions on graphics (TOG) 11(1), 92–99 (1992)
21. Scott-Webber, L.: In Sync – environmental behavior research and the design of learning spaces. The Society for College and University Planning, Michigan (2004)

The 9⁺-Intersection for Topological Relations between a Directed Line Segment and a Region

Yohei Kurata[1] and Max J. Egenhofer[2]

[1] SFB/TR8 Spatial Cognition, Universität Bremen
Postfach 330 440, 28334 Bremen, Germany
ykurata@informatik.uni-bremen.de

[2] National Center for Geographic Information and Analysis, and
Department of Spatial Information Science and Engineering,
University of Maine
Boardman Hall, Orono, ME 04469-5711, USA
max@spatial.maine.edu

Abstract. This paper develops a formal model of topological relations between a directed line segment (*DLine*) and a region in a two-dimensional space. Such model forms a foundation for characterizing movement patterns of an agent with respect to a region. The DLine-region relations are captured by the 9-intersection for line-region relations with further distinction of the line's boundary into two subparts (starting and ending points). This *9⁺-intersection* distinguishes 26 topological DLine-region relations. The relations' conceptual neighborhood graph takes the shape of a V-shaped tube, whose upper and lower halves are isomorphic to the conceptual neighborhood graph of 19 topological line-region relations. The conceptual neighborhood graph of the 26 DLine-region relations is applied to the iconic representation of movement patterns that satisfy a qualitative condition. By manipulating such iconic representations, the movement patterns that satisfy complex conditions are easily deduced.

1. Introduction

Movement of an agent with respect to an area, such as *entering*, *leaving*, and *going-through*, is modeled as a spatial relation between a directed line segment and a region. For instance, Figs. 1a and 1b illustrate two scenarios, *going abroad from Germany* and *being blocked by the cell wall*, by the combinations of a directed line segment and a region with different spatial relations. Similarly, how a person goes in and out of a room, a hazardous district, or any area of interest, is captured as a spatial relation between a directed line segment and a region. Among several types of spatial relations, topological relations are particularly important, because the topological relations capture how the agent moves between the inside and outside and how the agent crosses or touches the border, which are fundamental information when people conceptualize the movement. A model of such topological relations is, therefore, potentially useful for the information systems that concern spatio-dynamic behaviors, such as security monitoring systems, smart homes, mobile robots, and route navigation systems, where computers have to communicate qualitative information about spatial movement patterns with human users.

Fig. 1. (a) A directed line segment starts from the inside and ends at the outside of a Germany-shaped region, illustrating *going abroad from Germany*. (b) A directed line segment starts from the outside, touches the boundary, and ends at the outside of a cell-shaped region, illustrating *being blocked by the region.*

The goal of this paper is to develop a formal model of topological relations between a directed line segment and a region embedded in a two-dimensional space. Spatial relations between a directed line segment and a region, including topological relations, have not systematically studied, even though there are many studies on spatial relations between two directed line segments (Schlieder 1995; Clementini and Di Felice 1998; Moratz *et al.* 2000; Kurata and Egenhofer 2006) and those between a non-directed line segment and a region (Egenhofer and Herring 1991; Mark and Egenhofer 1994). Our model, which stands on these existing studies, distinguishes a set of topological relations between a directed line segment and a region in a formal way. This relation set forms a foundation for characterizing the movement patterns of an agent with respect to a region.

A non-branching, non-directed line segment is often called a *line* for short (Egenhofer and Herring (1991), Hadzilacos and Tryfona (1992), Clementini (1993), Paradias (1995), Schlieder (1995), Clementini (1998), Moratz (2000)). In contrast, this paper calls a non-branching directed line segment a *DLine* (Kurata and Egenhofer 2006). Accordingly, a spatial relation between a DLine and a region is called a *DLine-region relation*.

The remainder of this paper is structured as follows: Section 2 reviews related work on topological relations. Section 3 develops a formal model of topological DLine-region relations. Based on this model Section 4 identifies the complete set of DLine-region relations in a two-dimensional space, which are then schematized by a conceptual neighborhood graph in Section 5. Section 6 applies the conceptual neighborhood graph to the iconic representation of the movement patterns that satisfy certain qualitative conditions and demonstrates some benefits of this representation. Finally, Section 7 concludes with a discussion of future problems.

2. Related Work

Topological relations are spatial relations that are invariant under topological transformations, such as translation, rotation, and scaling (Egenhofer 1989). Topological relations are considered highly influential for people's conceptualizations of space (Lynch 1960; Egenhofer and Mark 1995b). Topological relations between two spatial objects (i.e., points, lines, and regions) and their lower-dimensional counterparts, time intervals, have been studied extensively, starting from Allen's (1983) thirteen interval relations. Based on the point-set topology (Alexandroff 1961), the *4-intersection* (Egenhofer and Franzosa 1991) formally captures topological relations between two spatial objects through the geometric intersections of the objects' interiors and boundaries. The *9-intersection* (Egenhofer and Herring 1991) further considers the intersections with respect to the objects' exteriors. In this model, topological relations between two spatial objects *A* and *B* are characterized by the *9-intersection matrix* (Eqn. 1), where A°, ∂A, and A^- are *A*'s interior, boundary, and exterior while B°, ∂B, and B^- are *B*'s interior, boundary, and exterior, respectively. Based on the presence or absence of the 3×3 types of geometric intersections, the 9-intersection distinguishes 19 topological relations between a line and a region in $\mathbf{R}^2$ (Egenhofer and Herring 1991) and 43 topological relations between a complex line and a complex region in $\mathbf{R}^2$ (Schneider and Behr 2006).

$$M(A,B)=\begin{pmatrix} A^\circ\cap B^\circ & A^\circ\cap\partial B & A^\circ\cap B^- \\ \partial A\cap B^\circ & \partial A\cap\partial B & \partial A\cap B^- \\ A^-\cap B^\circ & A^-\cap\partial B & A^-\cap B^- \end{pmatrix} \quad \textbf{(1)}$$

Another variation of the 4-intersection distinguishes explicitly the disconnected subparts of the interval's boundary or DLine's boundary (i.e., its starting point and ending point), distinguishing 16 relations between two intervals in a temporal cycle (Hornsby *et al.* 1999) and 68 relations between two DLines in $\mathbf{R}^2$ (Kurata and Egenhofer 2006).

A set of spatial relations is typically schematized by a *conceptual neighborhood graph* (Egenhofer and Al-Taha 1992; Freksa 1992a; Egenhofer and Mark 1995b; Papadias *et al.* 1995; Schlieder 1995; Hornsby *et al.* 1999; Egenhofer 2005; Van de Weghe and De Maeyer 2005; Kurata and Egenhofer 2006). In this graph, each node corresponds to a spatial relation and two nodes are linked if the corresponding relations are *conceptual neighbors* (Freksa 1992a). Different definitions of conceptual neighbors lead to different graphs. For instance, Egenhofer and Mark (1995a) derived two different conceptual neighborhood graphs of the 19 line-region relations based on the *smooth-transition* (Freksa 1992a), which requires the possibility of continuous transformation between the neighboring relations, and the *minimum topological distance* (Egenhofer and Al-Taha 1992), which requires minimum difference between the 9-intersection matrices of the neighboring relations.

The conceptual neighborhood graph has been applied to the analysis of spatial predicates in natural languages (Mark and Egenhofer 1994; Shariff *et al.* 1998). These linguistic studies show that line-region relations are often associated with a spatial predicate that assumes a spatial movement along the line, such as *going into*. This implies that people may recognize a line segment by imposing a virtual movement on it, despite the lack of the line's direction, just like such verbal expressions as "*the mountain range goes from Mexico to Canada*" evokes *fictive motion* (Talmy 1996).

3. The 9⁺-intersection for Topological DLine-Region Relations

This paper considers DLines that may curve, but have no loop. Such DLines are *simple lines* with direction, which is obtained through a continuous one-to-one mapping from [0, 1] to $\mathbf{R}^2$ (Schneider and Behr 2006). In the point-set topology (Alexandroff 1961), a simple line is considered a set of an infinite number of linearly aligned points, among which two distinctive end-points form the *boundary* and the other points form the *interior*. The *exterior* is the complement of the union of the boundary and the interior. Naturally, the interior, boundary, and exterior of a DLine are pairwise disjoint and jointly exhaustive in $\mathbf{R}^2$. A DLine categorizes its two end-points into a *starting point* and an *ending point*, which are also called a *tail* and a *head*, respectively (Kurata and Egenhofer 2006).

A *single-component region* is a connected, homogeneously two-dimensional 2-cell in $\mathbf{R}^2$ (Schneider and Behr 2006). A single-component region does not have two or more disconnected interiors, spikes, puncturing points, cuts, but may have holes. A *multi-component region* is the union of multiple disjoint single-component regions. In this paper, a *region* refers to a single-component region or a multi-component region. The interior, boundary, and exterior of a region are pairwise disjoint and jointly exhaustive in $\mathbf{R}^2$.

Let *D* and *R* be a DLine and a region, respectively. Uppercase letters are used because they are considered point sets. In the 9-intersection (Egenhofer and Herring 1991) their topological relation is captured through the geometric intersections of *D*'s three topological parts (i.e., interior D°, boundary ∂D, and exterior D^-) and *R*'s three topological parts (i.e., interior R°, boundary ∂R, and exterior R^-). These 3×3 types of intersections are concisely represented by the 9-intersection matrix in Eqn. 2.

$$M(D,R)=\begin{pmatrix} D^\circ \cap R^\circ & D^\circ \cap \partial R & D^\circ \cap R^- \\ \partial D \cap R^\circ & \partial D \cap \partial R & \partial D \cap R^- \\ D^- \cap R^\circ & D^- \cap \partial R & D^- \cap R^- \end{pmatrix} \qquad (2)$$

In our model, the intersections with respect to *D*'s boundary ∂D are further distinguished into the intersections with respect to *D*'s starting point $\partial_s D$ and those with respect to *D*'s ending point $\partial_e D$. Accordingly, the 9-intersection matrix in Eqn. 2 is extended to the matrix in Eqn. 3, which is called the *9⁺-intersection matrix* of the topological relation between the DLine *D* and the region *R*.

$$M^+(D,R)=\begin{pmatrix} D^\circ \cap R^\circ & D^\circ \cap \partial R & D^\circ \cap R^- \\ \begin{bmatrix} \partial_s D \cap R^\circ \\ \partial_e D \cap R^\circ \end{bmatrix} & \begin{bmatrix} \partial_s D \cap \partial R \\ \partial_e D \cap \partial R \end{bmatrix} & \begin{bmatrix} \partial_s D \cap R^- \\ \partial_e D \cap R^- \end{bmatrix} \\ D^- \cap R^\circ & D^- \cap \partial R & D^- \cap R^- \end{pmatrix} \qquad (3)$$

In general, the 9⁺-intersection captures topological relation between two spatial objects *A* and *B* through the geometric intersections of all subparts of *A*'s interior, boundary, and exterior and all subparts of *B*'s interior, boundary, and exterior. The subparts of the interior, boundary, and exterior are determined by their disconnections. For instance, a single-component region with n holes has $n + 1$ boundary subparts and $n + 1$ exterior subparts. On the other hand, we consider that a connected interior, boundary, or exterior consists of a single subpart. Accordingly, Eqn. 4 is the general form of the 9^+-intersection matrix, where A°_i}, $\partial_i A$, and A^{-_i} are the i-th subpart of A's interior, boundary, and exterior while B°_j}, $\partial_j B$, and B^{-_j} are the j-th subpart of B's interior, boundary, and exterior, respectively.

$$M^+(A,B)=\begin{pmatrix} \left[A^{\circ_i}\cap B^{\circ_j}\right] & \left[A^{\circ_i}\cap \partial_j B\right] & \left[A^{\circ_i}\cap B^{-_j}\right] \\ \left[\partial_i A\cap B^{\circ_j}\right] & \left[\partial_i A\cap \partial_j B\right] & \left[\partial_i A\cap B^{-_j}\right] \\ \left[A^{-_i}\cap B^{\circ_j}\right] & \left[A^{-_i}\cap \partial_j B\right] & \left[A^{-_i}\cap B^{-_j}\right] \end{pmatrix} \quad (4)$$

For instance, the DLine-region relation in Fig. 2b and the DLine-DLine relation in Fig. 2c are captured by the 9^+-intersection matrix shown in each figure.

For simplification, the 9^+-intersection matrices are represented by icons (Fig. 2b-2c). These icons are based on the iconic representation of the 9-intersection matrix by Mark and Egenhofer (1994) (Fig. 2b). Each icon has 3×3 cells, which correspond to the matrix's 3×3 elements. Each cell is marked out if the corresponding element is non-empty ($\neg\phi$). In our iconic representation, the icon's columns and rows are partitioned if their corresponding topological parts have multiple subparts. For instance, when visualizing the 9^+-intersection matrix of a DLine-region relation, the icon's second row is partitioned (Fig. 2b), such that the upper and lower halves correspond to the intersections with respect to the DLine's starting point and ending point, respectively. Similarly, for a DLine-DLine relation, both the second row and the second column of the icon are partitioned (Fig. 2c).

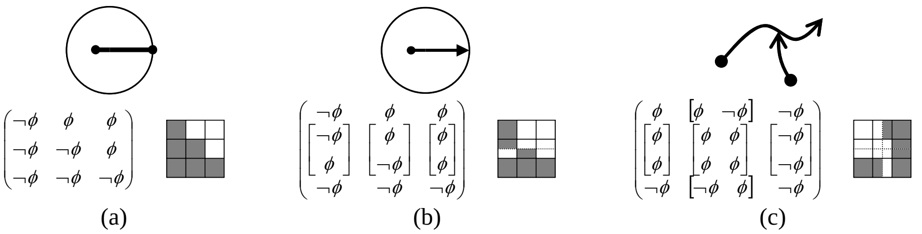


Fig. 2. Sample configurations of (a) a line and a region, (b) a DLine and a region, and (c) two DLines, together with (a) the 9-intersection and (b-c) the 9^+-intersection matrices which capture their topological relations. The icons visualize the pattern of these matrices.

4. Set of Topological DL-Region Relations

From the 19 topological line-region relations distinguished by the 9-intersection (Egenhofer and Herring 1991) we can derive a set of topological DLine-region relations through assigning directions to the lines. Among the 19 relations, twelve relations (Figs. 3a-l) are invariant to the line's direction, because the line's two end-points are located at the same part of the region, whereas seven relations (Fig. 3m-s) are variant to the line's direction. From each direction-variant relation we can derive two DLine-region relations through assigning different directions to the line. As a result, $12 + 7 \times 2 = 26$ topological DLine-region relations are obtained (Fig. 4).

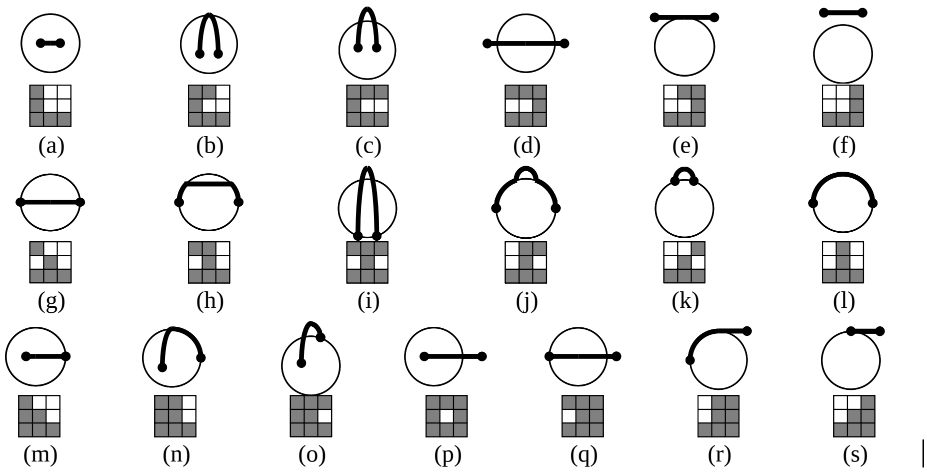


Fig. 3. Sample configurations of 19 topological line-region relations (Egenhofer and Herring 1991), together with the patterns of the corresponding 9-intersection matrices.

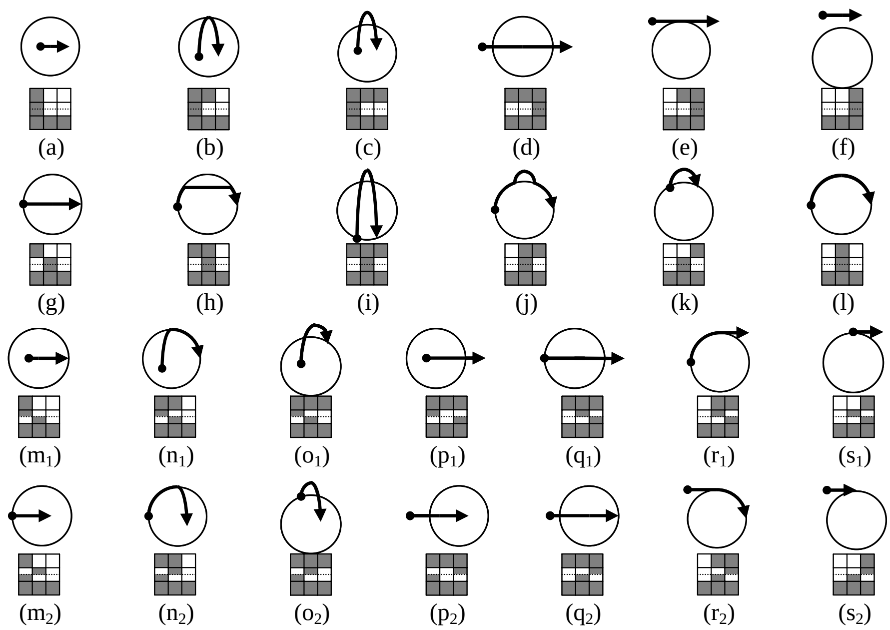


Fig. 4. Sample configurations of 26 topological DLine-region relations, together with the patterns of the corresponding 9^{+}-intersection matrices.

These 26 relations correspond to different patterns of the 9^+-intersection matrices (Fig. 4). This indicates that the 9^+-intersection distinguishes *at least* 26 DLine-region relations. Actually, there is no other DLine-region relation that the 9^+-intersection distinguishes. This is proven with the aid of the constraints on the 9^+-intersection matrix of DLine-region relations (Eqns. 6-14), which are derived systematically from the constraints on the 9-intersection matrix of line-region relations (Egenhofer and Herring 1991) considering the distinction of lines' two end-points.

- *D*'s exterior and *R*'s exterior intersect with each other.

$$D^- \cap R^- = \neg\phi \tag{5}$$

- If *D*'s interior is a subset of *R*'s closure then *D*'s both end-points must be a subset of *R*'s closure as well.

$$D^\circ \cap R^- = \phi \quad \rightarrow \quad \partial_s D \cap R^- = \phi \quad \wedge \quad \partial_e D \cap R^- = \phi \tag{6}$$

- *D*'s each end-point intersects at least one part of *R*.

$$\exists P, Q \in \{R^\circ, \partial R, R^-\} \quad \partial_s D \cap P = \neg\phi \quad \wedge \quad \partial_e D \cap Q = \neg\phi \tag{7}$$

- If *D*'s interior and *R*'s interior are disjoint then neither of *D*'s end-points can intersect with *R*'s interior.

$$D^\circ \cap R^\circ = \phi \quad \rightarrow \quad \partial_s D \cap R^\circ = \phi \quad \wedge \quad \partial_e D \cap R^\circ = \phi \tag{8}$$

- If *D*'s interior intersects with both *R*'s interior and exterior, then it must also intersect with *R*'s boundary.

$$D^\circ \cap R^\circ = \neg\phi \quad \wedge \quad D^\circ \cap R^- = \neg\phi \quad \rightarrow \quad D^\circ \cap \partial R = \neg\phi \tag{9}$$

- *D*'s each end-point intersects one part of *R*.

$$\begin{array}{l} \exists P_1 \in \{R^\circ, \partial R, R^-\}, \forall Q_1 \in \{R^\circ, \partial R, R^-\} \setminus P_1 \quad \partial_s D \cap P_1 = \neg\phi \quad \wedge \quad \partial_s D \cap Q_1 = \phi \\ \exists P_2 \in \{R^\circ, \partial R, R^-\}, \forall Q_2 \in \{R^\circ, \partial R, R^-\} \setminus P_2 \quad \partial_e D \cap P_2 = \neg\phi \quad \wedge \quad \partial_e D \cap Q_2 = \phi \end{array} \tag{10}$$

- *R*'s interior always intersects with *D*'s exterior.

$$D^- \cap R^\circ = \neg\phi \tag{11}$$

- *R*'s boundary always intersects with *D*'s exterior.

$$D^- \cap \partial R = \neg\phi \tag{12}$$

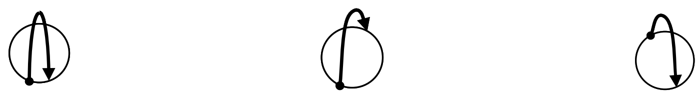
- *D*'s interior must intersects with at least one part of *R*.

$$\exists P \in \{R^\circ, \partial R, R^-\} \quad D^\circ \cap P = \neg\phi \tag{13}$$

The 9^+-intersection matrix may have $2^{12} = 4096$ patterns, since it has 12 elements with two possible values (empty or non-empty). Among these 4096 patterns, however, only 26 patterns satisfy the constraints in Eqns. 6-14. These 26 patterns are exactly same as the matrix patterns in Fig. 4. This indicates that the 26 patterns in Fig. 4 are the complete set of topological DLine-region relations distinguished by the 9^+-intersection.

In Fig. 4, the geometric configuration assigned to each DLine-region relation is merely an example. It is possible that other configurations, which are topologically different from the illustrated example, may also correspond to the same topological relation (Fig. 5). In this sense, the 26 topological relations categorize DLine-region configurations based on some topological characteristics (the presence or absence of 12 types of intersections), but not on their topological equivalence.

To describe the topological detail of DLine-region configurations, this paper also introduces an alternative notation by three-tuples, which trace the positions of a virtual agent moving along the DLine (Fig. 5a-c). The first and third element in each three-tuple represents the agent's starting and ending positions, while the second element represents the sequence of the agent's intermediate positions. *I*, *B*, *E* represent the positions in the region's interior, boundary, and exterior, respectively.



(a) $(B, IBEBI, B)$ (b) (B, IBE, B) (c) (B, EBI, I)

Fig. 5. Three configurations that correspond to the same DLine-region relation in Fig. 4i.

With such three-tuple notations, Table 1 summarizes all DLine-region configurations that belong to the 26 DLine-region relations. This table indicates that 17 DLine-region relations correspond to multiple configurations (i.e., they have multiple topological interpretations).

Table 1. Sets of DLine-region configurations that belong to each DLine-region relations in Fig. 4. Each set of configurations is described using the three-tuple notation, where [*X*] is an empty or *X*, Y^* is an arbitrary number of *Y*, and *Z*|*W* is *Z* or *W*, but not both.

Direction-Invariant Relations		Direction-Variant Relations	
(a)	(I, I, I)	(m_1)	(I, I, B)
(b)	$(I, IB[IB]^*I, I)$	(m_2)	(B, I, I)
(c)	$(I, IB[IB\|EB]^*E[BI\|BE]^*BI, I)$	(n_1)	$(I, IB[IB]^*[I], B)$
(d)	$(E, EB[IB\|EB]^*I[BI\|BE]^*BE, E)$	(n_2)	$(B, [I][BI]^*BI, I)$
(e)	$(E, EB[EB]^*E,E)$	(o_1)	$(I, IB[IB\|EB]^*E[BI\|BE]^*[B], B)$
(f)	(E, E, E)	(o_2)	$(B, [B][IB\|EB]^*E[BI\|EI]^*BI, I)$
(g)	(B, I, B)	(p_1)	$(I, IB[IB\|EB]^*E, E)$
(h)	$(B, [B]IB[IB]^*, B)$, $(B, [BI]^*BI[B], B)$	(p_2)	$(E, E[BI\|BE]^*BI, I)$
(i)	$(B, [B][IB\|EB]^*(IBE\|EBI)[BI\|BE]^*[B], B)$	(q_1)	$(B, [B][IB\|EB]^*I[BI\|BE]^*BE, E)$
(j)	$(B, [B]EB[EB]^*, B)$, $(B, [BE]^*BE[B], B)$	(q_2)	$(E, EB[IB\|EB]^*I[BI\|BE]^*[B], B)$
(k)	(B, E, B)	(r_1)	$(B, [E][BE]^*BE, E)$
(l)	(B, B, B)	(r_2)	$(E, EB[EB]^*[E], B)$
		(s_1)	(B, E, E)
		(s_2)	(E, E, B)

5. Conceptual Neighborhood Graphs for DL-Region Relations

We schematize the 26 DLine-region relations graphically, using a *conceptual neighborhood graph* (Freksa 1992a). In this graph, each node corresponds to a spatial relation, and two nodes are linked if the corresponding relations are similar relations called *conceptual neighbors*. This paper considers that two DLine-region relations are neighbors if one relation can be derived from another relation by moving either starting point, interior, or ending point of the DLine while keeping the presence or absence of the intersections with respect to the others (Fig. 6a). This transformation is a subset of the *smooth transitions* (Egenhofer and Mark 1995a), which also includes the transformation by moving one part of the DLine without keeping the presence or absence of the intersections with respect to the others (Fig. 6b).

Fig. 6. Smooth transitions between two DLine-region relations, derived by moving the DLine's ending point from the region's boundary to (a) interior and (b) exterior, respectively. In (b), the movement of the DLine's ending point generates the intersection of the DLine's interior and the region's boundary.

We identified 46 pairs of conceptual neighbors among the 26 DLine-region relations. By linking these neighbors, a conceptual neighborhood graph of the 26 topological DLine-region relations is developed (Fig. 7). The graph is non-planar and drawn three-dimensionally on a V-shaped tube, such that links do not cross. Clearly, this graph schematizes DLine-region relations based on their similarity.

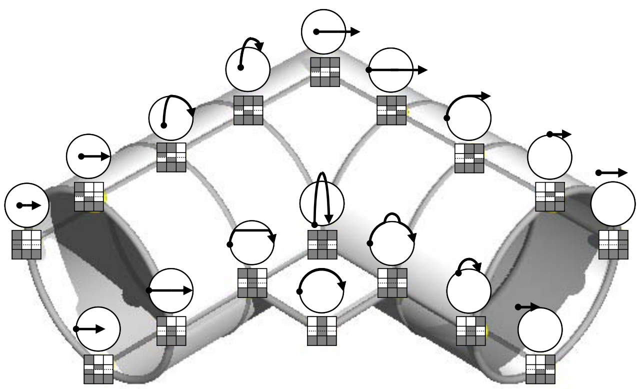


Fig. 7. A conceptual neighborhood graph of the 26 topological DLine-regions.

Fig. 8 shows the upper and lower halves of the conceptual neighborhood graph in Fig. 7. The relations with gray background in Fig. 8 correspond to the relations located at the top and bottom of the V-shaped tube in Fig. 7, respectively. Interestingly, the two subgraphs in Fig. 8 are isomorphic to the conceptual neighborhood graph of the 19 line-region relations (Fig. 9). Actually, the two subgraphs in Fig. 8 can be derived from the graph in Fig. 9 through assigning directions to the line in each relation. Since the line-region relations with gray background in Fig. 9 are variant to the line's direction, different directions assigned to the lines yield different conceptual neighborhood graphs.

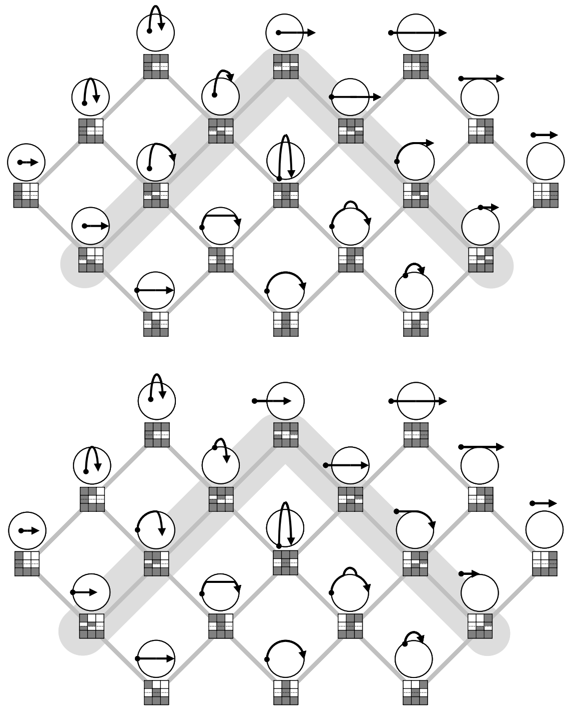


Fig. 8. Upper and lower halves of the conceptual neighborhood graph in Fig. 7. The relations with gray background correspond to the relation located at the top and bottom of the graph in Fig. 7, respectively.

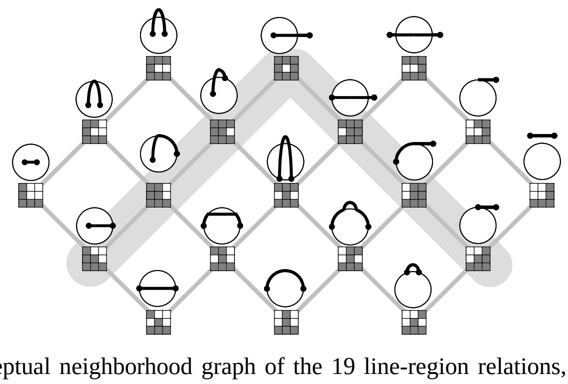


Fig. 9. The conceptual neighborhood graph of the 19 line-region relations, derived under the same definition of conceptual neighbors. The relations with gray background are variant to the line's direction.

The conceptual neighborhood graph of the 26 DLine-Region relations in Fig. 7 has the following unique properties:

- Pairs of vertically facing relations are derived from each other by exchanging upper and lower halves of the second row of the 9^+-intersection matrices (essentially reversing the DLine's direction).
- Pairs of relations located symmetrically across the front-to-back line penetrating the V-tube's center are derived from each other by flipping the 9^+-intersection matrices horizontally (essentially reversing the region's interior and exterior).
- The number of different elements in the 9^+-intersection matrix, called the *topological distance* (Egenhofer and Al-Taha 1992), is 1 between the horizontal neighbors and 2 between the other neighbors. This implies that the conceptual neighborhood graph in Fig. 7 cannot be obtained through linking all pairs of relations with minimum topological distance.

6. Modeling Movement Patterns with Respect to a Region

Topological DLine-region relations categorize the patterns of an agent's movement with respect to a region. For instance, the DLine-region relation in Fig. $4p_1$ corresponds to movement patterns that start from the region's interior, cross the region's boundary at least once, and end at the region's exterior. Such categorization is useful, because it highlights the topological characteristic of movement which highly influences people's conceptualization of movement, while abstracts less important detail. Another benefit of such categorization is that movement patterns that satisfy a certain qualitative condition, whose number is typically infinite, are captured by a finite set of DLine-region relations. For instance, a set of movement patterns that satisfy the qualitative condition, *starting from the region's interior*, is represented by the set of seven DLine-region relations (Figs. 4a, 4b, 4c, $4m_1$, $4n_1$, $4o_1$, and $4p_1$). Such summarized expression makes it easy for computers to process people's characterization of movement patterns.

To visually represent a set of DLine-region relations, we introduce iconic representations (Figs. 10a-h), which superimpose the two subgraphs in Fig. 8. The set of marked nodes indicates the set of DLine-region relations. Some nodes are partitioned, such that their upper and lower halves correspond to the different relations in the upper and lower subgraphs at the same position (i.e., the relations located at the top and bottom of the V-shaped tube in Fig. 7). For instance, the icon in Fig. 10a shows the set of seven relations in Fig. 4a, 4b, 4c, $4m_1$, $4n_1$, $4o_1$, and $4p_1$, which corresponds to the previous condition, *starting from the region's interior*. Here we consider ten basic qualitative conditions: (a) *starting from interior*, (b) *starting from boundary*, (c) *starting from exterior*, (d) *crossing boundary*, (d) *ending at interior*, (e) *ending at boundary*, (f) *ending at exterior*, and (g) *crossing/touching boundary*. The icons in Figs. 10a-h show the sets of DLine-region relations that represent all movement patterns satisfying each of these conditions.

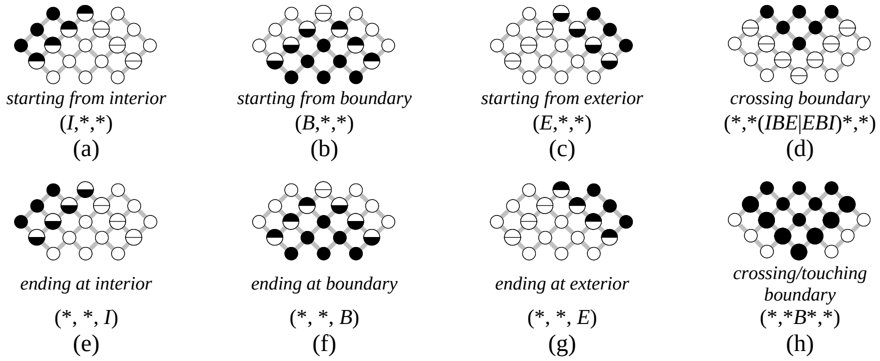


Fig. 10. Marked nodes in each icon indicate the set of DLine-region relations that represent all movement patterns satisfying each qualitative condition. Each condition is also described by the three-tuple notation (Section 4) with a wildcard symbol *.

The ten icons in Fig. 10a-h have the following unique properties:

- DLine-region relations that correspond to each qualitative condition form a connected subgraph. This is because the movement patterns that satisfy each condition has certain topological similarity, while the conceptual neighborhood graph schematizes the DLine-region relations based on their topological similarities.
- A non-directed condition, such as *crossing boundary*, yields a symmetric icon (Figs. 10d and 10h).
- A pair of conditions, interchangeable by exchanging *starting from* and *ending at* yield a pair of icons with reversed partitions (Figs. 10a and 10e, Figs. 10b and 10f, and Figs. 10c and 10g).
- A pair of conditions, interchangeable by exchanging *interior* and *exterior*, yields a pair of horizontally flipped icons with reversed partitions (Figs. 10a and 10c, and Figs. 10e and 10g).

A merit of this iconic representation is that the set of movement patterns that satisfy a complex condition is derived through simple manipulations on the icons. For instance, Fig. 11a shows the intersection of the icons in Figs. 10a and 10g, whose result indicates that only one DLine-region relation corresponds to the movement patterns satisfying *starting from interior and end at exterior* (Fig. 4p_1). Similarly, Figs. 11b-d show the union of two icons, the difference of two icons, and the complement of an icon, whose result correspond to the movement patterns satisfying *starting from inside or ending at outside*, *starting from inside but not ending at outside*, and *not starting from inside*, respectively. Such computation is particularly useful for integrating the qualitative characterizations of a movement pattern reported by multiple observers (e.g., different sensors).

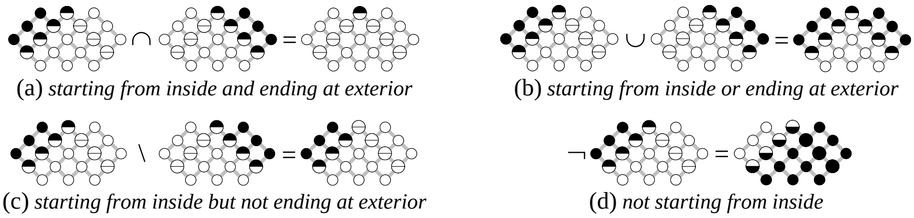


Fig. 11. Set operations on the icons for representing a set of DLine-region relations.

The iconic representation is also applicable to the classification of the region's boundary. For instance, the icon in Fig. 12a corresponds to an inapproachable boundary, like a prison's wall, which people normally cannot cross or even stand on. Conversely, the icon in Fig. 12e corresponds to a freely crossable boundary, like the border between a city square and roads around the square. In this way, the regions' boundaries can be graphically classified in terms of the movement patterns they accept (Fig. 12).

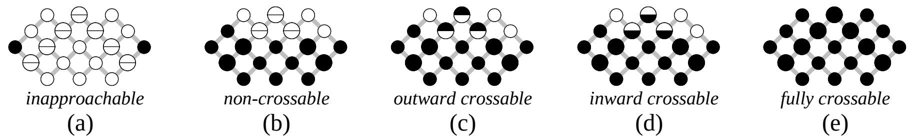


Fig. 12. Five types of regions' boundaries in terms of the movement patterns they accept.

7. Conclusion and Future Problems

This paper distinguished 26 topological DLine-region relations based on the 9^+-intersection. A conceptual neighborhood graph of these 26 relations took the shape of a V-shaped tube except one node. The 26 DLine-region relations qualitatively categorize the patterns of an agent's movement with respect to a region. It is a future question how to apply this model to the analysis of motional expressions in natural language (i.e., how people describes a movement), as well as the computer systems that communicate the information on spatial movement with human users, hopefully in a qualitative way.

In the current model, we can tell whether an agent crosses or touches the boundary, but cannot describe how many times the agent crosses or touches the boundary, or whether the agent crosses or touches the boundary instantly or stepwise. Also, we cannot describe, if the DLine and the region are disjoint, whether the agent goes left of, right of, toward, or against the region, even though people often emphasize such information when describing movement. It is, therefore, a future question to extend the current model in a meaningful way, incorporating other topological properties of DLine-region relations, such as the number and dimension (point or interval) of intersections, as well as non-topological properties such as distance and direction between the DLine and the region.

This paper did not discuss the composition of two DLine-region relations (Freksa 1992b; Egenhofer 1994). We can consider two types of compositions: the composition of two DLine-regions relations with a common DLine (i.e., all possible relations between two regions R_1 and R_2 when the relation between a DLine D and R_1 and the relation between D and R_2 are known) or that with a common region (i.e., all possible relations between two DLines D_1 and D_2 when the relation between D_1 and a region R and the relation between D_2 and R are known). The compositions of all pairs of 26 DLine-region relations yield two 26×26 composition tables, which will enrich the foundation of qualitative spatial reasoning.

The 9^+-intersection introduced in this paper provides a flexible and systematic framework for capturing topological relations between various spatial objects, including DLines, branching lines, regions with holes, and multi-component regions. The application of the 9^+-intersections for other spatial relations is highly potential for enriching the formal model of space configurations and the foundation of qualitative spatial reasoning.

Acknowledgments

This work was partially supported by the National Geospatial-Intelligence Agency in USA under grant number NMA201-01-1-2003 and NMA401-02-1-2009.

References

Alexandroff, P. (1961) *Elementary Concepts of Topology*. Dover, Mineola, NY.

Clementini, E. and Di Felice, P. (1998) Topological Invariants for Lines. *IEEE Transactions on Knowledge and Data Engineering*, **10**(1): 38-54.

Clementini, E., Di Felice, P., and van Oosterom, P. (1993) A Small Set of Formal Topological Relationships Suitable for End-User Interaction. in: Abel, D. and Ooi, B. C. (Eds.) *3rd International Symposium on Advances in Spatial Databases*, Singapore, Lecture Notes in Computer Science, 692, 277-295, Springer.

Egenhofer, M. (1989) A Formal Definition of Binary Topological Relationships. in: Litwin, W. and Schek, H.-J. (Eds.) *3rd International Conference on Foundations of Data Organization and Algorithms (FODO)*, Paris, France, Lecture Notes in Computer Science, 367, 457-472, Springer.

Egenhofer, M. (1994) Deriving the Composition of Binary Topological Relations. *Journal of Visual Languages and Computing*, **5**(2): 133-149.

Egenhofer, M. (2005) Spherical Topological Relations. *Journal on Data Semantics III*: 25-49.

Egenhofer, M. and Al-Taha, K. (1992) Reasoning About Gradual Changes of Topological Relationships. in: Frank, A., Campari, I., and Formentini, U. (Eds.) *Theories and methods of spatio-temporal reasoning in geographic space*, Pisa, Italy, Lecture Notes in Computer Science, 639, 196-219, Springer.

Egenhofer, M. and Franzosa, R. (1991) Point-Set Topological Spatial Relations. *International Journal of Geographical Information Systems*, **5**(2): 161-174.

Egenhofer, M. and Herring, J. (1991) Categorizing Binary Topological Relationships between Regions, Lines and Points in Geographic Databases. in: Egenhofer, M., Herring, J., Smith, T., and Park, K. (Eds.) *A Framework for the Definitions of Topological Relationships and*

an Algebraic Approach to Spatial Reasoning within This Framework, NCGIA Technical Reports 91-7. National Center for Geographic Information and Analysis, Santa Barbara, CA.
Egenhofer, M. and Mark, D. (1995a) Modeling Conceptual Neighborhoods of Topological Line-Region Relations. *International Journal of Geographical Information Systems*, **9**(5): 555-565.
Egenhofer, M. and Mark, D. (1995b) Naive Geography. in: Frank, A. and Kuhn, W. (Eds.) *COSIT '95*, Semmering, Austria, Lecture Notes in Computer Science, 988, 1-15, Springer.
Freksa, C. (1992a) Temporal Reasoning Based on Semi-Intervals. *Artificial Intelligence*, **54**: 199-227.
Freksa, C. (1992b) Using Orientation Information for Qualitative Spatial Reasoning. in: Frank, A., Campari, I., and Formentini, U. (Eds.) *The International Conference GIS - From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, New York, NY, Lecture Notes in Computer Science, 639, 162-178, Springer.
Hadzilacos, T. and Tryfona, N. (1992) Model for Expressing Topological Integrity Constraints in Geographic Databases. in: Frank, A., Campari, I., and Formentini, U. (Eds.) *The International Conference GIS—From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, New York, NY, Lecture Notes in Computer Science, 639, 252-268, Springer.
Hornsby, K., Egenhofer, M., and Hayes, P. (1999) Modeling Cyclic Change. in: Chen, P., Embley, D., Kouloumdjian, J., Liddle, S., and Roddick, J. (Eds.) *Advances in Conceptual Modeling*, Paris, France, Lecture Notes in Computer Science, 1227, 98-109, Springer.
Kurata, Y. and Egenhofer, M. (2006) The Head-Body-Tail Intersection for Spatial Relation between Directed Line Segments. in: *GIScience2006*, Münster, Germany, Lecture Notes in Computer Science, 4197, 269-286, Springer.
Kuratowski, K. (1930) Sur Le Probleme Des Courbes Gauches En Topologie. *Fundamenta Mathematicae*, **16**: 271-283.
Lynch, K. (1960) *The Image of the City*. MIT Press, Cambridge, MA.
Mark, D. and Egenhofer, M. (1994) Modeling Spatial Relations between Lines and Regions: Combining Formal Mathematical Models and Human Subjects Testing. *Cartography and Geographical Information Systems*, **21**(3): 195-212.
Moratz, R., Renz, J., and Wolter, D. (2000) Qualitative Spatial Reasoning About Line Segments. in: Horn, W. (Ed.) *14th European Conference on Artificial Intelligence*, Berlin, Germany, 234-238, IOS Press.
Papadias, D., Theodoridis, Y., Sellis, T., and Egenhofer, M. (1995) Topological Relations in the World of Minimum Bounding Rectangles: A Study with R-Trees. in: Carey, M. and Schneider, D. (Eds.), SIGMOD RECORD, 24(2), 92-103.
Schlieder, C. (1995) Reasoning About Ordering. in: Frank, A. and Kuhn, W. (Eds.) *COSIT '95*, Lecture Notes in Computer Science, 988, 341-349, Springer.
Schneider, M. and Behr, T. (2006) Topological Relationships between Complex Spatial Objects. *ACM Transactions on Database Systems*, **31**(1): 39-81.
Shariff, A., Egenhofer, M., and Mark, D. (1998) Natural-Language Spatial Relations between Linear and Areal Objects: The Topology and Metric of English-Language Terms. *International Journal of Geographical Information Science*, **12**(3): 215-246.
Talmy, L. (1996) Fictive Motion in Language and "Ception". in: Bloom, P., Peterson, M., Nadel, L., and Garrett, M. (Eds.) *Language and Space*. MIT Press, Cambridge, MA, 211-276.
Van de Weghe, N. and De Maeyer, P. (2005) Conceptual Neighbourhood Diagrams for Representing Moving Objects. in: *2nd International Workshop on Conceptual Modeling for Geographic Information Systems (CoMoGIS)*, Klagenfurt, Austria, Lecture Notes in Computer Science, 3770, 228-238, Springer.

Spatio-temporal configurations of dynamics points in a 1D space

Pierre Hallot[1], Roland Billen[1]

[1] Geomatics Unit, University of Liège, Allée du 6 Août, 17 B-4000 Liège Belgium
{P.Hallot@ulg.ac.be, rbillen@ulg.ac.be}

Abstract. This paper describes a spatio-temporal configurations building approach, which has been applied to dynamics points in 1-dimensional space. In this approach, a temporal logic, Allen's time intervals, is crossed with a new spatial logic based on topology called spatial states. These spatial states are derived from topological relationships and a new concept of degenerate topological relationships. This work is the first step of a PhD research aiming to create a generalized spatio-temporal reasoning model based on topological relationships between spatio-temporal histories .

Keywords: Spatio-temporal modelling, spatio-temporal relationships, lifelines, spatio-temporal configuration, spatial states, degenerate topological relationships.

1 Introduction

For years now, several research communities (GIS, AI, etc.) have investigated spatio-temporal representations and reasoning. It was a logical evolution after putting so much effort in (qualitative) spatial reasoning and temporal reasoning. Indeed, there is a lot of applications where spatio-temporal reasoning is or could be beneficial: movement description, monitoring objects, region evolution, trajectory calculus, epistemology, crime mapping, on board-GPS analyses, etc. In behavior and monitoring interpretation for instance, spatio-temporal reasoning could be used to reason about the interaction of people with their environment or to describe motion patterns of moving objects (peoples, animals, vehicles…) [1-7].

So far, different types of spatio-temporal reasoning models have been developed. Some of them combine spatial and temporal logic; they "temporalize" spatial reasoning models. Others try to create spatio-temporal mereotopology directly from the spatio-temporal histories of life-lines [8]. Following this latter approach, we aim to develop a generalised spatio-temporal calculus based on spatio-temporal histories. The underlying idea is to extract spatio-temporal information by applying topological calculi on life-lines (e.g. considering a life-line as a line in 2D geometrical space). As a preliminary mandatory study, we wish to build the entire set of a specific kind of spatio-temporal configurations mixing topological and temporal information. For this purpose, we use Allen's time interval and a new spatial logic based on topology called spatial states. We have decided to start with spatio-temporal configurations between two moving points in a 1D space. This will be extended to 2D and 3D spaces and later, extended to other spatial objects (lines, regions and bodies). Further the establishment of such exhaustive configurations, our aim is to obtain in the future a framework allowing evaluating the relevance of spatio-temporal models. In other words, checking if a given model allows or not to retrieve all the possible spatio-temporal configurations.

The paper is structured as follow. First we make a brief description of models and concepts used thereafter. Then, we expose our general research objectives. After, we develop the approach used to build the spatio-temporal configurations and finally, we conclude.

2 Spatial, temporal and spatio-temporal reasoning

2.1 Spatial Reasoning

Most common spatial reasoning models are based on topology. We wish to cite in this section the 9-i model [9] and the RCC model [10, 11]. The former one is based on the study of intersections between spatial objects topological primitives (see figure 1). The latter is based on the Clarke's connectivity relationship. Both of them gave equivalent sets of topological relationships for regions.

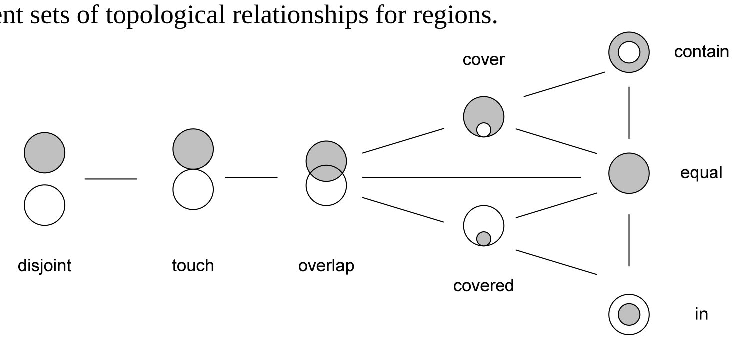


Fig. 1. Topological relationships between two regions in conceptual neighbourhood diagram [12].

2.2 Temporal Reasoning

Allen's time interval reasoning is the most well-known reasoning model used in time modelling. His theory of action and time proposes a formalism based on a temporal logic which is used to represent and reason about events, action, beliefs, intentions, causality, and serve as a framework for solving problems [13]. The time primitive used by Allen is the interval. On this basis, he introduced a set of thirteen mutually exclusive binary relations between intervals (figure 2). The time is assumed to be linear, dense and consequently infinite in the past and future.

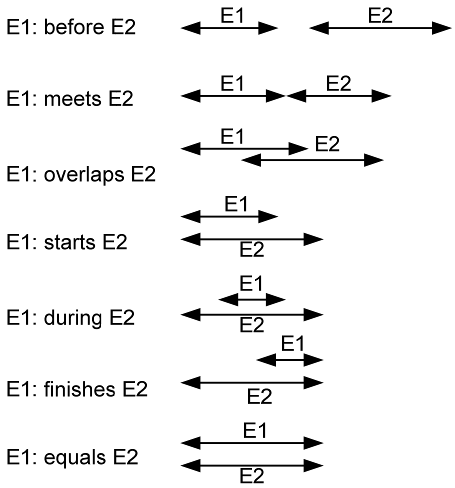


Fig. 2. Thirteen interval relationships defined by Allen (from [14]).

2.3 Spatio-temporal Reasoning

A logical evolution after putting so much effort in spatial and temporal reasoning was to combine them to obtain spatio-temporal reasoning models [15,16]. Spatio-temporal representing and reasoning can be envisaged in two different ways [8]. First one is the combination of a spatial logic with a temporal logic. Some spatial snapshots are combined in a temporal reasoning to derivate spatio-temporal information. The second one is to view the world as spatio-temporal histories and create new reasoning based on spatio-temporal entities [17].

Some of the most achieved realizations illustrate this duality. First, Wolter and Zakharyaschev propose in [15] to combine the constraints formalism RCC-8 with the Propositional Temporal Logic (PLT) [13]. Gerevini and Nebel combine the RCC-8 with Allen's Interval Calculus that is closer in spirit from RCC-8 than PLT, focusing on the computational complexity of such approach.

Combining both approaches, Claramunt and Jiang cross topological relationships and Allen's time intervals in [1, 18, 19] to deduce spatio-temporal histories of static objects (segments and regions). They have defined a temporal region as a region of space valid for a convex temporal interval (see figure 3). This idea can be used for regions but also for points.

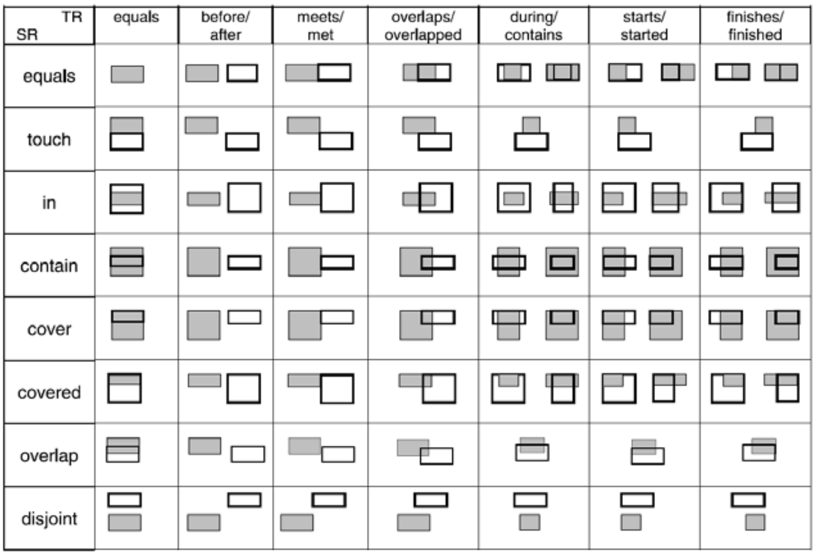


Fig. 3. Visual presentation of relationships in a 2-dimensional space (from [1]).

Finally, Muller considers space-time histories of objects as primitive entities to analyse directly spatio-temporal shapes or histories. He defines a specific space-time to characterise classes of spatial changes [20], which is the first full mereotopological theory based on space-time as a primitive. More recently the works of Hazarika concern a better understanding of spatio-temporal histories continuity [8]. It is worth mentioning a new model, the QTC [6, 21] dealing with direction, speed and acceleration information between moving points.

3 General research objectives

Our research is inspired originally by Claramunt's and Muller's works and considers spatio-temporal space as a primitive space. Our main research objective is to use topology to express relationships between spatio-temporal histories of moving objects. Indeed, by considering a primitive space (spatial and temporal dimensions are not differentiates) [17], we end up considering two lines (the life-lines) and their topological relationships (see [22] for preliminary research objectives). In other words, lifelines are just considered as normal lines in a 2D space. It is therefore possible to analyse topological relationships between them (topological relationships between lines in a 2D space). We aim to end up with a generalized spatio-temporal calculus based on a set of topological relationships between spatio-temporal histories, which should beneficiated of existing topological calculi

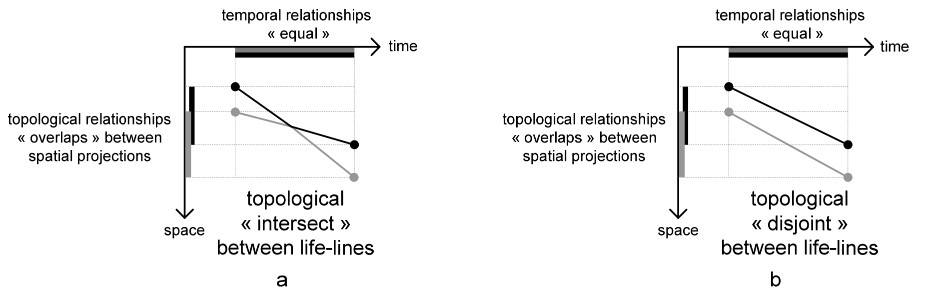


Fig. 4. Examples of spatio-temporal information extraction based on topological relationships between life-lines. This space-time representation is explained in section 4.

Figure 4 illustrates the type of information we wish to extract from such model. In figure 4 a, the "intersect" topological relationship between the two lifelines indicates without ambiguity that there is collision between the two points (spatial and temporal meeting between the two objects), when in figure 4 b., the "disjoint" topological relationship indicates no collision. Note that studying the projections of life-lines on the temporal and spatial axis does not allow differentiating the two behaviors and therefore does not provide enough information to detect a collision.

Beyond these examples, we believe that others topological relationships might have spatio-temporal meaning and could of some use for others analyses as crime mapping or epistemology. We think also that one of the major interests of this approach could be the generalization of all the possible configurations into a smaller set of topological relationships (33 in the lines case [9]). Assuming that enough spatio-temporal meaning would be associated to these relationships, we could use existing topological models and calculi and hopefully increase speed analysis and understanding of spatio-temporal configurations.

The aim of the present paper is to build the entire set of spatio-temporal configurations between two points in a one dimensional space. Beyond the interest of getting these configurations *per se*, this will help us to study the relevance of different spatio-temporal models to retrieve spatio-temporal configurations. This will also be useful when studying generalisation processes.

The next sections describe our spatio-temporal configurations building approach.

4 Building of a set of spatio-temporal configurations

This section presents an approach allowing extracting spatio-temporal configurations between two dynamic points from topological and temporal information. Considering degenerate notions of topological relationships between two points (see section 4.1) and the well-known Allen time intervals, we derive all the possible (in respect to these concepts at least) spatio-temporal configurations between two dynamic points. At this stage of our research, we have decided to start with a simple case; dynamic points in a 1D space. Points are the simplest spatial objects (0D) and they could not move in a space lower than 1D. We assume that points can not go back in the past, i.e. the temporal dimension is oriented (in accordance with Allen's theory).

Practically, with one spatial and one temporal dimension, we can plot a 2-dimensional space with one dimension attributed to each axis. This space is called a temporal space in accordance to Claramunt [18].

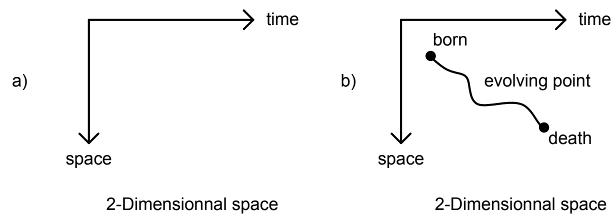


Fig. 5. 2-dimensionnal temporal space with the evolution of a 1-dimensional object.

The existence of a point in this space will be represented by a line-segment. The beginning and the end of the line-segment correspond respectively to the "born" and the "death" of the point-life. This representation is called spatio-temporal history or life-line in the dynamic's point case. Both terms will be used in this paper. All the future representations will be plotted with the same convention and orientation axis as in figure 5 Note that we do not want to impose continuity of spatio-temporal histories. This assumption could be added for specific applications if needed.

4.1 Topological relationships and degenerate topological relationships

The spatial relationships considered here are topological relationships. First, we know that they are two possible topological relationships between two points; disjoint or equal (figure 6, cases 1 and 2).

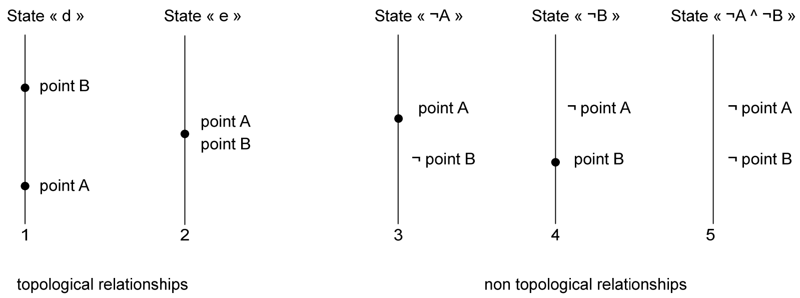


Fig. 6. Representation of the 5 different topological and degenerate topological states between points.

To fully encompass spatio-temporal information complexity, we wish to propose degenerate cases of topological relationships between points. The underlying idea is that at certain moments in time, when considering the life-line of two points A and B, point A or point B might not exist. In such cases, binary topological relationships are no longer valid. Therefore, we propose to consider three other "states" in addition to the two topological "states" (disjoint and equal) which cover all the cases of existence or non existence of points. In this context, a "state" is a particular relationship between objects at a given time. This concept is therefore time independent. The cases 3 to 5 from figure 6 illustrate the three "non-topological" states: "¬B" when point B does not exist, "¬A" when point A does not exist and "¬A ∧ ¬B" when none of them exist.

The set of states "d", "e", "¬A", "¬B", "¬A ∧ ¬B" is a Jointly Exhaustive and Pairwise Disjoint (JEPD) set of topological and degenerate topological relationships.
In the decision tree (figure 7), one can find also the state "t" which means that the two points exist and have a topological relationship and the state "¬t" gathering the non topological states. We believe that such concepts correspond to a lot of real cases, just mention the analysis of moving GPS antennas with some cycle slips.

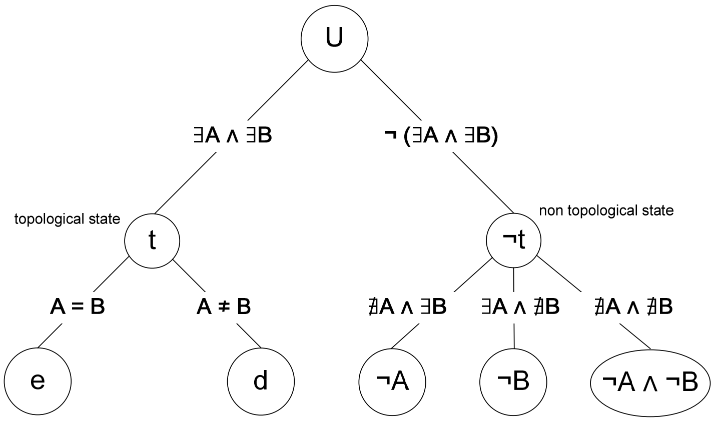


Fig. 7. Decision tree representing the JEPD set of spatial states.

A spatio-temporal configuration can be seen as a succession of different states in time. The study of spatio-temporal histories successive states transitions is out of the scope of this paper. In the following, when representing the life-line of a dynamic point A, we will join successive states where the point A exists.

4.2 Combination of states: the tuple

The method we used to obtain the entire spatio-temporal configuration set is based on the mapping of spatial information (different states) and a temporal logic (Allen's time interval in this case). To be able to combine them, we compose states in a structure called the "tuple". A tuple is defined as a combination of *n* states, where *n* is an integer and represents the level of the tuple. Let ε be the set of possible states values: "e", "d", "¬A", "¬B" and "¬A ∧ ¬B", a tuple of level *n* is denoted as $t^n\{\varepsilon_1, \ldots, \varepsilon_n\}$ with $\varepsilon_1, \ldots, \varepsilon_n \in \varepsilon$. The major interest of this combination is that there is no order between the different states ($\varepsilon_1, \ldots, \varepsilon_n$). Indeed, if ordered the combination of different states may include temporal information, e.g. the succession of the three states "e", "d" and "$\neg_A$" in time lead to a temporal relationship "starts" only (figure 8). The order of the states in tuple is obtained by crossing it with temporal relationships.

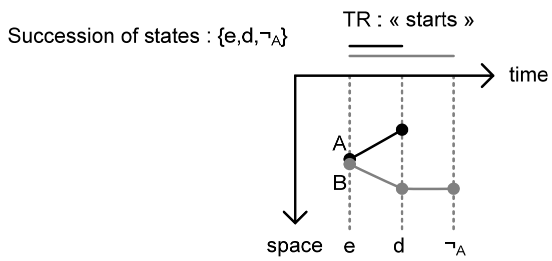


Fig. 8. Succession of states {e,d, $\neg_A$} inducting temporal relationships.

Note that the states "¬t" and "t" could be used to provide a greater level of generalization.

4.3 Spatio-temporal configuration: mapping of tuples with time

Crossing spatial states tuple axis S with Allen's intervals axis T, we can map a spatio-temporal space containing spatio-temporal configurations. Figure 9 shows an example of creation. Let's consider the tuple t_3{e,d,$\neg_A$} combined with temporal relationships "starts". Theoretically, it corresponds to 6 possible arrangements of spatial states: {e,d, $\neg_A$}, {d,e, $\neg_A$}, {$\neg_A$,e,d}, {$\neg_A$,d,e}, {e,$\neg_A$,d}, {d,$\neg_A$,e}. By combining these states with Allen's time intervals, we need to impose continuous life-line, the last 2 cases {e, $\neg_{A,}$ d}, {d, $\neg_{A,}$ e} must be withdrawn. In this particular case, the combination of the tuple t_3{e,d,$\neg_A$} with the temporal relationships "start" lead us to select the two cases where the $\neg_A$ state is at the end of the state's succession (squared in white on figure 9).

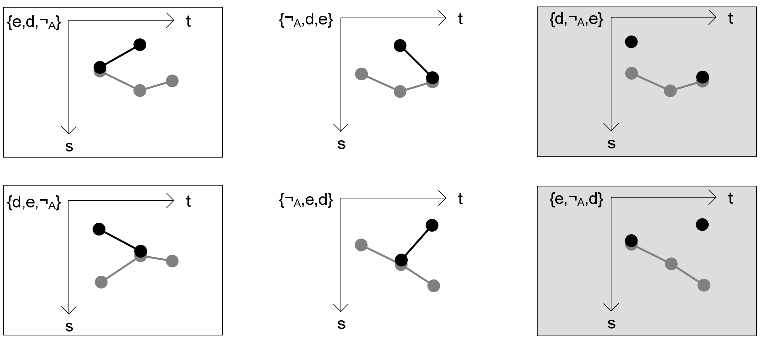


Fig. 9. Possible combination of the 3-tuple t^3{e,d, $\neg_A$}, the spatio-temporal configuration squared in white are the only two valid when crossing the tuple with temporal relationships "starts".

In a similar way, we have derived all the spatio-temporal configurations (279 for level 4) for dynamic points in 1D with Allen's intervals continuity assumption. It appears that it was necessary to consider level 4 tuples and combining them with the entire set of temporal relationships. A level less than 4 cannot be combined with temporal relationships as "overlaps" or "overlapped". Working with upper levels than 4 seems to be just a combination of smaller levels, however for future analyses we believe that considering level 6 tuple would be necessary. Figure 10 presents an extract of spatial configurations derived from level 4 tuple.

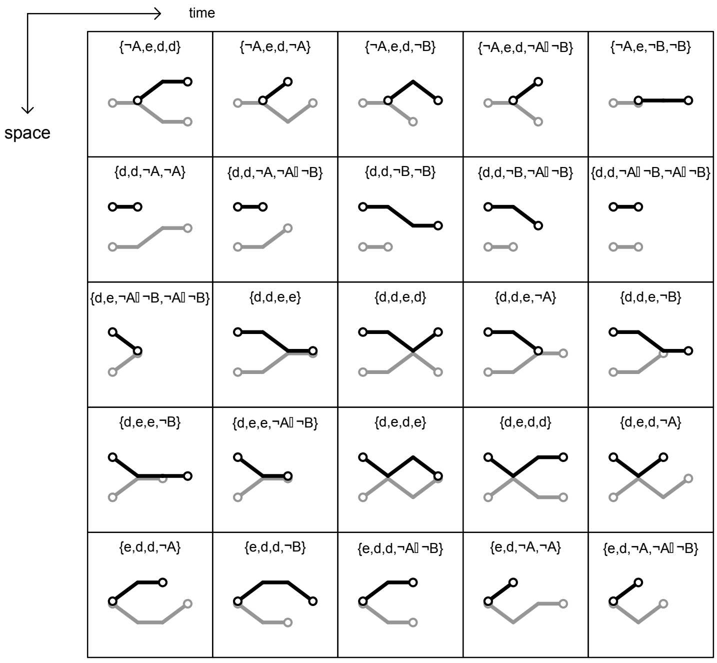


Fig. 10. Extract of spatio-temporal configuration generated from tuple t^4 with continuity assumption.

The non continuous histories can be derived from the continuous histories configurations. Figure 11 represents an extract of the possible non continuous histories; the all set of (625) being accessible at the following address: http://www.geo.ulg.ac.be/hallot/.

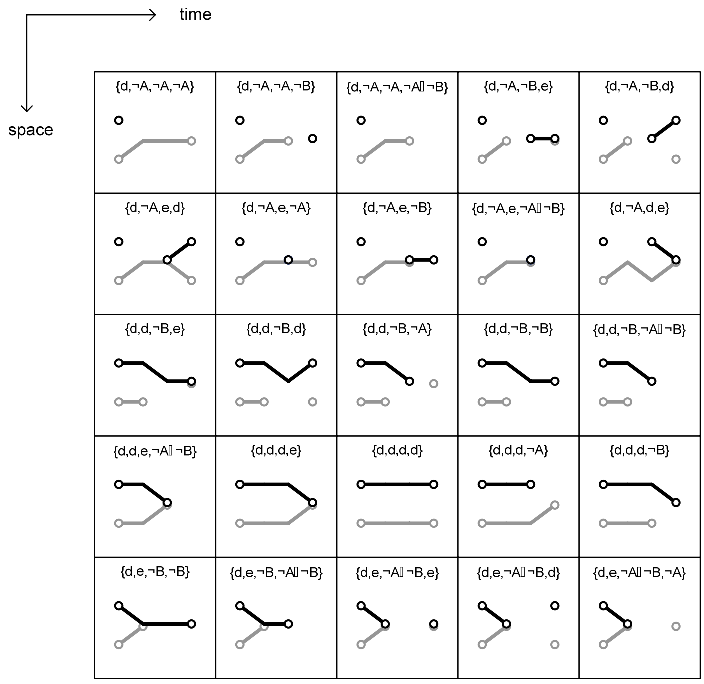


Fig. 11. Extract of spatio-temporal configuration generated from tuple t^4 without continuity assumption.

5 Conclusions

Spatio-temporal reasoning models aim to describe the real world dynamic phenomena's. They can be of two kinds: either they mix spatial or temporal reasoning model or they describe directly new spatio-temporal mereotopology [8]. In this paper, we have developed an innovative approach using known spatial (topology) and temporal (Allen's time intervals) logics to build a specific set of possible spatio-temporal configurations. Building this set of spatio-temporal configurations is the first step of a global research aiming to develop a generalized spatio-temporal reasoning model. Such model, briefly sketched in this paper, aims to extract spatio-temporal information from life-lines by considering primitive space topological calculi. We wish to end up with a set of topological relationships containing enough spatio-temporal meaning to perform relevant spatio-temporal analyses. Getting these configurations *is a necessary step to* study the relevance of such kind of spatio-temporal model.

We start from the definition of degenerate topological relationships between two points allowing relationships between non coexistent points. Combined with topological relationships, we obtain a JEPD set of spatial states which are particular relationships between objects at a given time. Spatio-temporal histories can be seen as a succession of states. After, we define a time free combination of states called "tuple". This new representation of two moving points spatiality is crossed with a temporal logic (Allen's time intervals) to create the entire set of spatio-temporal configurations (279 for level 4).

In the future we wish to extend the spatio-temporal configurations to higher dimensions and to other types of spatial objects. Then, we plan to develop further the generalized model and testing its relevance using real data (GPS).

Finally, we believe that our approach could be also complementary to existing qualitative spatial reasoning models. For instance, the Qualitative Trajectory Calculus [6, 21, 25] is based on analysis of direction, speed and acceleration between two dynamic points. Such calculus can only be used when the two points are coexisting. Our approach could be a nice preliminary analysis to select only the cases where QTC can be used.

6 Bibliography

1. Claramunt, C., Jiang, B.: An integrated representation of spatial and temporal relationships between evolving regions. Journal of Geographical Systems **3** (2001) 411-428
2. Claramunt, C., Jiang, B., Bargiela, A.: A new framework for the visualisation of urban traffic data. Transportation Research Part C **8** (2000) 167-184
3. Erwig, M., Schneider, M.: A visual language for the evolution of spatial relationships and its translation into a spatio-temporal calculus. Journal of Visual Languages and Computing **14** (2003) 181-211
4. Sinha, G., Mark, D.M.: Measuring similarity between geospatial lifelines in studies of environmental health. Journal of Geographical Systems **7** (2005) 115-136
5. Thériault, M., Claramunt, C., Villeneuve, P.Y.: A Spatio-temporal taxonomy for the Representation of Spatial Set Behaviours. In: Bèohlen, M.H., Jensen, C.S., Scholl, M.O. (eds.): Spatio-temporal database management : International Workshop STDBM'99. Springer, Edinburgh, Scotland (1999) 18
6. Van de Weghe, N., Cohn, A.G., De Tre, G., De Maeyer, P.: A Qualitative Trajectory Calculus as a Basis for Representing Moving Objects in Geographical Information Systems. Control and Cybernetics **35** (2006)
7. Laube, P., Kreveld, M., Imfeld, S.: Finding REMO — Detecting Relative Motion Patterns in Geospatial Lifelines. Developments in Spatial Data Handling (2005) 201-215
8. Hazarika, S.M.: Qualitative Spatial Change: Space-time Histories and Continuity. School of Computing, Vol. PhD. University of Leeds, Leeds (2005) 250
9. Egenhofer, M., Herring, J.: Categorizing Binary Topological Relations Between Regions, Lines and Points in Geographic Databases. Technical Report. Department of Surveying Engineering, University of Maine (1990) 28
10. Cohn, A.G.: Calculi for Qualitative Spatial Reasoning. In: Clamet, J., Campell, J., Pfalzgraf, J. (eds.): Artificial Intelligence and Symbolic Mathematical Computation. Springer-Verlag, Berlin (1996) 124-143
11. Cohn, A.G., Bennett, B., Gooday, J.M., Gotts, N.M.: Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. Geoinformatica **1** (1997) 275-316
12. Freksa, C.: Conceptual Neighborhood and its role in temporal and spatial reasoning. In: Singh, M., Travé-Massuyès, L. (eds.): Proceedings of the IMACS Workshop on Decision Support System and Qualitative Reasoning, Amsterdam, Holland (1991) 181-187
13. Pani, A.K., Bhattacharjee, G.P.: Temporal representation and reasoning in artificial intelligence: A review. Mathematical and Computer Modelling **34** (2001) 50-80
14. Allen, J.F.: Time and time again: the many ways to represent time. International Journal of Intelligent Systems **6** (1991) 341-355
15. Wolter, F., Zakharyaschev, M.: Spatio-temporal representation and reasoning based on RCC-8. Seventh Conference on Principles of Knowledge Representation and Reasoning, KR2000. Morgan Kafmann, Breckenridge, USA (2000) 3-14
16. Gerevini, A., Nebel, B.: Qualitative Spatio-Temporal Reasoning with RCC-8 and Allen's Interval Calculus: Computational Complexity. ECAI 2002. IOS Press (2002) 312-316
17. Muller, P.: Éléments d'une théorie du mouvement pour la formalisation du raisonnement spatio-temporel de sens commun. Institut de recherche en informatique de Toulouse. Université Paul Sabatier, Toulouse (1998) 219

18. Claramunt, C., Jiang, B.: A representation of relationships in temporal spaces. In: Atkinson, P., Martin, D. (eds.): Innovations in GIS VII: GeoComputation, Vol. 7. Taylor & Francis, London (2000) 41-53
19. Claramunt, C., Thériault, M., Parent, C.: A qualitative representation of evolving spatial entities in Two-dimensional Topological Spaces. In: Carver, S. (ed.): Innvation in GIS V, Vol. 5. Taylor & Francis, London (1998) 119-129
20. Muller, P.: Topological Spatio-Temporal Reasoning and Representation. Computational Intelligence **18** (2002) 420-450
21. Van de Weghe, N.: Representing and Reasoning about Moving Objects: A Qualitative Approach (Volume I). Department of Geography - Faculty of Sciences. Ghent University, Ghent (2004) 168
22. Hallot, P.: Spatio-temporal relationships in a primitive space: an attempt to simplify spatio-temporal analysis. In: Raubal, M., Miller, H.J., Frank, A., Goodchild, M. (eds.): Fourth International Conference Geographic Information Science, GIScience 2006, Vol. 28. IfGI prints, Münster (2006) 277-283
23. Allen, J.F.: Towards a general theory of action and time. Artificial Intelligence **23** (1984) 123-154
24. Allen, J.F.: Maintaining Knowledge about Temporal Intervals. Communications of the ACM - The University of Rochester **26** (1983) 832-843
25. Van de Weghe, N., Cohn, A.G., Bogaert, P., De Maeyer, P.: Representation of moving objects along road network. 12th International Conference on Geoinformatics - Geospatial Information Research: Bridging the Pacific and Atlantic, University of Gäve, Sweden (2004) 187-194

How to Handle Incomplete Knowledge Concerning Moving Objects

Nico Van de Weghe[1], Peter Bogaert[1], Anthony G. Cohn[2], Matthias Delafontaine[1], Leen De Temmerman[1], Tijs Neutens[1], Philippe De Maeyer[1], and Frank Witlox[1]

[1]Department of Geography, Ghent University, Krijgslaan 281 (S8), B-9000 Ghent, Belgium
Corresponding author: nico.vandeweghe@ugent.be
[2] School of Computing, University of Leeds, Leeds LS2 9JT, United Kingdom

Abstract. In this paper, we present a way of how to handle incomplete knowledge concerning moving objects. Our approach is based on the basic Qualitative Trajectory Calculus (QTC_B), which is a calculus for handling interactions between moving point objects (MPO's). Without elaborating on the domain of linguistics, we show that QTC_B is well-fitted to represent spatio-temporal natural language. Illustrative examples on how to deal with incomplete knowledge are presented.

1. Introduction

In the last two decades, qualitative formalisms suited to express qualitative temporal (e.g.: Freksa, C., 1992a) and spatial relationships (e.g.: Randell, D., Cui, Z., and Cohn, A.G., 1992) between entities have gained wide acceptance as a useful way of abstracting from the real world. Only in recent years, attention has been extended to applications that involve spatio-temporal data. Nevertheless, a variety of research communities have been studying movements of objects, e.g.: Wolfson, O., Xu, B., Chamberlain, S., and Jiang, L., 1998; Erwig, M., Güting, R.H., Schneider, M., and Vazirgiannis, M., 1999; Fernyhough, J.H., Cohn, A.G., and Hogg, D.C., 2000; Nabil, M., Ngu A., and Shepherd A.J., 2001; Pfoser, D., 2002. Until now, the spatio-temporal community has paid little attention to the qualitative aspects.

Apart from some limiting cases, such as a car accident and a predator catching a prey, where moving objects *meet*, mobile objects are represented by the relation *disjoint* in calculi defining topological relations, such as RCC (Randell, D., Cui, Z., and Cohn, A.G., 1992). This approach ignores some important aspects of reasoning about continuously moving physical objects. For example, given two trains on a railroad, it is of the utmost importance to know their movement with respect to each other, in order to detect whether or not they could crash in the near future. Thus, the inherent property with topological theories is that they put all *disjoint* relations into one undifferentiated set. Therefore, a challenging question remained largely unaddressed: 'How do we handle changes in movement between moving objects, if there is no change in their topological relationship?' With this in mind, and starting from the idea that the enormous complexity of interacting real world objects can be described by the relations between pairs of interacting point objects being constantly *disjoint*, the Qualitative Trajectory Calculus (QTC) was introduced by Van de Weghe Van de Weghe, N., 2004). QTC is a theory for representing and reasoning about movements of objects in a qualitative framework, able to differentiate between groups of disconnected objects. Depending on the level of detail and the number of spatial dimensions, different types of QTC were defined all belonging to QTC-Basic (QTC_B) (Van de Weghe, N., Cohn, A.G., De Tré, B., and De Maeyer, Ph., 2006) or QTC-Double Cross (QTC_C) (Van de Weghe, N., Cohn, A.G., De Maeyer, Ph., and Witlox, F., 2005). The reasoning power of QTC has been worked out, applying important reasoning techniques, such as conceptual neighbourhood diagrams (Van de Weghe, N. and De Maeyer, Ph., 2005) and composition tables (Van de Weghe, N., Kuijpers, B., Bogaert, P., and De Maeyer, Ph., 2005). In this paper, the focus is on the feasibility of QTC_B to handle incomplete knowledge[1]. Without elaborating on the domain of linguistics, we show that QTC_B is well-fitted to represent spatio-temporal natural language.

After an explanation of incomplete knowledge and how it is related to qualitative reasoning, a brief overview of QTC_B is presented. Section 4 presents illustrative examples on how to handle incomplete knowledge within the different types of QTC_B. Section 5 concludes the paper and gives some directions for further research.

2. Qualitative Reasoning and Incomplete Knowledge

Reasoning can be performed on quantitative as well as on qualitative information. According to Goyal (Goyal, R.K., 2000), a predefined unit of a quantity is used, typically when working with quantitative information. In the qualitative approach, continuous information is discretised by landmarks separating neighbouring open intervals, resulting in discrete quantity spaces (Weld, D.S. and de Kleer, J., 1990). The major idea in the qualitative approach is that only relevant distinctions are made (Clementini, E., Di Felice, P., and D. Hernandez, 1997). Thus, qualitative reasoning only studies the essence of information, represented as a small set of symbols such as the quantity space {–, 0, +} consisting of the landmark value 0 and its neighbouring open intervals $]-\infty,0[$ and $]0,\infty[$ represented respectively by the symbol – and + (Cohn, A.G. and Hazarika, S.M., 2001).

Not always everything has to be known about a situation to make inferences which are important for the specific study (Frank, A.U., 1996). Obviously in such situations sometimes information lacks for giving complete answers to queries. However, like Freksa (Freksa, C., 1992a, p.203) states, '*a partial answer may be better than no answer at all.*' By abstracting away from metrical details, qualitative representations are much more appropriate for handling such incomplete knowledge than quantitative methods (Cristani, M., Cohn, A.G., and Bennett, B., 2000).

The development of the Qualitative Trajectory Calculus (QTC) has been inspired by some important qualitative calculi in temporal and spatial reasoning, especially the temporal Semi-Interval Calculus (Freksa, C., 1992) and the spatial Double-Cross

---

[1] Knowledge only containing one relation in a specific calculus is called complete or fine knowledge. A union of fine relations results in incomplete knowledge.

Calculus (Freksa, C., 1992b; Zimmermann, K. and Freksa, C., 1996). Central in these theories is the specific attention to incomplete knowledge, for example produced by natural language expressions. In combination with the inherent capability of the qualitative calculi lying at the basis of QTC, one might expect that QTC ought to be able to handle incomplete knowledge.

3. The Qualitative Trajectory Calculus – Basic (QTC_B)

In this section, an informal account of the Qualitative Trajectory Calculus – Basic (QTC_B) is presented. For a formal axiomatisation, we refer to (Van de Weghe, N., 2004). Continuous time for QTC_B is assumed. In general, QTC_B compares positions of two objects at different moments in time. The movement of the first object (called *k*) with respect to the second object (called *l*) is studied by comparing the distance between *l* at the current time point (denoted *t*) and *k* during the period immediately before the current time point (denoted t^-), with the distance between *l* at *t* and *k* during the period immediately after the current time point (denoted t^+). In addition, the movement of *l* with respect to *k* is studied by comparing the distance between *k* at *t* and *l* at t^-, with the distance between *k* at *t* and *l* at t^+. QTC_{B1D} handles the qualitative movement of two constantly *disjoint* point objects restricted to 1D. Because the movement is restricted to 1D, the velocity vector of an object only has two possible directions, with the intermediate case where the object stands still. Hence, the direction of the movement of each object can be described by one single qualitative variable. Both degrees of freedom can be further subdivided according to the relative speed of the objects. This subdivision results in redundant information because the relative speed of *k* with respect to *l* is the inverse of the relative speed of *l* with respect to *k*. By reducing the continuum to the qualitative values –, 0 and +, the underlying continuous system can be described discretely. We introduce the following notation for QTC_{B1D}:

$x|t$ denotes the position of an object *x* at time *t*,
$d(u,v)$ denotes the distance between two positions *u* and *v*,
$v_x|t$ denotes the speed of *x* at time *t*,
$t_1 < t_2$ denotes that t_1 is temporally before t_2.

A movement is presented in QTC_{B1D} using the following four conditions (C):

C1. Movement of *k* with respect to the position of *l* at *t* (distance constraint):
–: *k* is moving towards *l*:

$$\exists t_1\, (t_1 < t \wedge \forall\, t^-\, (t_1 < t^- < t \;\rightarrow\; d(k|t^-, l|t) > d(k|t, l|t)))\ \wedge$$
$$\exists t_2\, (t < t_2 \wedge \forall\, t^+\, (t < t^+ < t_2 \;\rightarrow\; d(k|t, l|t) > d(k|t^+, l|t)))$$

+: *k* is moving away from *l*:

$$\exists t_1\, (t_1 < t \wedge \forall\, t^-\, (t_1 < t^- < t \;\rightarrow\; d(k|t^-, l|t) < d(k|t, l|t)))\ \wedge$$
$$\exists t_2\, (t < t_2 \wedge \forall\, t^+\, (t < t^+ < t_2 \;\rightarrow\; d(k|t, l|t) < d(k|t^+, l|t)))$$

0: *k* is stable with respect to *l* (all other cases):
all other cases

C2. The movement of *l* with respect to the position of *k* at *t* (distance constraint) can be described as in C1 with *k* and *l* interchanged.

C3. Relative speed of *k* at *t* with respect to *l* at *t* (which dually represents the relative speed of *l* at *t* with respect to *k* at *t*) (speed constraint):
$-$: $v_k|t < v_l|t$ $+$: $v_k|t > v_l|t$ 0: $v_k|t = v_l|t$

Accordingly, a qualitative trajectory pair can be represented by a label consisting of two or three characters, for respectively QTC_{BL1} (QTC_B of level one) only handling the changing distance between two objects and QTC_{BL2} (QTC_B of level two) also taking into account the third label representing the relative speed of both object with respect to each other. In theory, there should be 27 (3^3) *B12-relations* (QTC relations of level two in 1D). As illustrated in Fig. 1A, 10 relations are impossible (e.g. relation 2b: if object *k* moves towards object *l* and object *l* stands still, then $v_k < v_l$ is impossible). Therefore, we get only 17 B12-relations. Each icon in Fig. 1A represents one single relation, and therefore is called a *relation icon*, in this particular case a *B12-relation icon*. The left and the right dot of the B12-relation icon respectively represent the positions of *k* and *l*. The line segments represent whether each object can be moving towards or away from the other. A dot is filled if the object can be stationary, and open if an object cannot be stationary. The representations are no more than icons, in which we assume that *k* is on the left side of *l*.

The approach for 1D can be successfully used for higher dimensions by denoting the Euclidean distance between a pair of point objects as being the only dimension. This way 2D and even 3D movements can be reduced to 1D movements. To emphasise that we are working on 2D movements, the theory is called QTC_{B2D}. The definitions for the 2D movement are the same as the definitions for the 1D movement. In contrast with QTC_{B12}, there are 27 potential *B22-relations*, represented as 27 *B22-relation icons* in Fig. 1B. If, for example, the first character of the B22-relation is 0, then the first object stands still or can move tangentially with the second object. The icons contain line segments with the point object in the middle of it. The line segment stands for the opportunity to move to both sides of the point object. A filled dot represents the case when the object can be stationary. An open dot means that the object cannot be stationary. The icons also contain crescents with the point object in the middle of its straight border. If a crescent is used, then the movement starts in the dot and ends somewhere on the curved side of the crescent. It is important that the crescent is an open polygon: the straight boundary of a crescent is an element of another relation. Of major importance is that, in contrast to $QTC_{B1D,}$ all 27 relations are possible. The reason for this is quite straightforward. In 1D, an object can only move along a straight line. On the other hand, in 2D an object can move throughout the complete 2D space, being a higher dimension than the 1D distance. Therefore, there is a higher degree of freedom in B2D-movements compared to B1D-movements, resulting in the different number of possible relations.

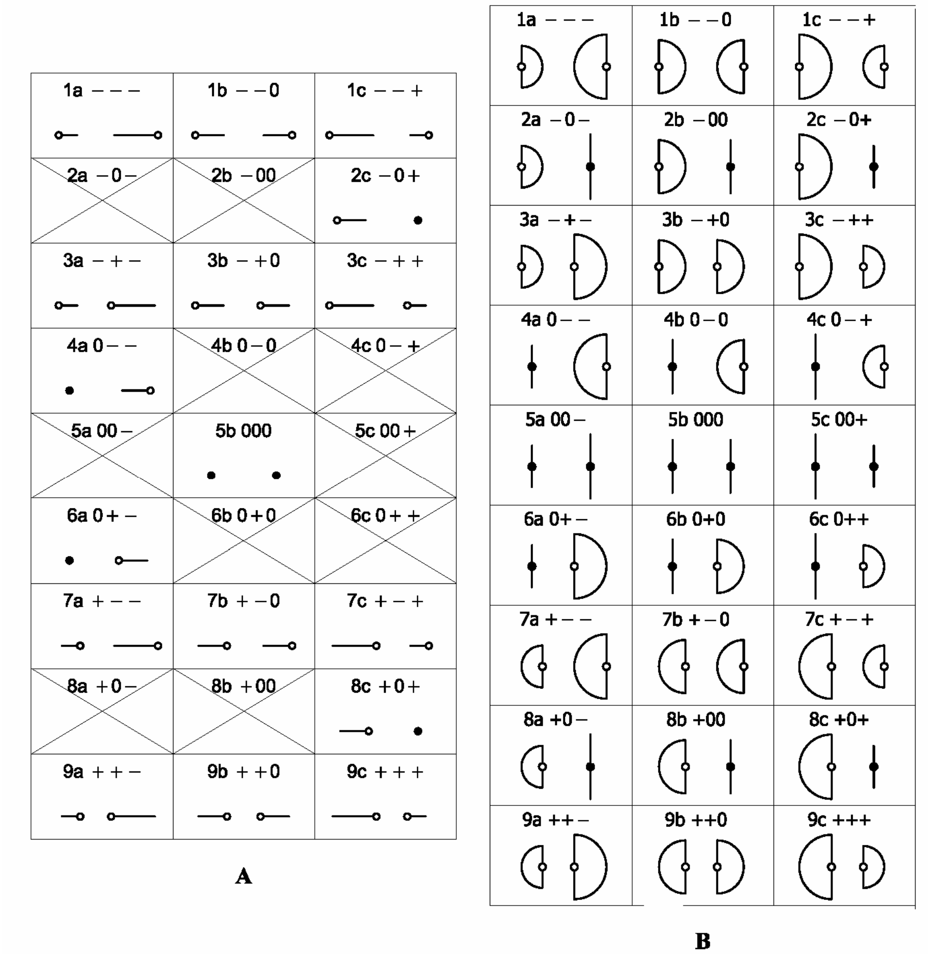


Fig. 1: (A) B12-relation icons, (B) B22-relation icons

4 Incomplete Knowledge about Moving Objects Handled Naturally

In common with qualitative spatial and temporal calculi, we need to consider that we do not always have complete knowledge about which relation holds between a pair of moving objects. In this section, illustrative examples on how to handle incomplete knowledge within QTC_B are presented. *Expressions in natural language* (*Ex*), about the movement of two objects (*k* and *l*) with respect to each other, are considered. We determine which QTC_B relations hold for each particular expression. We use the standard notation for *implication* and *equivalence*:

$a \rightarrow b$: if *a*, then *b*

$a \leftrightarrow b$: if and only if *a*, then *b*

as well as the standard notations for the following set operations:

$a \setminus b$: *a* minus *b*

$a \cap b$: intersection of *a* and *b*

4.1. From Fine to Incomplete Knowledge and Vice Versa

This example starts from *Ex1* forming fine knowledge concerning moving objects, and relaxes the constraints in order to get incomplete knowledge (*Ex2*, *Ex3* and *Ex4a*). Thereafter, the inverse approach is discussed. Starting from several incomplete constraints (*Ex4a*, *Ex4b*, *Ex4c*, and *Ex4d*), fine knowledge will be generated by the intersection of the incomplete solutions. The example is worked out for 1D and 2D.

4.1.1. From Fine to Incomplete Knowledge

Ex1: *k* is moving towards *l*, which in turn is moving away from *k*, both objects moving along the same straight line and having the same speed.

$$Ex1 \rightarrow (-\,+)_{B11} \text{ and } Ex1 \leftrightarrow (-\,+\,0)_{B12}$$

QTC_{B11} does not give full detail, because $(-\,+)_{B11}$ also contains situations where *k* and *l* have a different speed. Therefore, it is more appropriate to work at level two, which incorporates the speed variable.

$$Ex1 \rightarrow (-\,+)_{B21} \text{ and } Ex1 \rightarrow (-\,+\,0)_{B22}$$

At first sight, QTC_{B2D} and QTC_{B1D} give the same result. However, there is only an implication ($\rightarrow$) between *Ex1* and $(-\,+\,0)_{B22}$, because $(-\,+\,0)_{B22}$ does not consider the restriction in *Ex1* that both objects are moving along the same straight line, which was an implicit restriction for movements in 1D.

Ex2: *k* is moving towards *l*, which in turn is moving away from *k*, both objects moving along the same straight line.

$$Ex2 \leftrightarrow (-\,+)_{B11} \text{ and } Ex2 \leftrightarrow (-\,+\,A^{2})_{B12}$$

The only difference between *Ex1* and *Ex2* is the speed constraint, which is not given in *Ex2*. In contradiction to *Ex1*, we have in *Ex2*: if $(-\,+)_{B11}$ is true, then *Ex2* must be true. *Ex2* is thus totally covered by QTC_{B11}. The difference between *Ex1* and *Ex2* has perhaps more implications for QTC_{B12}, since $(-\,+\,A)_{B12}$ consists of a disjunction of solutions:

The following statement is false: $Ex2 \rightarrow a$ (with $a \in (-\,+\,A)_{B12}$)

The following statement is true: $a \rightarrow Ex2$ (with $a \in (-\,+\,A)_{B12}$)

$$Ex2 \rightarrow (-\,+)_{B21} \text{ and } Ex2 \rightarrow (-\,+\,A)_{B22}$$

[2] A qualitative variable *A* (*B*, *C*, …) stands for the set {–, 0, +}

Ex2 represented in QTC_{B22} gives no extra information compared to *Ex2* represented in QTC_{B21} since the third character of QTC_{B22}, differentiating QTC_{B22} from QTC_{B21}, can have all qualitative values. Note that there is only an implication (→) between *Ex2* and $(- +)_{B21}$, because $(- +)_{B21}$ does not consider the restriction in *Ex2* that both objects are moving along the same straight line, which was an implicit restriction for movements in 1D. The same applies to the implication between *Ex2* and $(- + A)_{B22}$.

Ex3: *k* is moving towards *l*, which in turn is moving away from *k*.

$$Ex3 \leftrightarrow (- +)_{B11} \text{ and } Ex3 \leftrightarrow (- + A)_{B12}$$

Compared to *Ex2*, the objects do not need to move along a straight line. However, this constraint is straightforward, since we are working in 1D.

$$Ex3 \leftrightarrow (- +)_{B21} \text{ and } Ex3 \leftrightarrow (- + A)_{B22}$$

In contrast to the 1D movement, the constraint that both objects have to move on the same straight line (or in 1D) is important in 2D. In *Ex3*, this constraint is deleted, which results in an important extension of the solution set. This extension can be seen in the formulae; on the one hand one gets an implication between *Ex2* and the B2D relations, on the other hand one gets an equivalence between *Ex3* and the B2D relations. This extension can be easily seen by comparing the relation icons for $(- + A)_{B21}$ in Fig. 1A with those for $(- + A)_{B22}$ in Fig. 1B.

Ex4a: *k* is moving towards *l*.

$$Ex4a \leftrightarrow (- A)_{B11} \text{ and } Ex4a \leftrightarrow (- A\ B)_{B12}$$

This expression does not state whether *l* is moving. Because this expression is less complete than *Ex3*, it is obvious that we cannot distinguish QTC_{B11} from QTC_{B12}. However, note in QTC_{B12} that when *l* is not moving, only $(- 0 +)_{B12}$ holds, because $(- 0 -)_{B12}$ and $(- 0\ 0)_{B12}$ are impossible in 1D.

$$Ex4a \leftrightarrow (- A)_{B21} \text{and } Ex4a \leftrightarrow (- A\ B)_{B22}$$

As could be expected, there is no difference between for QTC_{B11} and QTC_{B21}.

4.1.2. From Incomplete to Fine Knowledge

Now, let us start from four expressions (*Ex4a*, *Ex4b*, *Ex4c*, and *Ex4d*), which together form the fine compound expression *Ex1*:

Ex4a: *k* is moving towards *l*.

$$Ex4a \leftrightarrow (- A)_{B11}$$
$$Ex4a \leftrightarrow (- A\ B)_{B12}$$
$$Ex4a \leftrightarrow (- A)_{B21}$$

$Ex4a \leftrightarrow (- A\ B)_{B22}$

Ex4b: *l* is moving away from *k*.

$$Ex4b \leftrightarrow (A\ +)_{B11}$$
$$Ex4b \leftrightarrow (A + B)_{B12}$$
$$Ex4b \leftrightarrow (A\ +)_{B21}$$
$$Ex4b \leftrightarrow (A + B)_{B22}$$

Again, there is no difference between the representations of QTC_{B11} and QTC_{B21}, and those of QTC_{B12} and QTC_{B22}.

Ex4c: *k* and *l* are moving along the same straight line.

$$Ex4c \leftrightarrow (A^{*}\ B^{*})_{B11}$$ [3]
$$Ex4c \leftrightarrow (A^{*}\ B^{*}\ C)_{B12}$$
$$Ex4c \rightarrow (A^{*}\ B^{*})_{B21}$$
$$Ex4c \rightarrow (A^{*}\ B^{*}\ C)_{B22}$$

Because it is specified that both objects are moving, neither of the two objects may stand still.

Ex4d: *k* and *l* have the same speed.

$$Ex4d \rightarrow (A^{*}\ B^{*},\ 0\ 0)_{B11}$$
$$Ex4d \leftrightarrow (A^{*}\ B^{*}\ 0,\ 0\ 0\ 0)_{B12}$$

It is not specified whether the speed has to be higher than zero. Therefore, $(0\ 0)_{B11}$ and $(0\ 0\ 0)_{B12}$ are possibilities. However, since the speed of both objects has to be the same, it is impossible to have a pair of objects where only one object is moving.

$$Ex4d \rightarrow (A\ B)_{B21} \text{ and } Ex4d \leftrightarrow (A\ B\ 0)_{B22}$$

In contrast to QTC_{B11}, every relation is possible in QTC_{B21}, which is a direct result of specifications concerning the exclusive B22-relations.

4.1.3. Overall Result

The intersection of the four solution sets of the expressions *Ex4a*, *Ex4b*, *Ex4c*, and *Ex4d*, gives $(- +)_{B11}$ and $(- + 0)_{B12}$. One can state that the intersection of the solution sets of the components of a compound expression is the same as the solution set of the compound expression.

---

[3] A qualitative variable *A** (*B**, *C**, ...) stands for the set {−, +}

$(-\ A)_{B11} \cap (A\ +)_{B11} \cap (A^*\ B^*)_{B11} \cap (A^*\ B^*,\ 0\ 0)_{B11} = (-\ +)_{B11}$
$(-\ A\ B)_{B12} \cap (A + B)_{B12} \cap (A^*\ B^*\ C)_{B21} \cap (A^*\ B^*\ 0,\ 0\ 0\ 0)_{B12} = (-\ +\ 0)_{B12}$

The intersection of the four solution sets for QTC_{B2D} of each expression is respectively $(-\ +)_{B21}$ and $(-\ +\ 0)_{B22}$. Again (cf. QTC_{B1D}), the intersection of the solution sets of the components of a compound expression is the same as the solution set of the compound expression.

$(-\ A)_{B21} \cap (A\ +)_{B21} \cap (A^*\ B^*)_{B21} \cap (A\ B)_{B21} = (-\ +)_{B21}$
$(-\ A\ B)_{B22} \cap (A + B)_{B22} \cap (A^*\ B^*\ C)_{B22} \cap (A\ B\ 0)_{B22} = (-\ +\ 0)_{B22}$

4.2. How Many Objects Are Moving?

If we say that an object is moving, we can have interpretation problems; do we mean that at least one of the objects is moving, or do we mean that exactly one object is moving? This ambiguity can be overcome by QTC_B.

Ex5: At least one of the objects is moving.

$Ex5 \leftrightarrow (A\ B) \setminus (0\ 0)_{B11}$
$Ex5 \leftrightarrow (A\ B\ C) \setminus (0\ 0\ 0)_{B12}$
$Ex5 \rightarrow (A\ B)_{B21}$
$Ex5 \rightarrow (A\ B\ C)_{B22}$

Due to this expression, it is possible that only one object is moving or it could be that both objects are moving. Note that for QTC_{B21} and QTC_{B22}, the relations where the first and the second character are zero do not need to be excluded since objects can move tangentially when both the first and the second are 0 in 2D.

Ex6: Exactly one of the objects is moving.

$Ex6 \leftrightarrow (A^*\ 0{,}0\ B^*)_{B11}$
$Ex6 \leftrightarrow (A^*0\ B,\ 0\ A^*\ B)_{B12}$
$Ex6 \rightarrow (A\ 0,\ 0\ A)_{B21}$
$Ex6 \rightarrow (A\ 0\ B,\ 0\ A\ B)_{B22}$

Note again the subtle difference between QTC_{B1D} and QTC_{B2D}. In QTC_{B1D}, an object can only move when a character is different from 0. In QTC_{B2D}, an object can move if a character is 0. Note that $(A^*B^*)_{B21}$ is impossible since here both objects are moving.

5. Conclusion

Based on several illustrative examples, the ability of handling incomplete knowledge and natural language expressions within QTC_B is studied. In further research, the possibilities of QTC-Double Cross (QTC_C) to handle incomplete knowledge will be discussed. Since QTC_C considers additionally the direction in which an object is moving with respect to the line segment between the two objects, this calculus is more expressive and will involve more complex reasoning. Note for example that, in contrast with $QTC_{C2D,}$ it is not possible in QTC_{B2D} to denote whether two objects are moving along the same straight line. This will be possible. In the future, we will continue to explore the bridge between natural language, perception and formal ontologies of moving objects.

References

Clementini, E., Di Felice, P., and D. Hernandez, 1997, Qualitative representation of positional information, *AI*, 95(2), 317-356.

Cohn, A.G. and Hazarika, S.M., 2001, Qualitative spatial representation and reasoning: an overview, *Fundamenta Informaticae*, 46(1-2), 1-29.

Cristani, M., Cohn, A.G., and Bennett, B., 2000, Spatial locations via morpho-mereology, *Conf. on Principles of Knowledge Representation and Reasoning*, 15-25.

Erwig, M., Güting, R.H., Schneider, M., and Vazirgiannis, M., 1999, Spatio-temporal data types: an approach to modelling objects in databases, *Geoinformatica*, 3(3), 269-296.

Fernyhough, J.H., Cohn, A.G., and Hogg, D.C., 2000, Constructing qualitative event models automatically from video input, *Image and Vision Computing*, 18(2), 81-103.

Frank, A.U., 1996, Qualitative spatial reasoning: cardinal directions as an example, *International Journal of Geographical Information Science*, 10(3), 269-290.

Freksa, C., 1992a, Temporal reasoning based on semi-intervals, *AI*, 54, 199-227.

Freksa, C., 1992b, Using orientation information for qualitative spatial reasoning, *Int. Conf. on Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, *LNCS*, (639), 162-178.

Goyal, R.K., 2000, *Similarity Assessment for Cardinal Directions between Extended Spatial Objects*, PhD thesis, USA, University of Maine, 167 pp.

Nabil, M., Ngu A., and Shepherd A.J., 2001, Modelling and retrieval of moving objects, *Multimedia Tools and Applications*, 13(1), 35-71.

Pfoser, D., 2002, Indexing the trajectories of moving objects, *IEEE Data Engineering Bulletin*, 25(2), 3-9.

Randell, D., Cui, Z., and Cohn, A.G., 1992, A spatial logic based on regions and connection, *Conf. on Knowledge Representation and Reasoning*, 165-176.

Van de Weghe, N. and De Maeyer, Ph., 2005, Conceptual neighbourhood diagrams for representing moving objects, *Int. Workshop on Conceptual Modeling for Geographic Information System*, *LNCS*, (3770), 228-238.

Van de Weghe, N., Cohn, A.G., De Tré, B., and De Maeyer, Ph., 2006, A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems, *Control and Cybernetics*. (Accepted for publication)

Van de Weghe, N., 2004, *Representing and Reasoning about Moving Objects: A Qualitative Approach*, PhD thesis, Belgium, Ghent University, 268 pp.

Van de Weghe, N., Cohn, A.G., De Maeyer, Ph., and Witlox, F., 2005, Representing moving objects in computer-based expert systems: the overtake event example, *Expert Systems with Applications*, 29(4), 977-983.

Van de Weghe, N., Kuijpers, B., Bogaert, P., and De Maeyer, Ph., 2005, A qualitative trajectory calculus and the composition of its relations, *Int. Conf. on Geospatial Semantics*, *LNCS*, (3799), 60-76.

Weld, D.S. and de Kleer, J., 1990, *Readings in Qualitative Reasoning about Physical Systems*. Morgan Kaufmann, 720 pp.

Wolfson, O., Xu, B., Chamberlain, S., and Jiang, L., 1998, Moving object databases: issues and solutions, *Int. Conf. on Scientific and Statistical Database Management*, 111-122.

Zimmermann, K. and Freksa, C., 1996, Qualitative spatial reasoning using orientation, distance, and path knowledge, *Applied Intelligence*, 6(1), 49-58.

Exploring Context-Sensitivity in Spatial Intention Recognition

Peter Kiefer and Christoph Schlieder

Laboratory for Semantic Information Technologies
Otto-Friedrich-University Bamberg
96045 Bamberg, Germany
{peter.kiefer,christoph.schlieder}@wiai.uni-bamberg.de

Abstract. In its most general form, the problem of inferring the intentions of a mobile user from his or her spatial behavior is equivalent to the plan recognition problem which is known to be intractable. Tractable special cases of the problem are therefore of great practical interest. Using formal grammars, intention recognition problems can be stated as parsing problems in a way that makes the connection between expressiveness and complexity explicit. We argue that context free grammars are not sufficiently expressive to handle important use cases. Furthermore, we identify three types of constraints on the grammar's productions that may arise in spatial intention recognition: rule-at-location constraints, rule-rule-constraints, and complex rule-location-constraints. Finally we show that Tree Adjoining Grammars can be used to handle rule-rule-constraints.

1 Introduction

Intention recognition is the problem of inferring a person's intentions to act from observations of that person's behavior. It can undoubtedly make an information system more intelligent. The system should automatically provide services that support the user's intentions, thus minimizing the necessary human computer interaction (information push). Especially applications used under conditions where the possibilities of human computer interaction are rather limited (such as biking or technical maintenance) should adapt to the limited cognitive resources of a user [1].

The intention recognition problem is also known in the literature as plan recognition problem [2]. Many current plan recognition approaches deal with the intractable form of the problem, and are thus rather suited for running on a server than on a stand-alone mobile client. We focus on the latter since we believe that there is a need for stand-alone mobile applications due to still expensive data tariffs and privacy issues.

Assigning intentions to a sequence of incoming behaviors can be interpreted as a pattern recognition problem. A formalism that has proven well for pattern recognition in areas like machine vision ([3, 4]) are Context Free Grammars (CFG) and their probabilistic counterparts. In addition to polynomial parsing algorithms, such production systems have the advantage of being modular, generative, and intuitive.

Grammatical approaches are also used for intention recognition, but current research in this area agrees that the context-freeness of CFGs is not sufficiently expressive for most intention recognition domains. It seems at least a contradiction in terms to build a system that is context *aware* with a formalism that is context *free*. *Probabilistic State Dependent Grammars* (PSDG) ([5]), for instance, define the probability of a production rule as dependent on a general context variable (state) which is stored additional to the parse tree. This approach is very general and directed towards a large class of plan recognition problems which makes inference not efficient enough for our problem. *Spatially Grounded Intentional Systems* (SGIS) ([6]) include spatial context by making the applicability of a rule dependent on the current spatial region. This specific focus on spatial context makes this approach interesting for stand-alone mobile intention recognition of spatio-temporal behavior. A third line of argumentation targets at utilizing grammar formalisms with more context sensitivity from natural language processing (NLP) ([7]). *Mildly Context Sensitive Grammars* (MCSGs), a class of grammatical formalisms well-known in NLP since the 1980s, are proposed due to structural similarities to intention recognition problems. MCSGs are more expressive than CFGs while still being polynomially parsable which makes them especially attractive for algorithms running on mobile devices with low computational power.

This paper strengthens the argument made in [7] and tries to combine it with the spatial constraints formulated in SGIS [6]. We formulate three kinds of constraints on behaviors typical for intention recognition on spatio-temporal trajectories of a moving agent. This approach is new, in that it combines MCSGs, plan/intention recognition and spatial knowledge modeled by a domain expert. The goal consists in finding a formalism that has enough expressive power needed for the kind of context sensitivity imposed by our domain, while restricting the complexity of inference through spatial knowledge.

The outline of this paper is as follows. Section 2 introduces intention recognition with Context Free Grammars (CFG) and the problem of ambiguity. It is shown how spatial knowledge, like that encoded in SGIS, can help to resolve these ambiguities. Section 3 discusses the need for more context sensitive representations with special consideration of spatial knowledge. An example drawn from a location-based game is formalized using Tree Adjoining Grammars (TAG) which fall in the class of MCSGs. Section 4 sketches connections to related work in intention recognition and motion pattern analysis not mentioned in sections before. This paper describes work in progress, so we finally outline issues for our future work in Section 5.

Production Rules P			
$Trip$	$\rightarrow$	$Pick$ driving $Drop$	(1)
$Pick$	$\rightarrow$	parking	(2)
	|	parking driving parking	(3)
$Drop$	$\rightarrow$	parking	(4)
	|	parking driving parking	(5)

Start	Destination	Observed Behavior Sequence
$P_S \neq P_1$	$P_D \neq P_2$	parking driving parking driving parking driving parking
$P_S = P_1$	$P_D \neq P_2$	parking driving parking driving parking
$P_S \neq P_1$	$P_D = P_2$	parking driving parking driving parking
$P_S = P_1$	$P_D = P_2$	parking driving parking

Fig. 1. A very simple context free production system for intention recognition (top) and the language it defines (bottom).

2 Reducing Ambiguity by Adding Spatial Knowledge

2.1 Ambiguity: A Very Simple Example

To illustrate the basic ideas of this section we refer to the following simple example: suppose we observe an agent on a car trip from P_S to P_D. We know that she is going to pick up a friend at some point P_1 on that trip, and that she is going to drop her passenger at some other point P_2. The observations we get from our sensors are the behaviors $B = \{parking, driving\}$ while the concrete events of getting on or off the car remain hidden. The behaviors B are used as terminals in our context free production system. We want to decide for any element in the behavior sequence whether the driver is still on the way to the pick up point P_1 (having the intention $Pick$), or already accompanied by her friend (having the intention $Drop$). A context free production system sufficiently expressive for this simple use case is listed in Figure 1 (top). A top level intention $Trip$ is introduced as starting symbol, which yields in a set of intentions (non-terminals) $I = \{Trip, Pick, Drop\}$. Note that the two agents might have the same starting point ($P_S = P_1$), or the same destination ($P_D = P_2$). In Figure 1 (top), this is expressed by rules (2) and (4). Rule (2), for instance, states that the start of the journey and the pick up may coincide. Figure 1 (bottom) presents the language defined by our context free production system, i.e. all possible behavior sequences that can be created with rules (1)–(5). From an intention recognition perspective, these are all behavior sequences that can be recognized by the simple grammar. The first two columns indicate which of the four possible cases is covered by the accordant line, with respect to coinciding start/pick-up and destination/drop points.

Formally, we now have defined all necessary components of a Context Free Grammar $CFG_{simple} = (B, I, P, Trip)$, which we may also call an *intentional system* [6]. Recognizing the agent's intentions in this formalism consists in determining the correct parse tree for a sequence of behaviors. Algorithms for

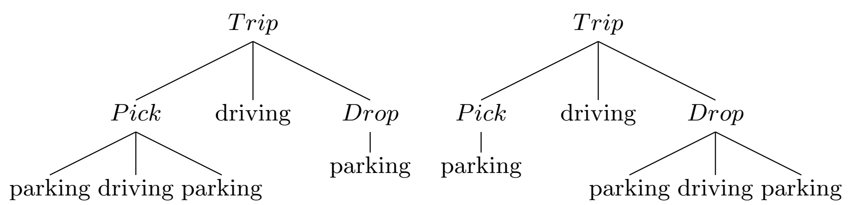


Fig. 2. Ambiguity: two possible parse trees for the same sequence of behaviors.

parsing CFGs, most based on chart-based parsers like the Earley Algorithm [8], are well-known and run in polynomial time. For each behavior b of an observation sequence, we can define the (direct) parent intention in the parse tree as the currently active intention. When building a context-aware service it is those active intentions that must be mapped to an appropriate information service. For instance, the third behavior in the parse trees in Figure 2 (parking) has an active intention *Pick* in the left tree, and *Drop* in the right tree.

The parse trees from Figure 2 visualize the ambiguity that may occur in our example. Without further knowledge we cannot decide which of the two trees is more plausible for the given behavior sequence. The problem of ambiguity is well-known in NLP and typically solved by assigning probabilities to the production rules, yielding in a Probabilistic Context Free Grammar (PCFG) (see [9] for an overview and examples). Rule probabilities for a PCFG in NLP are typically learned from large annotated text corpora. Assigning rule probabilities for an intention recognizing PCFG is much more complicated, especially because behavioral corpora do not exist for most domains. Instead of choosing a probabilistic approach, we rather try to discriminate between different behavior interpretations by using world knowledge, especially knowledge about the spatial structure of the environment, to assist the parsing process.

2.2 Disambiguation with Spatial Constraints

We have already introduced *intentional systems* which are CFGs with behaviors as terminals, and intentions as non-terminals. As a next step, we add spatial knowledge to the production rules and obtain *spatially grounded intentional systems* (SGIS), also defined in [6], as follows:

Definition 1. *Let R denote a set of spatial regions, B a set of behaviors, and I a set of intentions, all three sets being finite. A spatially grounded intentional system $A = (B, I, P, S, G)$ is an intentional system (B, I, P, S) together with a relation $G \subseteq P \times R$ describing the regions in which a production rule is applicable.*

SGISs exploit the fact that possible interpretations for an agent's behavior depend on the space he is currently located or moving in. This basic intuition of connection between intention and place has a long tradition and already appears

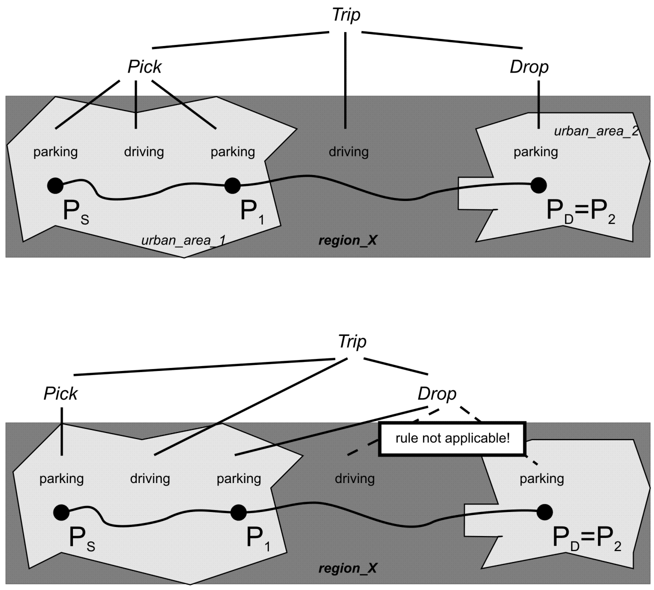


Fig. 3. Spatial disambiguation.

in the first article that has formulated the plan recognition problem [2] where the concept of $LOCATION$ was integrated as basic concept in the $World\ Domain$. For instance, a certain behavior in [2] can only be interpreted as plan $TAKE$, if both, the actor and the object of that possible $TAKE$–plan, are at the same location.

In our simple example, we could have some additional knowledge about the structure of space in which the two friends are traveling. For instance, they might be driving from *urban_area_1* to *urban_area_2* which are both part of $region_X$ (refer to Figure 3). We choose the spatial grounding of the production rules introduced in Figure 1 (top) as follows: rules (2) and (3) are grounded in region *urban_area_1*, rules (4) and (5) are grounded in region *urban_area_2*. Grounded means that the rule may only be applied if all of its basic behaviors (leaf nodes) take place in one of the selected areas. Thus, the production (1) must be grounded in all regions because the trip spans over all spatial regions from R.

Figure 3 shows how the spatial grounding in our simple example helps to disambiguate during parsing. For this given connection between behavior and space, the right parse tree of Figure 2 is impossible because rule (5) cannot be applied for regions outside of *urban_area_2*. Summarizing this section we can say that SGIS reduce the ambiguities that may occur when parsing a behavior sequence by modeling additional knowledge about the space where the behavior occurs.

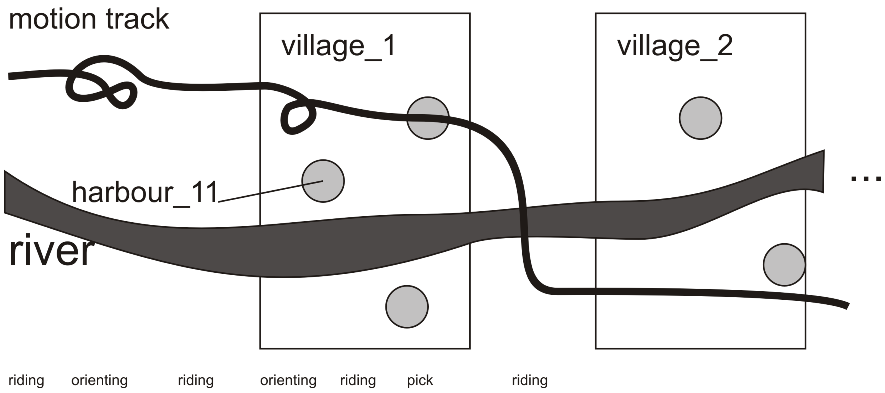


Fig. 4. Spatial behavior in the FluPa-Game.

3 Towards More Context Sensitivity

3.1 A Real World Example: The FluPa-Game

As in [6], we take a location-based game as use case for intention recognition. We may consider a location-based game "a game which is supported by localization technology and integrates the position of (one or several) players as main game element into its rules" [10]. In other words, players moving in a geographic gaming area are observed by localization technology, like GPS, which yields in a motion track as an input for intention analysis. The real advantage of choosing a game as use case are the limited possibilities of interaction. Especially when a game is played by bike, the user needs both hands for navigation. This is an ideal use case for implementing an information push mechanism.

A special case of location-based games are linear location-based games [11]. These are games where players are not free to move in the gaming area in an arbitrary manner, but follow a linear geographic feature, for instance a river or a railway line. This use case arises whenever the player's main intent consists in reaching a certain destination, like on a one-day bicycle tour from one city to another. The logics of such a game must ensure that players are not required to visit a location twice, for that would mean covering an extra distance. The first implemented, at least to our knowledge, bicycle game for linear trips was developed during the FluPa project at the Laboratory for Semantic Information Technologies, Otto-Friedrich-University Bamberg. The FluPa-Game has been implemented for Personal Digital Assistants (PDA). The GPS readings are taken from an external GPS receiver which transmits its data via Bluetooth. The implementation has just finished at the time of writing, so no user studies have been conducted yet.

We will use variations of the FluPa-Game in the rest of this section, so we need to sketch the main idea of the game concept: each of the two adversaries plays the part of a skipper on the river Regnitz (although, certainly, in reality

Production Rules for the FluPa-Game			
$PlayFluPa$	$\rightarrow$	$TreatVillage\ PlayFluPa$	(1)
	$\mid$	$SkipVillage\ PlayFluPa$	(2)
$SkipVillage$	$\rightarrow$	riding	(3)
$TreatVillage$	$\rightarrow$	riding $TreatVillage$	(4)
	$\mid$	riding $FindWay\ TreatVillage$	(5)
	$\mid$	$ChooseHarbour\ ReachHarbour\ ChangeGoods$	(6)
$FindWay$	$\rightarrow$	orienting	(7)
$ChooseHarbour$	$\rightarrow$	orienting	(8)
$ReachHarbour$	$\rightarrow$	riding	(9)
	$\mid$	riding $ReachHarbour$	(10)
	$\mid$	riding $FindWay\ ReachHarbour$	(11)
$ChangeGoods$	$\rightarrow$	picking	(12)
	$\mid$	dropping	(13)
	$\mid$	dropping picking	(14)

Fig. 5. Intentional system for the FluPa-Game.

they are moving by bike). It is basically a game around a transport optimization problem where you have to transport goods from one place to another to earn money. You have a certain capacity on your boat which you may not exceed with your freight. Before you may transport some goods, you must sign a delivery order of the form "4 sacks of grain from Eggolsheim to Forchheim for 16 gold pieces". The available delivery orders are optimized in a way that guarantees some competition between the players. Stations on the way are ordered linearly along the Regnitz, so you can only accept and execute delivery orders that lie ahead of you.

The types of spatial regions we can use for intention recognition are displayed in Figure 4: the regions of type *village* are ordered linearly along the river, and all part of the global region *game_area*. Each *village* contains at least two possible pick-up points, called *harbours*. All *harbours* inside one *village* are treated equal, in a sense that a player may reach any of them to trigger the same game action. An example intentional system for the intentions we want to support is given in Figure 5: a player decides before reaching a *village* if she wants to trigger some game actions there ($TreatVillage$) or just skip the village. Treating a village means: finding the village area, choosing one of the *harbours*, and going to the selected one (which again might include some wayfinding). Finally, the player must decide if to pick some goods, drop them, or do both actions. The motional behaviors are the result of a preprocessing of the GPS track which is segmented along the borders of the spatial regions (see [12] for details). The subscripts in Figure 4 give a hint of how a very simple preprocessing could look like. Picking and dropping events are taken directly from the user input (typing on the PDA).

Like in the simple example of section 2, we also have ambiguity for this grammar. For instance, the behavior sequence *riding, orienting, riding, orienting, riding, picking* is ambiguous, due to the two possible assignments of $FindWay$

Production Rules			
S	$\rightarrow$	$NP\ VP$	(1)
NP	$\rightarrow$	$Pronoun$	(2)
VP	$\rightarrow$	$Verb$	(3)
$Pronoun$	$\rightarrow$	I	(4)
	$\vert$	He	(5)
$Verb$	$\rightarrow$	live	(6)
	$\vert$	lives	(7)

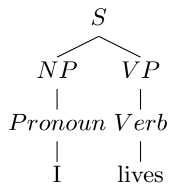


Fig. 6. A simple CFG from NLP (left) and one tree generated by this CFG (right).

(find way to the *village* or to the *harbour*). We will not present the corresponding parse trees here. It is also easy to find an appropriate spatial grounding for the rules that can help us in this example. Our point is that, through the rules of this specific game, we can formulate even more constraints on the applicability of certain production rules than the spatial grounding constraints.

3.2 A Parallel to NLP: Cross-Dependency Constraints

Most standard textbooks in NLP elaborate on syntax processing using CFGs like that shown in Figure 6 (left), which is certainly a very simplified one. Typical for NLP are the so-called preterminals which are nonterminals that can only have one terminal as child. These are typically used for word classes. For instance, $Pronoun$ and $Verb$ in Figure 6 are preterminals. Especially for preterminals, natural languages pose congruency constraints that are not covered by CFGs. The parse tree in Figure 6 (right), for instance, was deliberately chosen to generate a non-grammatical sentence. If the pronoun is chosen as “I”, the verb must certainly be “live”. This cannot be expressed in a CFG where the application of any production is by definition free of any other productions chosen for symbols in the tree that are not parent. In some natural languages, constraints like these can be even more complicated and show crossing dependencies such as in the pattern ABCABC.

In intention recognition, cross-serial dependencies like these can also occur. For intention recognition from spatio-temporal behaviors we can formulate three kinds of constraints, two of which may be cross-serial constraints:

1. **Rule-at-location constraints:** Certain productions may only be applied in certain regions.We described these constraints in section 2. These constraints are purely spatial constraints and covered by the SGIS formalism. They cannot be cross-serial because their influence is restricted to one subtree.
2. **Rule-rule constraints:** If a certain rule has been chosen, another rule must be chosen some time later in the parse tree. These constraints are equivalent to those in NLP and purely temporal constraints, because they do not pose any constraints on the regions where the rules are applied. These constraints can be cross-serial.

Production Rules			
$PlayFluPa$	$\rightarrow$	skip $PlayFluPa$	(1)
	$\mid$	pick $PlayFluPa$	(2)
	$\mid$	drop $PlayFluPa$	(3)

Fig. 7. Simplified intentional system for the FluPa-Game.

3. **Complex rule-location constraints**: If a certain rule has been applied in some region R, this restricts the freedom of applying another rule later in the parse tree. These constraints are spatial and temporal, and can also be cross-serial.

The examples we give for constraints 2 and 3 are drawn from the FluPa-Game. However, we do not need the full complexity of the rule set from Figure 5, because the most important aspect in this context is the choice between skipping a location, picking and dropping. Thus, for reasons of clarity, we use a simplified version of the FluPa grammar, as can be seen in Figure 7. Note that cross-dependencies appear just the same in the original full grammar from Figure 5 which we introduced for illustrative purposes.

A **rule-rule constraint** in the FluPa-Game is imposed by the fact that you cannot drop something before you have picked it somewhere else. In grammatical terms, that means that for any position in the input string, the number of 'drop' up to that point must not be greater than the number of 'pick'. It is obvious that this constraint can be handled with a non-deterministic pushdown automaton and thus be handled by a CFG. Thus, with a CFG we can avoid the application of rule (3) from Figure 7 when our stack (= the virtual boat in the game) is empty. However, intention recognition is not only about deciding if a certain input string belongs to our language, but rather about parsing the structure imposed by the rules. In other words, it does not suffice to decide that "pick pick drop pick drop drop" belongs to our language, but it will in many use cases also be important to associate the pairs of "pick" and "drop" that belong together. For instance, imagine "pick" and "drop" were only preterminals which in a second step had to be mapped to concrete goods, like "pick-grain" or "drop-fish". This long-ranging association between pairs cannot be expressed by a parse tree from a CFG.

A **complex rule-location constraint** in the FluPa-Game is given by the transport orders. For instance, if the player has picked the 4 sacks of grain in Eggolsheim (see the example transport order in section 3.1), the production that may be choosen in Forchheim is constrained to a "drop". One could also image typed constraints of the form: if you pick a sack of grain at a location of type *grain-pick-up-point*, you must choose the drop production at a location of type *grain-drop-point*.

3.3 Mildly Context Sensitive Grammars

Climbing up the well-known Chomsky hierarchy we come from regular grammars (type 3), over CFGs (type 2) to context-sensitive grammars (CSG, type 1). NLP has extended this hierarchy by adding new levels between CSGs and CFGs. The first formalism in between were *Indexed Grammars* developed by Aho [13]. In this formalism, a stack is attached to each node in the parse tree and handed on to all children when a production is executed, after a possible push or pop operation. A restricted version of Indexed Grammars are *Linear Indexed Grammars* (LIG). They were first used, but not yet designated as LIG, in [14]. An important restriction of LIGs is that the stack is handed on only to one child node. This allows efficient parsing in polynomial time.

Other grammatical formalisms that fall in between CSGs and CFGs are *Tree Adjoining Grammars* (TAG, [15]), Head Driven Phrase Structure Grammars (HDPSG, [16]) and Combinatory Categorial Grammars (CCG, [17]). Interestingly, these other formalisms have been shown to be weakly equivalent to LIGs, i.e. their expressiveness allows them to produce the same languages ([18]). Thus, LIG, TAG, HDPSG, and CCG can be subsumed under a common class of language formalisms, called *Mildly Context Sensitive Grammars* (MCSG)[1]. Joshi [19] was the first to call this class MCSG and defined common properties: limited cross-serial dependencies, constant growth, and polynomial parsing.

The latter, polynomial parsing, obviously is an important argument for using MCSGs in our use case. Cross-serial dependencies can be important in many use cases as explained above. The various MCSG formalisms vary in the kind of cross-serial dependencies they support, for instance regarding to the number of symbols in a string that may be part of the same depend-relation[2]. The constant growth property is explained in [7] as: "this means that if there is a plan of length n then there is another plan of length at most n+K where K is a constant for the specific domain". This restrains the growth of the parsing sequence and is one main reason why these formalisms stay polynomially parsable. As the authors in [7], we see no use case in intention recognition that speaks against this property. For instance, an exponential growth in the possible length of plans cannot be observed in any reasonable domain.

3.4 Example: Reformulating the FluPa-Game as a TAG

Tree-Adjoining-Grammars (also called Tree-Adjunct-Grammars) are a MCSG formalism which is - different than other grammars - rather a tree generative system than a string manipulating system. That means, the elementary units that are manupulated by operations are not strings of terminals and nonterminals, but trees. An introduction to TAG can be found in [15]. In comparison to other MCFGs, TAGs are a rather intuitive way of modeling grammatical knowledge. Additionally, a parser for TAGS is available under GNU General Public License (XTAG, [20]).

The trees in a TAG are divided into initial trees and auxiliary trees (see Figure 8). Initial trees are headed by the starting symbol and have exclusively

[1] Geib and Steedman [7] also call them *LIG-equivalents*.

[2] Although in this paper we only present examples of two dependent symbols, one can easily imagine cases where three or more symbols in an input string share one depend-relation.

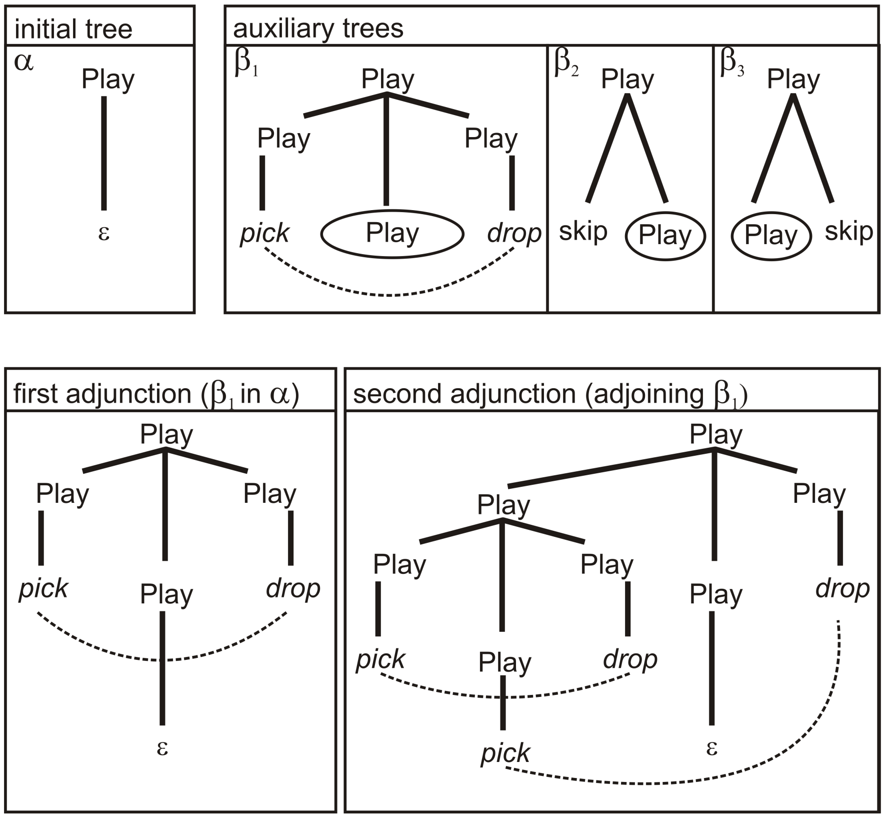


Fig. 8. A Tree-Adjoining-Grammar for the FluPa-Game (top) and two adjoining operations (bottom).

terminals in their leaf nodes. In other words: the initial tree itself already forms a correct word of the grammar through its terminals. Auxiliary trees have a head node, and a number of leaf nodes which are all but one terminal nodes. One of the leaf nodes is the so-called substitution node which has the same non-terminal as the head node (marked with a circle in Figure 8). On these trees, an operation called "adjunction" (or "adjoining") is defined. Through this operation, a non-terminal may be replaced by an auxiliary tree which has the same non-terminal as head and substitution node.

For instance, in the first step in Figure 8, the auxiliary tree β_1 is adjoined into the starting symbol *Play*. Thus, the original *Play* and ϵ from the initial tree are pushed down and appended to the substitution node. After the second adjunction, we get the final tree on the right bottom in Figure 8.

We have not mentioned the dotted lines yet: these are dependencies between nodes. TAGs allow these dependencies to be formulated on all initial and auxiliar trees. Through the adjoining operation, the dependency just gets stretched but not destroyed. Thus, in our example TAG we can generate any cross-dependency of pick and drop just like needed in the FluPa-game use case. On these dependency links, different types of constraints can be formulated [21]. For instance, if pick and drop were pre-terminals, we could formulate a *selective adjoining constraint* stating that the drop pre-terminal must be terminalized to the corresponding symbol to the pick pre-terminal (e.g. pick-grain, drop-grain). This constraint is later used by the parsing algorithm.

4 Related Work

We have already cited [7] who recently made an argument for using MCSGs in plan recognition. As in our work, they argued with possible cross-dependencies in plan recognition. However, it was left open how exactly the correspondence between cross-dependencies in NLP and plan recognition can look like for specific domains. Adding spatial knowledge was also not addressed in that paper.

The plan recognition community has spent some interest in probabilistic network based approaches (Bayesian approaches, [22]). One recent work in this area chooses a Hierachical Markov Model and Particle Filtering to predict a user's changes in transportation mode, like getting on or off a bus [23]. They evaluate their methods using a system called Opportunity Knocks which aims at helping cognitively-impaired people to use public transportation safely. Different to our approach, their mobile client sends data to an inference engine on a server so that weaker demands on the computational complexity of the algorithm hold. Another probabilistic network based approach, the Abstract Hidden Markov Model, is chosen by [24]. Again, they evaluate their algorithms not on a mobile client, but in an intelligent office environment.

A probabilistic model is also used in [25], here with the Dempster-Shafer theory of evidential reasoning. This approach is interesting for it formalizes spatial knowledge as spatial conceptual maps. The system collects evidence for a set of candidate locations from which the most probable is chosen. The user's future trajectory is predicted, not his or her intention.

Inferring the goals of a moving human agent is also the aim of [26]. However, the focus here is on detecting "inexplicable behavior" in an observed environment (like a car park), not on explaining the high level intentions. Possible goals are restrained to "reaching a certain point P". Sufficient for these simple goals, a representational formalism with low expressiveness (a state-transition diagram) is chosen.

An ex-post analysis of spatio-temporal trajectories that have occurred in a RoboCup game is the use case of [27]. This paper is related to our approach in that the motion patterns are topologically contextualized by cutting them at the edges of spatial regions. However, not the changing intentions of an agent at a certain time are analyzed, but the overall characterization of a game, for instance "BalancedWingPlaying(TeamA) = true".

In the extended FluPa-Game example from Figure 5 we assumed a stream of incoming preprocessed behaviors, namely orienting and riding. These behaviors can be even much more complex, as in the museum example from [28] (e.g. crossing, visiting, ant-like-touring, grasshopper-like-touring). To obtain such behaviors, qualitative characteristics need to be extracted from an observed motion path (or motion segment). Methods for this task have been developed in the area of spatial cognition, e.g. [29].

The types of behaviors that are of interest for a specific domain can generally be obtained in two different ways: one way is to model them by hand by asking a domain expert. On the other hand, we could detect them from a set of observed trajectories, e.g. when a number of FluPa-Games has produced a collection of recorded tracks. The task of automatic motion pattern detection is one concern of the spatio-temporal data mining community, see for instance [30].

On the lowest preprocessing level, before we can even get any reliable motion paths, we must deal with sensor noise. Players in the FluPa-Game, for example, are positioned with imprecise GPS measurements. Although newest GPS chip sets deliver quite high precision and accuracy, the position might still be noisy in situations with few satellites in view, for instance in the forest. To solve this problem, Bayesian Filtering techniques may help us [31]. For our use case of processing everything on-device, the benefits of Bayesian Filtering and its costs need to be evaluated deliberately.

5 Discussion and Future Work

In this paper, we have argued that CFGs are not sufficiently expressive for many use cases of intention recognition of spatio-temporal behavior. We identified three kinds of constraints on productions rules that may occur. We have shown that with TAGs we can express rule-rule constraints, while with SGIS we can formalize rule-at-location constraints.

What still remains open is the representation of complex rule-location constraints. It is not yet clear if all complex constraints that can occur are expressible with the same formalism, or if this category needs a further refinement. Furthermore, other grammatical formalisms than TAG need to be considered. An interesting issue we have skipped in this paper are probabilistic grammars. Even with spatial grounding of production rules, ambiguity in an input sequence may occur. For these cases, adding probabilities to our grammar might help.

Finally, we plan to evaluate the use of MCFGs with spatial information in user studies with the FluPa-Game. Other use cases are also planned, for instance a mobile tourist guide.

Acknowledgements

We wish to thank Klaus Stein for the discussions on the problem of intention recognition. Further, we would like to thank the FluPa-team (http://www.kinf.wiai.uni-bamberg.de/flupa/) for their help with the game concept and implementation.

The FluPa-project was funded by the Wasserwirtschaftsamt Kronach (watershed management office Kronach) and the Umweltstation Lias-Grube Unterstürmig e.V. (office for environmental education in Unterstürmig).

References

1. Baus, J., Krueger, A., Wahlster, W.: A resource-adaptive mobile navigation system. In: Proc. 7th International Conference on Intelligent User Interfaces, San Francisco, USA, ACM Press (2002) 15–22
2. Schmidt, C., Sridharan, N., Goodson, J.: The plan recognition problem: An intersection of psychology and artificial intelligence. Artificial Intelligence **11**(1-2) (1978) 45–83
3. Chanda, G., Dellaert, F.: Grammatical methods in computer vision: An overview. Technical Report GIT-GVU-04-29, College of Computing, Georgia Institute of Technology, Atlanta, GA, USA (2004) ftp://ftp.cc.gatech.edu/pub/gvu/tr/2004/04-29.pdf.
4. Ivanov, Y., Bobick, A.: Recognition of visual activities and interactions by stochastic parsing. IEEE transactions on pattern analysis and machine intelligence (PAMI) **22**(8) (2000) 852–872
5. Pynadath, D.V., Wellman, M.P.: Probabilistic state-dependent grammars for plan recognition. In: Proceedings of the 16th Annual Conference on Uncertainty in Artificial Intelligence. (2000) 507–514
6. Schlieder, C.: Representing the meaning of spatial behavior by spatially grounded intentional systems. In: GeoSpatial Semantics, First International Conference. Volume 3799 of Lecture Notes in Computer Science., Springer (2005) 30–44
7. Geib, C.W., Steedman, M.: On natural language processing and plan recognition. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI). (2007) 1612–1617
8. Earley, J.: An efficient context-free parsing algorithm. Communications of the ACM **13**(2) (1970) 94–102
9. Jurafsky, D., Martin, J.H., eds.: Lexicalized and Probabilistic Parsing. In: Speech and Language Processing. Prentice Hall, Upper saddle River, New Jersey, USA (2000) 447–476
10. Kiefer, P., Matyas, S., Schlieder, C.: Geogames - integrating edutainment content in location-based games. In Magerkurth, C., Roecker, C., eds.: Pervasive Games - Applications. (2007) to appear.
11. Kiefer, P., Matyas, S., Schlieder, C.: Playing on a line: Location-based games for linear trips. In: International Conference on Advances in Computer Entertainment Tecnology (ACE 2007). (2007) to appear.
12. Stein, K., Schlieder, C.: Recognition of intentional behavior in spatial partonomies. In: ECAI 2004 Worskhop 15: Spatial and Temporal Reasoning (16th European Conference on Artificial Intelligence). (2005)
13. Aho, A.V.: Indexed grammars - an extension of context-free grammars. Journal of the ACM **15**(4) (1968) 647–671
14. Gazdar, G.: Applicability of indexed grammars to natural languages. In: Natural Language Parsing and Linguistic Theories. D. Reidel, Dordrecht (1988)
15. Joshi, A.K., Schabes, Y.: Tree-adjoining grammars. In Rozenberg, G., Salomaa, A., eds.: Handbook of Formal Languages. Volume 3. Springer, Berlin, New York (1997) 69–124

16. Sag, I.A., Pollard, C.: Head-driven phrase structure grammar: An informal synopsis. CSLI Report 87-79, Stanford University, Stanford University (1987)
17. Steedman, M.: Dependency and coordination in the grammar of dutch and english. Language **61**(3) (1985) 523–568
18. Vijay-Shanker, K., Weir, D.: The equivalence of four extensions of context-free grammars. Mathematical Systems Theory **27**(6) (1994) 511–546
19. Joshi, A.K.: Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In Dowty, D.R., Karttunen, L., Zwicky, A.M., eds.: Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives. Cambridge University Press, Cambridge (1985) 206–250
20. XTAG Research Group: A lexicalized tree adjoining grammar for english. Technical Report IRCS-01-03, IRCS, University of Pennsylvania (2001)
21. Joshi, A.K.: Introduction to tree adjoining grammars. In Manaster-Ramer, A., ed.: Mathematics of Language, Amsterdam/Philadelphia, John Benjamins (1987) 87–114
22. Charniak, E., Goldman, R.P.: A bayesian model of plan recognition. Artificial Intelligence **64**(1) (1993) 53–79
23. Liao, L., Patterson, D.J., Fox, D., Kautz, H.: Learning and inferring transportation routines. Artificial Intelligence **171**(5-6) (2007) 311–331
24. Bui, H.H.: A general model for online probabilistic plan recognition. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). (2003)
25. Samaan, N., Karmouch, A.: A mobility prediction architecture based on contextual knowledge and spatial conceptual maps. IEEE Transactions on Mobile Computing **4**(6) (2005) 537–551
26. Dee, H., Hogg, D.: Detecting inexplicable behaviour. In: Proceedings of the British Machine Vision Conference, The British Machine Vision Association (2004) 477–486
27. Gottfried, B., Witte, J.: Representing spatial activities by spatially contextualised motion patterns. In Lakemeyer, G. et al., ed.: RoboCup 2007, International Symposium,, Springer (2007) 329–336
28. Schlieder, C., Werner, A.: Interpretation of intentional behavior in spatial partonomies. In: Spatial Cognition III, Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Learning. Volume 2685 of Lecture Notes in Computer Science., Springer (2003) 401–414
29. Musto, A., Stein, K., Eisenkolb, A., Röfer, T., Brauer, W., Schill, K.: From motion observation to qualitative motion representation. In: Spatial Cognition II, Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications, LNCS 1849, London, UK, Springer (2000) 115–126
30. Laube, P., van Krefeld, M., Imfeld, S.: Finding remo - detecting relative motion patterns in geospatial lifelines. In: Developments in Spatial Data Handling, Proceedings of the 11th International Symposium on Spatial Data Handling. (2004) 201–215
31. Fox, D., Hightower, J., Liao, L., Schulz, D., Borriello, G.: Bayesian filtering for location estimation. IEEE Pervasive Computing **2**(3) (2003) 24–33

Incremental Generation of Abductive Explanations for Tactical Behavior

Thomas Wagner, Tjorben Bogon, and Carsten Elfers

Center for Computing Technologies (TZI), Universität Bremen, D-28359 Bremen
{twagner, tbogon, celfers}@tzi.de

Abstract. According to the expert literature on (human) soccer, e.g., the tactical behavior of a soccer team should differ significantly with respect to the tactics and strategy of the opponent team. In the offensive phase the attacking team is usually able to *actively select* an appropriate tactic with limited regard to the opponent strategy. In contrast, in the defensive phase the more passive *recognition* of tactical patterns of the behavior of the opponent team is crucial for success. In this paper we present a qualitative, formal, abductive approach, based on a uniform representation of soccer tactics that allows to recognize/explain the tactical and strategical behavior of opponent teams based on past (usually incomplete) observations.

1 Introduction

Abductive inference, i.e., the inference to reason from *observation* to *cause*, has already been proposed by Charles S. Pierce [10] in the early 1920 as the third fundamental inference next to induction and deduction which accounts to the generation of new knowledge. Nevertheless, in contrast to the latter inferences abduction[1] has gained only limited attention. Recent advances in a wide range of scientific research fields ranging from robotics to ubiquitous and intelligent environments have imposed strong requirements in the generation of environmental context, especially in the face of (inter)acting agents. The environmental context is to a large extent characterized by the spatial representation of the objects and by the actions of the agents who (inter)act within the environment. While the generation of the spatial context has gained strong interest the recognition of action (from complex to simple) and intentions (underlying these actions) is especially in the area of ubiquitous and intelligent environments an open and demanding task. Nevertheless the recognition of intentions and actions of collaborating and competing agents is an essentiell task for (semi)active assistance. In order to provide appropriate active assistance the intelligent environment needs to how either *what* an agents intends to do or *how* an agents intents to realize his intentions. The requirements differ to the intention recognition process differ significantly from many other areas of application like diagnosis (see following section 2). In this paper we present a new approach to plan recognition based on the works of Eiter and Makino [4] that is specially suited to the demands of (active) assistance systems.

[1] The use of abductive inference is not necessarily limited to logical representations

The rest of the paper is organized as follows. In section 2 we describe in more detail the role of tactics in human soccer and try to motivate the relevance of incrementality and configurable context sensitivity. Furthermore we motivate how the recognition process is related to abduction. In section 3 we shortly describe the previous work on plan recognition with a focus on abductive methods. The succeeding section 4 shows on the one hand how a declarative representation can be transformed into horn clause (in a uniform fashion) and how qualitative action description can be generated based on 2D- and 3D simulation data. In section 4.4 we introduce our optimization and extensions of the *Eiter et al.*-approach. After the results in section 5, we finally conclude the paper with a discussion and an outlook on future work in section 6.

2 Motivation

One of the most important tasks to be accomplished in an intelligent environment is the generation of a consistent representation of the static (e.g., the spatial representation of the static environment) and dynamic actions and entities (e.g., most importantly the agents acting within). The generation of the *situational context*[2] is one of the most fundamental and demanding tasks, since it provides the basis for every kind of assistance that e.g., an intelligent environment will be able to provide. At least two assistance scenarios can be distinguished: (1) *passive* assistance, where the acting agent actively asks for support/information for a specific task and (2) *active* assistance, where the environment actively provides support without explicit user request. The latter task is of special interest, since it allows for a fundamentally new range of application: e.g., support of elderly and/or handicapped people or support of children (avoiding dangerous situation), ... Active support is not limited to these applications but is also interesting for everyday scenarios, like pre-heating the oven, defrost foot (both, if required). Active support is not only one of the most interesting scenarios for assistance but also one of the most demanding. It requires the intelligent environment to make to fundamental decisions: (a) *when* and (b) *how* support should be provided. Both questions can only be answered with respect to the specific environmental conditions and the intentions of the users. Especially the latter one imposes new requirements to the generation of the *sc* in terms of plan-/and intention recognition. In contrast to many other approaches to plan recognition like diagnosis we are not interested in a single hypotheses which tries to predict future behavior as precisely as possible (which appears to be least extremely difficult due to the indeterministic nature). Instead we require a plan/intention recognition process that accounts for,

Different levels of granularity: A precise (but therefore possibly incorrect) prediction is usually not necessary. Assume an intelligent environment wants to provide support by pre-heat the oven. In this case we do not want to know whether he is making (marinated beef or meat loaf). The level of granularity strongly depends on the assistance services available and may differ significantly with respect to the specific situations[3].

[2] *sc* for short

[3] Can be handled at the representational level.

Focused Recognition: Additionally, we are not interested in any prediction. We need a goal directed process that allows on the recognition of actions and intentions we are interested in, e.g., identifying certain risks (e.g. in child care scenarios) or identifying scenarios that allow for active support.

Explanatory coherence: The intelligent environment should be able to explain not only what it attempts to do but also *why*. This is especially important in supporting elderly or handicapped people. An explanation is additionally useful to find out whether the provided support is accepted in future situations, e.g., in order to improve user profile.

Time Efficiency: Efficiency may play a crucial role depending on the specific area of application. Especially in risk avoidance scenarios efficiency is an essential property.

Incrementality: Generating explanations should be designed as an incremental process.

Based on the RoboCup-domain which provides us an interesting scenario for a highly dynamic scenario with competitive as well as cooperating intentional agents we will show how plan recognition can applied incrementally and efficiently.

3 Related Work

The generation of explanations of observed events/behavior has gained much interest within the research community and resulted in various applications like diagnosis [9], natural language understanding [5] and plan recognition [2]. The methods under consideration vary from probabilistic methods like bayes networks [1], classification-based approaches [3] to the already mentioned logic-based abductive methods [2]. One aspect most of these approaches have in common are the constraints of their application: most approaches focus on static scenarios with an precise model of behavior (e.g., *closed world assumption*). An popular exception is the work of [1] which applys probabilistic reasoning to an online-dungeon game but also grounded on a complete (predefined and also limited) number of actions and valid combinations. An additional approach that overcomes these limitations at least to some extent is the work of Intille and Bobick [6] who apply their classification-based approach to the multi-agent scenario of *American football*-domain. Although the *American football*-domain appears to be highly related to the *RoboCup*-scenario the differences (within the Intille-approach) are significant. The *American football*-domain provides a complete, predefined taxonomic playbook which specifies all possible (allowed) patterns of behavior and therefore allows for classification-based approaches. Additionally, they are able to use manually generated data without noise. In contrast, in the soccer domain tactics and strategies describe behavior on much higher level that allows a wide range of variations that cannot be specified in all detail. Furthermore, the given observations have to be assumed to be uncomplete due to sensor limitation.

An alternative, logic-based approach that has been proposed for a wide range of applications is abduction (see above). Abduction has been introduced by C.S. Pierce [10] as a third kind of logic inference next to induction and deduction and has gained much interest in the late 80'th and early 90'th. Abduction does not rely on complete observations but instead supports to infer missing knowledge i.e., premisses. The abductive inference process can be generally be decomposed in two steps:

1. generation of all abductive explanation
2. selection of the most appropriate explanation

Several proposals have been made for the time consuming generation process depending on the underlying representation[4]. A serious problem for the use of abduction especially in time critical applications like RoboCup is that the generation of abductive explanations has been proofed to be NP-hard (in the general case - some time ago) [11]. Nevertheless, more recently Eiter et al. [4] developed an efficient algorithm that allows to generate all explanations of positive queries based on a logic horn-clause representation.

4 Abduction-based Generation of Explanations of Tactical Behavior

Observations are essential elements for plan recognition. In this approach the necessary information are recognized actions from a team of soccer players. The next section briefly describes the used process to gather these observations. Thereafter the usage of them for explaining tactical behavior is described.

4.1 Qualitative Action Recognition

In the following monotonicity based and threshold based qualitative propositions are distinguished (cf. [7]). Both *types of propositions* share several common properties, as the moment they have been satisfied for the first time, further called *StartTime* and the moment they stopped being satisfied, further called *FinishTime*. Additionally a copy of the world model is stored at *StartTime* and at *FinishTime*, further called *StartTimeWorldModel* and *FinishTimeWorldModel*. Thereby a proposition is defined as a tupel *Proposition[StartTime, FinishTime, StartTimeWorldModel, FinishTimeWorldModel, Type]*. *Monotonicity based Propositions* are satisfied if either a sequence of given values are monotonically increasing or monotonically decreasing, depending on their specified type through the proposition. E.g. the *ApproachingBall* proposition is satisfied as long as the distance from an agent to the ball is continues decreasing. *Threshold based Propositions* are satisfied as long as a given sequence of values under-run a specified threshold, e.g. if the distance from an agent to the ball does not overrun a threshold of 4 meters, the BallDribbleRange proposition is fulfilled. Figure 4.1 illustrates used propositions with their type and describtion.

The given propsitions are a qualitative description of the world which allow all considered actions to be defined as in expression 1 - 4. Each action inherits the common properties of the previously specified propositions, *StartTime*, *FinishTime* and so on.

[4] Abduction does not necessarily have to rely on an logic representation.

qual. proposition	type	satisfied when
ApproachingBall	Monotonically Decreasing	decreasing distance from agent to ball
DisapproachingBall	Monotonically Increasing	increasing distance from agent to ball
BallKickable	Threshold	ball is kickable for agent
BallDribbleRange	Threshold	ball is near agent
KeepingDirection	Threshold	minor direction changes of agent
Stopped	Threshold	agent not moving

Fig. 1. Used qualitative propositions for action recognition

According to the actions' name the GetBall condition (1) recognizes if an agent becomes the ball carrier:

$$ApproachingBall \wedge BallKickable \wedge \neg BallOwnedSoon_{LastValue} \tag{1}$$

The Dribble condition (2) recognizes if an agent dribbles the ball:

$$\begin{aligned} &BallDribbleRange \wedge (ApproachingBall_{StartTime} \\ &\geq BallDribbleRange_{StartTime}) \wedge (DisapproachingBall_{StartTime} \\ &\geq BallDribbleRange_{StartTime}) \end{aligned} \tag{2}$$

The pass condition (3) recognises if possibly a pass has been performed:

$$\neg BallDribbleRange \wedge DisapproachingBall \wedge BallOwnedSoon \tag{3}$$

The Move condition (4) recognises movements:

$$KeepingDirection \wedge \neg BallDribbleRange \wedge \neg Stopped \tag{4}$$

The common properties of the action conditions are assigned by checking the action conditions' satisfaction. *StartTime* and *StartTimeWorldModel* are assigned at the moment the action condition becomes satisfied, accordingly *FinishTime* and *FinishTime-WorldModel* are assigned at the moment the action condition stops beeing satisfied. At this time an action has been identified and temporal segmented. E.g. the *StartTime* of a move action is set to the first time, the regarded agent is keeping his direction, no ball is in dribble range and it is still moving (cf. 4). Analogical the *FinishTime* is set to the time the agent stops keeping its direction, a ball is getting near the agent (in dribble range) or the agent stops walking. For further application the identified actions are stored to a list. With the propositions' copys of the world model at *StartTime* and at *FinishTime*, it is ensured that all needed information is available to determine necessary parameters of the propositions.
In order to complete recognitions and delete unnecessary information, two *Modificators* have been implemented. *Modificators* are modules working on the previously created list of recognized actions. The first modificator is the *Pass Modificator*, searching in the list of actions for pass actions and get ball actions in a temporal relationship to complete the recognized pass action with information about the pass receiver, e.g. if it has been recognized that agent 4 performed a pass and agent 5 a get ball back-to-back, the pass action will be updated by the pass destination, in this case agent 5 and its position while

performing the get ball. Another modificator is the *Cut Modificator* that cuts off all gathered recognized actions when the ball is possessed by the team that is not regarded. This avoids the list of recognized actions expanding too much by deleting unnecessary information. After accomplishing this procedure all necessary information have been produced for following processing.

4.2 Generation of Tactical Knowledgebase

Our knowledge base is based on the book of soccer tactics from Lucchesi [8]. The key assumption which is essential for the use of abductive reasoning is that the (logical) implication can be interpreted not only as an inference from *cause to effect*, but also as an inference from *effect to cause*[5]. Following this pattern each single action within a complex tactical pattern can be interpreted as the cause for a possible sequence of successional actions and may itself be the effect of a previous action and therefore sequences of actions are modeled strictly as sequences of implications. The situation depicted in figure 2 is in the first step described as a sequence of implications (see figure 4) which is in the second step transfered into a *horn* representation[6]. Two limitations had to be considered (1) no cyclic horn theories and (2) no expressions like $x \rightarrow 0$ or $1 \rightarrow x$ are allowed[7]. The resulting horn-clause knowledge base is described in figure 4.

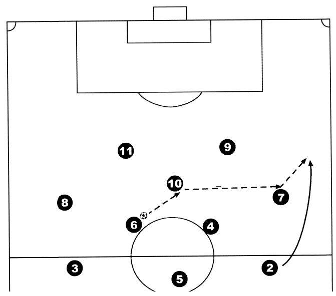


Fig. 2. A left-right Sidechange of a Human Soccer Tactic

(1) {¬PassPlayer6ToPlayer10, ¬MovePlayer2ToRightOppMid, RecievePassPlayer10}
(2) {¬RecievePassPlayer10, PassPlayer10ToPlayer7}
(3) {¬PassPlayer10ToPlayer7, RecievePassPlayer7}
(4) {¬RecievePassPlayer7, PassPlayer7ToPlayer2}
(5) {¬PassPlayer7ToPlayer2, RecievePassPlayer2}
(6) {¬HaveBallPlayer6, ¬IsFreePlayer10, PassPlayer6ToPlayer10}
(7) {¬IsFreeRightOppMid, MovePlayer2ToRightOppMid}
(8) {¬HaveBallPlayer10, ¬IsFreePlayer7, PassPlayer10ToPlayer7}
(9) {¬HaveBallPlayer7, ¬IsFreePlayer2, PassPlayer7ToPlayer2}

Fig. 3. Extracted Clauses from the Tactic.

[5] It should be mentioned that reasoning from *effect* to *cause* is *non-monotonic*.

[6] Due to space limitation we leave out the corresponding horn representation (quite straightforward)

[7] The first condition is a prerequisite of the algorithm of [4], the latter helps to ensures a consistent knowledge base

(1) PassPlayer6ToPlayer10 ∧ MovePlayer2ToRightOppMid → RecievePassPlayer10
(2) RecievePassPlayer10 → PassPlayer10ToPlayer7
(3) PassPlayer10ToPlayer7 → RecievePassPlayer7
(4) RecievePassPlayer7 → PassPlayer7ToPlayer2
(5) PassPlayer7ToPlayer2 → RecievePassPlayer2
(6) HaveBallPlayer6 ∧ IsFreePlayer10 → PassPlayer6ToPlayer10
(7) IsFreeRightOppMid → MovePlayer2ToRightOppMid
(8) HaveBallPlayer10 ∧ IsFreePlayer7 → PassPlayer10ToPlayer7
(9) HaveBallPlayer7 ∧ IsFreePlayer2 → PassPlayer7ToPlayer2

Fig. 4. Extracted Implication from the Tactic

4.3 The basic Algorithm

The basic algorithm can be decomposed into two main steps: (1) calculation of *prime implicants* and (2) calculation of abductive explanations[8] The first step results in a knowledge base of *prime implicants*.

In our example all clauses are also *prime implicants*. The following generations of abductive explanations will be done on this representation. The general idea is quite simple: in the first step the positive request clause σ is used to look up in the *consequences* of the set of all *prime implicants* (in the following *pi*). If σ is found in some prime clause ρ the corresponding *antecedents* of ρ as annotated as the first solution. Based on the first (simple) solution the algorithm tries systematically to find a more fundamental explanation by trying to find a new resolvent between ρ and some different clause in the set of *pi*. Given a new resolvent ρ is found *pi* is expanded by ρ and the same procedure is applied to *pi'* until no changes occur and therefore all solutions have been calculated.

4.4 Optimizing the Algorithm of Eiter and Makino

Although the described algorithm has not only been proved to be complete and correct it is also the only abductive algorithm known to be efficient (non NP-hard). In contrast to various different approaches to the generation of explanations an abductive algorithm is very robust with respect to redundant actions which is a serious problem in soccer in general[9]. Furthermore, an abductive algorithm can easily be applied to different knowledge bases at different levels of granularity. Nevertheless, efficiency is still a very serious problem for the application in a *RoboCup*-domain[10]. In order to improve efficiency various optimizations have been applied. It should be noted that the following

[8] Due to the space limitations we will only describe the more essential second part shortly. For more details please refer to [4].

[9] Even if a team is strictly using declarative tactical patterns of behavior, redundant actions result due to necessary adaptations as a result of unexpected opponent behavior

[10] On a more complex knowledge base the basic algorithm took 21 sec.! As we will see in section 5, with all optimization and with some restrictions the performance can be improved to 22ms.

Algorithm OPT-EXPLANATIONS;
Input: A Horn CNF φ , a positive letter *q* and Observation *obs*
Output: All nontrivial explanations of *q* from φ
Step 1. φ^* := $\emptyset$, *S* := $\emptyset$, and *O* := $\emptyset$;
OPT1.
if (*JustificationStruct* $\in \varphi$) $\wedge$ $(solutions\,(q) \in JustificationStruct)$ **then**
  S $= solutions\,(q)$;
  return;
end

Step 2.
OPT2.
if *usePreBuildPrimes = true;*
then φ^* = PreBuildPrimes;
else **foreach** c $\in \varphi$ **do**
  add any prime implicate *c'* $\subseteq$ *c* of φ to φ^*;
end
OPT3.
Mapping (q, φ^*) ;
foreach c' $\in \varphi^*$ *with* P(c') = $\{q\}$ *and* N(c') $\notin$ S **do**
  output *N(c')*;
  OPT4.
  S := S $\cup$ $\{$*N(c')*$\}\setminus$ *obs* ;
  O := O $\cup$ $\{$(c,c') $|$ c $\in \varphi^*$ $\}$;
end

Step 3. **while** *some* (c_1,c_2) $\in$ O *exists* **do**
  O:=*O* $\setminus$ $\{(c_1,c_2)\}$;
  if *(1)* q $\notin$N(c_1);
  (2) P(c_1) = $\{r\}$ $\subseteq$ N(c_2);
  OPT5.
  if *useSatisfyTest* = *true* **then**
    (3) $\varphi^* \cup$ N(c_1) $\cup$ N(c_2) $\setminus$ P(c_1) *is satisfiable*
  end
  then
    c := resolvent of c_1 and c_2;
    compute any prime implicate c' $\subseteq$ c of φ;
    if N*(c')* $\notin$ S **then**
      output *N*(c');
      OPT4.
      S:= S $\cup$ $\{$*N*(c')$\}\setminus$ *obs* ;
      O := O $\cup$ $\{$(c,c') $|$ c $\in \varphi^*\}$
    end
  end
end

Algorithm 1: (Optimized) Eiter et al. algorithm for generating all explanations optimizations may not be reasonable in all domains. I.e., although the optimized algorithm is (of course) domain-independent the optimizations can only be applied with some restrictions. Therefore they are optional with respect to different domains[11].

The complexity of the algorithm is given by $O(e * m * n * \|\varphi\|)$, whereas e denotes the number of solutions, m the number of clauses in φ and n the number of literals. The first significant improvement can be achieved by separating the complete knowledge base into different separat knowledge bases. As a matter of consequence, we get a special knowledge base for *counter attack on the right wing*, *counter attack on in the center*, ... This modularisation has an interesting advantage: as assumed in the motivation (see section2) a player is usually only interested in explanations that leads to an improved/adopted behavior: E.g., a right wing defender is specially interested in counter attack on the right wing and not on the left one. The modularisation allows him to focus on context sensitive, role specific tactic explanations.

Additionally, four different optimizations have been applied:

1. Pre-calculation of prime implicants: Independently of the specific request the basic algorithm calculates all prime implicants. This process is done before runtime in our realization and just has to be loaded together with the knowledge-base. The level of improvement is strongly dependent on the complexity of the causal relations and will lead at least in complex models to significant improvements.

2. Use of additional observations: In most of the cases a player has made different observations that account to a specific tactic. These observations can be used to improve the abductive reasoning process by skipping proofs. Since a player has already observed an action it is not necessary to find out under which conditions the observation is true.

3. Skipping satisfiability-test: The above handling of observations has an additional interesting side effect: In the classic abductive approach new observations are expanded in the knowledge base which in case of false observations *may* lead to an inconsistency. In order to avoid to an inconsistent knowledge base a satisfiability-test would be needed. But since observation are never expanded in the knowledge base the satisfiability-test can be skipped.

4. Pre-calculation of possible solutions: A static knowledge base offers additional advantages: it allows the pre-calculation of all possible solutions! A possible disadvantage can be the increase in memory usage which is minimal in this domain due to the modularisation of the knowledge base.

In addition to the improvements in efficiency the basic algorithm had to be adopted in order to increase robustness. Although tactical pattern described in Lucchesi [8] are strictly associated with specific tactical roles these assumption does rarely hold in the *RoboCup*-domain. Therefore we provided the abductive algorithm with a flexible role association method. The use of this method allows to detect tactical behavior independently from specific player numbers or roles and increases the robustness significantly. The obvious drawback is a decrease in efficiency since all player-number configurations have to be considered in the role assignment. The preliminary results of our extentions are described in the following section. The detailed modified algorithm is depicted in algorithm 1.

[11] And they are also optional in our implementation.

5 Experiments and Preliminary Results

Before we tried to integrate the modified algorithm in our 3D-team we evaluated the efficiency in different scenarios. In the first test scenario described in table 1 we wanted to evaluate the effect of pre-calculated solutions under (1) the varying condition of request complexity: simple vs. complex and (2) under varying goal: whether the agent wants to know if a plan is possible at all or whether he wants to get a list of all possible (opponent-) plans. The request complexity is simply changed by the action selection. In the case we observe an action that appends very late in a possible plan, the algorithm will find significant more solutions than in the inverse case. In the following tests we used 26 different plans in each case. The table 1 presents some interesting results.

Test 1	with Pre-Calculation		without Pre-Calculation	
	simp. Query	komp. Query	simpl. Query	komp. Query
all Time	62,8 ms	55,7 ms	69,3 ms	74,2 ms
∅ Time for one Plan	4,7 ms	4,5 ms	6,3 ms	10,5 ms
processed Plans	4 P.	3 P.	4 P.	3 P.
∅ compute Possible Plans	44,2 ms	42 ms	43,8 ms	42,2 ms

Table 1. Finds all Solution and search all Plans with a possible Solutions.

First, the calculation whether a single plan is possible is highly efficient and can be done in 4,7 ms to 10,5 ms with respect to the specific conditions. Interestingly, the pre-calculation of results is significantly more efficient than without but the difference is surprisingly small. Two main reasons can be found: (1) the complexity of our tactical model is quite low (in terms the capability of the algorithm). The efficiency decreases in the case of no pre-calculations only for complex requests. (2) The modularisation of the knowledge base appears to be highly efficient. In the case that all possible plans should be calculated (which represents the case of non-modularisation) the run-time requirements are significant higher. - But still efficient enough to be used e.g., in the 3D-simulation league, as it can be seen in table 2. In table 2 we used a 3D-trainer agent

	iterativ Test									
	all Queries					Queries with Solutions				
	min	max	∅	∅ over	Tests	min	max	∅	∅ over	Tests
P28	2	7	4,66	0,6	22,6	2	7	4,66	0,6	22,6
P53	0	4	1,06	6	182,2	1	4	2,62	6	55,8
P99	0	7	2,24	1,2	49,2	1	7	3,96	1	24,8
P130	0	6	2,6	0,6	53,4	2	6	4,6	0,2	16,6

Table 2. Used Cycles in 2D-League, watch a game with a Traineragent in 3d-League

who has been restricted to use at maximum 80ms in order to simulate cycles. In the test the coach was required to detect whether a single plan is possible and to find all possible solutions. These conditions have been tested on four different varying plans (P28, P53, P99, P130). The test mainly showed two important results: (1) The modularisation of the knowledge base provides a basis for incremental abductive reasoning. The algorithm in our implementation can interrupt the calculation process at (relative) fixed time steps in order to allow other tasks within a single cycle (instead using complete cycles). (2) Depending on the specific condition the algorithm requires at most between 4 to 7 cycles for all solutions. These results can also be approved under different conditions e.g., used by a 3D- or a 2D-player. The detailed description of all generated results is out of the scope of this paper (space limitations). - If the reviewers consider one of the other mentioned results more interesting then we will integrate them.

6 Summary and Discussion

The role of strategy and tactic is becoming more and more important especially in the simulation- and the small-size league. The use of more complex tactics of an offensive team will require that defensive teams are at least to some extent able to detect the set of possible opponent tactical patterns in order to coordinate defensive behavior. Nevertheless, the use strategical and tactical knowledge is not limited to a specific domain but is instead relevant in any domain that is related to the cooperative and/competitive dynamic interaction between natural and artificial agents.

In this paper we presented an approach to symbolic plan recognition (more precisely generation of explanation for opponent behavior) at all relevant stages: from the generation of qualitative action- and world descriptions based on [7] to the generation of abductive explanations of these observed behavior. The algorithm of Eiter and Makino [4] has been adopted to the specific requirements of highly dynamic domains. We showed that the adopted algorithm can efficiently be used for explanation generation. Furthermore, the algorithm can be used in an incremental fashion which is especially useful for monitoring scenarios. An additional characteristic is the robustness with respect to redundant/false observations/actions which have to be handled in any physically grounded scenario. The modularisation of the knowledge base allows for role- and context sensitive requests but does not prohibit the generation of complete solutions, i.e., without respect to role and context e.g., for the trainer.

Besides the application of the modified algorithm in different domains the hypotheses generation is still an open task. Although we claim that it will be sufficient in many situations to identity possible tactical behavior (with respect to role and context) it is clear that there also exits situations where we would like a single prediction, i.e., the selection of a single hypothesis out of the set of possible explanations. Various solutions may be considered, varying from probabilistic to symbolic approaches proposed in the abduction community [9].

References

1. David W. Albrecht, Ingrid Zukerman, and An E. Nicholson. Bayesian models for keyhole plan recognition in an adventure game. *User Modeling and User-Adapted Interaction*, 8(1-2):5–47, 1998.

2. Douglas E. Appelt and Martha E. Pollack. Weighted abduction for plan ascription. *User Modeling and User-Adapted Interaction*, 2(1-2):1–25, 1991.
3. Dorit Avrahami-Zilberbrand and Gal A. Kaminka. Fast and complete symbolic plan recognition. In *IJCAI*, pages 653–658, 2005.
4. Thomas Eiter and Kazuhisa Makino. On computing all abductive explanations. In *Eighteenth national conference on Artificial intelligence*, pages 62–67, Menlo Park, CA, USA, 2002. American Association for Artificial Intelligence.
5. Jerry R. Hobbs, Mark Stickel, Paul Martin, and Douglas D. Edwards. Interpretation as abduction. In *26th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, pages 95–103, Buffalo, New York, 1988.
6. S. Intille and A. Bobick. Recognizing planned, multi-person action. In *Computer Vision and Image Understanding*, volume 81, pages 414–445, 2001.
7. Andreas D. Lattner, Andrea Miene, Ubbo Visser, and Otthein Herzog. Sequential pattern mining for situation and behavior prediction in simulated robotic soccer. In *RoboCup*, pages 118–129, 2005.
8. Massimo Lucchesi. *Choaching the 3-4-1-2 and 4-2-3-1*. Reedswain Publishing, edizioni nuova prhomos edition, 2001.
9. Hwee Tou Ng and Raymond J. Mooney. On the role of coherence in abductive explanation. In *National Conference on Artificial Intelligence*, pages 337–342, 1990.
10. Charles Sanders Peirce. *Collected Papers of Charles Sanders Peirce*. Harvard University Press, 1931.
11. Bart Selman and Hector J. Levesque. Support set selection for abductive and default reasoning. *Artif. Intell.*, 82(1-2):259–272, 1996.

From Distributed Vision Networks to Human Behavior Interpretation

Hamid Aghajan and Chen Wu

Department of Electrical Engineering
Stanford University, Stanford CA, 94305, USA

Abstract. Analysing human behavior is a key step in smart home applications. Many reasoning approaches utilize information of location and posture of the occupant in qualitative assessment of the user's status and events. In this paper, we propose a vision-based framework to provide quantitative information of the user's posture which can be used to deduct qualitative representations for high-level reasoning. Furthermore, our approach is motivated by potentials introduced by interactions between the vision module and the high-level reasoning module. While quantitative knowledge from the vision network can either complement or provide specific qualitative distinctions for AI-based problems, these qualitative representations can offer clues to direct the vision network to adjust its processing operation according to the interpretation state. The paper outlines potentials for such interactions and describes two vision-based fusion mechanisms. The first employs an opportunistic approach to recover the full-parameterized human model by the vision network, while the second employs directed deductions from vision to address a particular smart home application in fall detection.

1 Introduction

The increasing interest in understanding human behaviors and events in a camera context has heightened the need for gesture analysis of image sequences. Gesture recognition problems have been extensively studied in Human Computer Interactions (HCI), where often a set of pre-defined gestures are used for delivering instructions to machines [1, 2]. However, "passive gestures" predominate in behavior descriptions in many applications. Some traditional application examples include surveillance and security applications, while more novel applications arise in emergency detection in clinical environments [3], video conferencing [4, 5], and multimedia and gaming applications. Some approaches to analyzing passive gestures have been investigated in [6, 7].

In a multi-camera network, access to multiple sources of visual data often allows for making more comprehensive interpretations of events and gestures. It also creates a pervasive sensing environment for applications where it is impractical for the users to wear sensors. Having access to interpretations of posture and gesture elements obtained from visual data over time enables higher-level

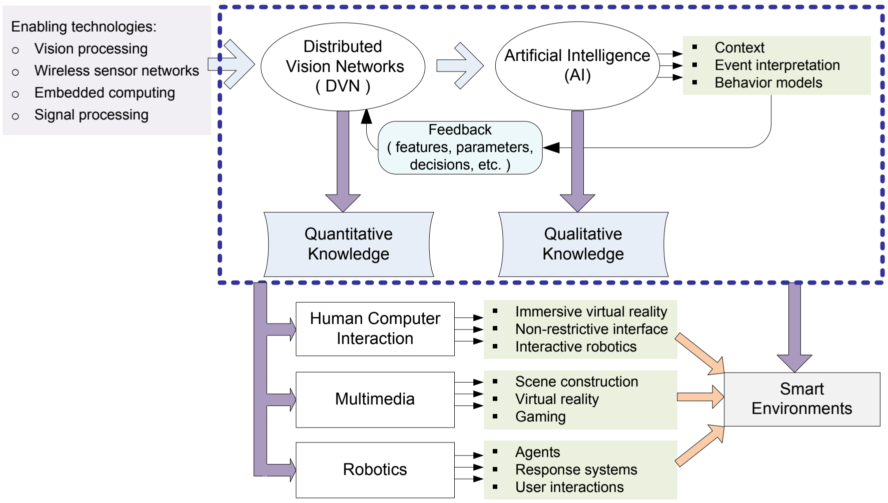


Fig. 1. The relationship between vision networks and high-level AI reasoning, and a variety of novel applications enabled by both.

reasoning modules to deduct the user's actions, context, and behavior models, and decide upon suitable actions or responses to the situation.

Our notion of the role a vision network can play in enabling novel intelligent applications derives from the potential interactions between the various disciplines outlined in Fig. 1. The vision network offers access to quantitative knowledge about the events of interest such as the location and other attributes of a human subject. Such quantitative knowledge can either complement or provide specific qualitative distinctions for AI-based problems. On the other hand, we may not intend to extract all the detailed quantitative knowledge available in visual data since often a coarse qualitative representation may be sufficient in addressing the application [8]. In turn, qualitative representations can offer clues to the features of interest to be derived from the visual data allowing the vision network to adjust its processing operation according to the interpretation state. Hence, the interaction between the vision processing module and the reasoning module can in principle enable both sides to function more effectively. For example, in a human gesture analysis application, the observed elements of gesture extracted by the vision module can assist the AI-based reasoning module in its interpretative tasks, while the deductions made by the high-level reasoning system can provide feedback to the vision system from the available context or behavior model knowledge.

In this paper we introduce a model-based data fusion framework for human posture analysis using opportunistic use of manifold sources of vision-based information obtained from the camera network in a principled way. The framework spans the three dimensions of time (each camera collecting data over time), space (different camera views), and feature levels (selecting and fusing different feature

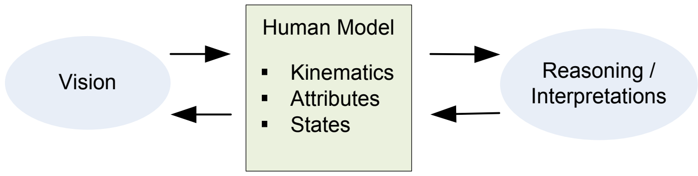


Fig. 2. The human model bridges the vision module and the reasoning module, as the interactive embodiment.

subsets). Furthermore, the paper outlines potentials for interaction between the distributed vision network and the high-level reasoning system.

The structure of the vision-based processing operation has been designed in such a way that the lower-level functions as well as other in-node processing operations will utilize feedback from higher levels of processing. While feedback mechanisms have been studied in active vision areas, our approach aims to incorporate interactions between the vision and the AI operations as the source of active vision feedback. To facilitate such interactions, we introduce a human model as the convergence point and a bridge for the two sides, enabling both to incorporate the results of their deductions into a single merging entity. For the vision network, the human model acts as the embodiment of the fused visual data contributed by the multiple cameras over observation periods. For the AI-based functions, the human model acts as a carrier of all the sensed data from which gesture interpretations can be deducted over time through rule-based methods or mapping to training data sets of interesting gestures. Fig. 2 illustrates this concept in a concise way.

In Section 2 we outline the different interactions between the vision and AI modules as well as the temporal and spatial model-based feedback mechanisms employed in our vision analysis approach. Section 3 presents details and examples for our model-based and opportunistic feature fusion mecahnisms in human posture analysis. In Section 4 an example collaborative vision-based scheme for deriving qualitative assessment for fall detection is described. Section 5 offers some concluding remarks and the topics of current investigation.

2 The Framework

Fig. 3 shows the relationship between the low-level vision processing, which occurs in the camera nodes, the instantaneous state resulting from camera collaboration in the visual domain, and the high-level behavior interpretation which is performed in the AI module. The feedback elements provided by the AI module help the vision processing system to direct its processing effort towards handling the more interesting features and attributes.

The concept of feedback flow from higher-level processing units to the lower-level modules also applies when considering the vision network itself. Within each camera, temporal accumulation of features over a period of time can for

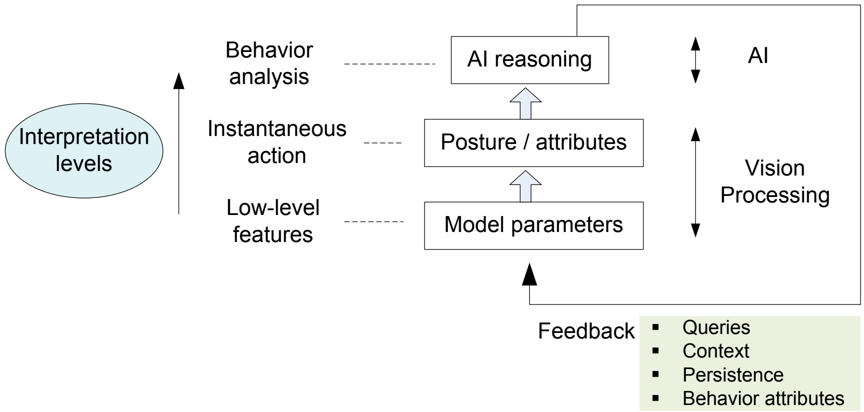


Fig. 3. Interpretation focuses on different levels from vision to behavior reasoning.

example enable the camera to examine the persistence of those features, or to avoid re-initialization of local parameters. In the network of cameras, spatial fusion of data in any of the forms of merged estimates or a collective decision, or in our model-based approach in the form of updates from body part tracking, can provide feedback information to each camera. The feedback can for example be in the form of indicating the features of interest that need to be tracked by the camera, or as initialization parameters for the local segmentation functions. Fig. 4 illustrates the different feedback paths within the vision processing unit.

3 Collaborative Vision Network

We introduce a generic opportunistic fusion approach in multi-camera networks in order to both employ the rich visual information provided by cameras and incorporate learned knowledge of the subject into active vision analysis. The opportunistic fusion is composed of three dimensions, space, time and feature levels. For human gesture analysis in a multi-camera network, spatial collaboration between multi-view cameras naturally facilitates solving occlusions. It is especially advantageous for gesture analysis since human body is self-occlusive. Moreover, temporal and feature fusion help to gain subject-specific knowledge, such as the current gesture and subject appearance. This knowledge is in turn used for a more actively directed vision analysis.

3.1 The 3D Human Body Model

Fitting human models to images or videos has been an interesting topic for which a variety of methods have been developed. Usually assuming a dynamic model (such as walking)[9, 10] will greatly help us to predict and validate the posture estimates. But tracking can easily fail in case of sudden motions or other

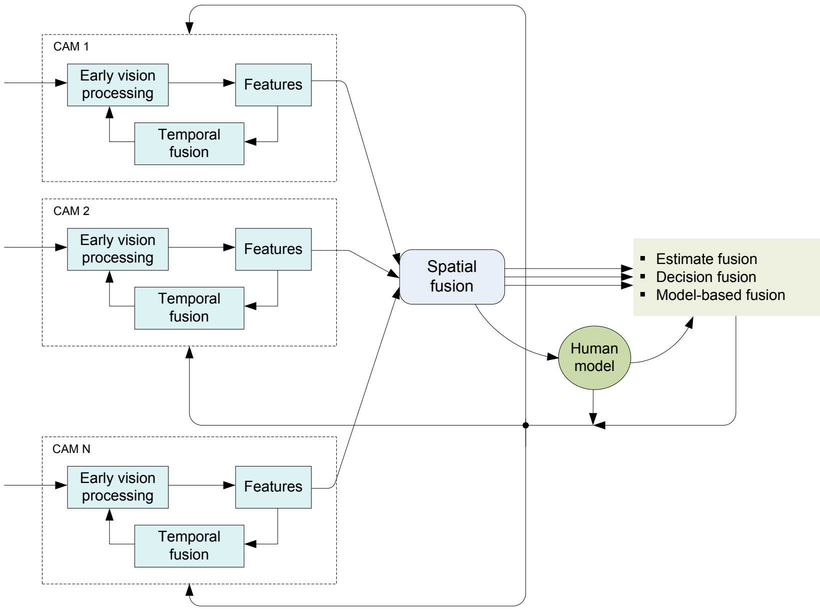


Fig. 4. Different feedback paths within distributed vision processing units.

movements that differ much from the dynamic model. Therefore we always need to be aware of the balance between the limited dynamics and the capability to discover more diversified postures. For multi-view scenarios, a 3D model can be reconstructed by combining observation from different views [11, 12]. Most methods start from silhouettes in different cameras, then points occupied by the subject can be estimated, and finally a 3D model with principle body parts is fit in the 3D space [13]. The approach above is relatively "clean" since the only image component it is based on are the silhouettes. But at the same time the 3D voxel reconstruction is sensitive to the quality of the silhouettes and accuracy of camera calibrations. It is not difficult to find situations where background subtraction for silhouettes suffers for quality or is almost impossible (clustered, complex background, and the subject is wearing clothes with similar colors to the background). Another aspect of the human model fitting problem is the choice of image features. All human model fitting methods are based on some image features as targets to fit the model. Most of them are based on generic features such as silhouettes or edges [14, 12]. Some use skin colors but those methods are prone to failure in some situations since lighting usually has big influence in colors and skin color varies from person to person.

In our work, we aim to incorporate appearance attributes adaptively learned from the network for initialization of segmentation, because usually color or texture regions are easier to find than generic features such as edges. Another emphasis of our work is that images from a single camera are first reduced to short descriptions and then reconstruction of the 3D human model is based on

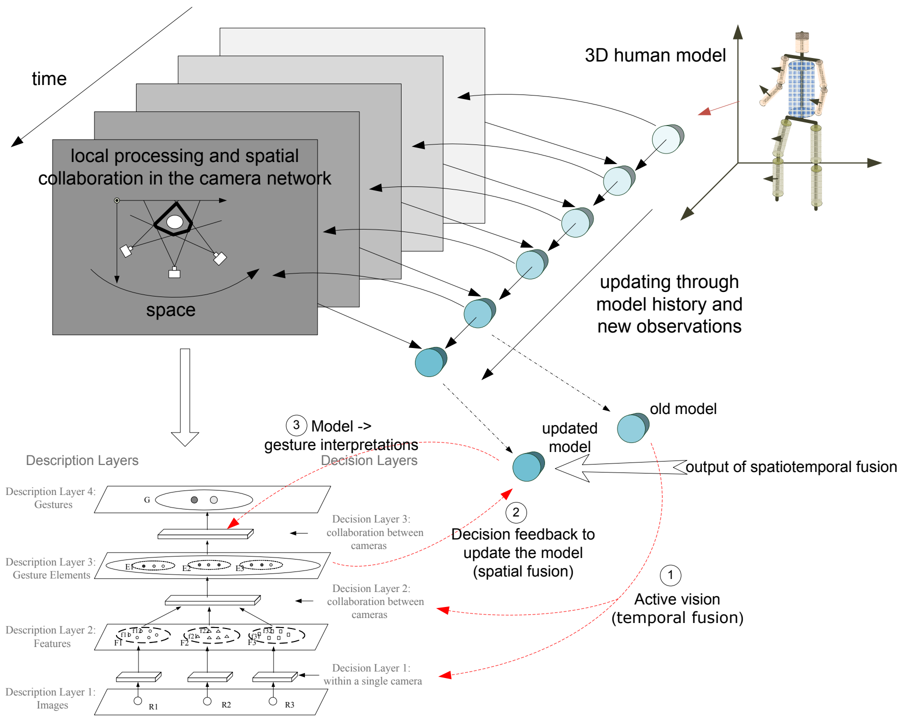


Fig. 5. Spatiotemporal fusion for human gesture analysis.

descriptions collected from multiple cameras. Therefore concise descriptions are the expected outputs from image segmentation.

In our approach a 3D human body model embodies up-to-date information from both current and historical observations of all cameras in a concise way. It has the following components: 1. Geometric configuration: body part lengths, angles. 2. Color or texture of body parts. 3. Motion of body parts. The three components are all updated from the three dimensions of space, time and features of the opportunistic fusion.

Apart from providing flexibility in gesture interpretations, the 3D human model also plays significant roles in the vision analysis process. First, the total size of parameters to reconstruct the model is very small compared to the raw images, and affordable through communication. For each camera, only segment descriptions are needed for collaboratively reconstructing the 3D model. Second, the model is a converging point of spatiotemporal and feature fusion. All the parameters it maintains are updated from the three dimensions of space, time and features of the opportunistic fusion. In sufficient confidence levels, parameters of the 3D human body model are again used as feedback to aid subsequent vision analysis. Third, although predefined appearance attributes are generally not reliable, adaptively learned appearance attributes can be used to identify the person

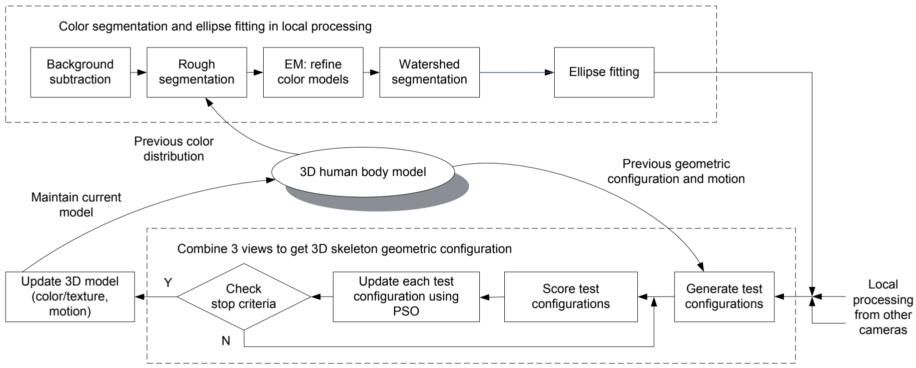


Fig. 6. Algorithm flowchart for 3D human skeleton model reconstruction.

or body parts. Those attributes are usually more distinguishable than generic features such as edges once correctly discovered.

The 3D model maps to the Gesture Elements layer in the layered architecture for gesture analysis (lower left part of Fig. 5) we proposed in [15]. However, here it not only assumes spatial collaboration between cameras, but also connects decisions from history observations with current observations.

3.2 The Opportunistic Fusion Mechanisms

The opportunistic fusion framework for gesture analysis is shown in Fig. 5. On the top of Fig. 5 are spatial fusion modules. In parallel is the progression of the 3D human body model. Suppose now it is t_0, and we have the model with the collection of parameters as M_0. At the next instance t_1, the current model M_0 is input to the spatial fusion module for t_1, and the output decisions are used to update M_0 from which we get the new 3D model M_1.

Now we look into a specific spatial fusion module (the lower part of Fig. 5) for the detailed process. In the bottom layer of the layered gesture analysis, image features are extracted from local processing. Distinct features (e.g. colors) specific for the subject are registered in the current model M_0 and are used for analysis, which may be much easier than always looking for patterns of the generic features (arrow ① in Fig. 5). After local processing, data is shared between cameras to derive for a new estimate of the model. Parameters in M_0 specify a smaller space of possible M_1's. Then decisions from spatial fusion of cameras are used to update M_0 to get the new model M_1 (arrow ② in Fig. 5). Therefore for every update of the model M, it combines space (spatial collaboration between cameras), time (the previous model M_0) and feature levels (choice of image features in local processing from both new observations and subject-specific attributes in M_0). Finally the new model M_1 is used for high-level gesture deductions in a certain scenario (arrow ② in Fig. 5).

An implementation for the 3D human body posture estimation is illustrated in Fig. 6. Local processing in single cameras include segmentation and ellipse fitting for a concise parametrization of segments. For spatial collaboration, ellipses from all cameras are merged to find the geometric configuration of the 3D skeleton model.

3.3 In-Node Feature Extraction

The goal of local processing in a single camera is to reduce raw images/videos to simple descriptions so that they can be efficiently transmitted between cameras. The output of the algorithm will be ellipses fitted from segments and the mean color of the segments. As shown in the upper part of Fig. 6, local processing includes image segmentation for the subject and ellipse fitting to the extracted segments.

We assume the subject is characterized by a distinct color distribution. Foreground area is obtained through background subtraction. Pixels with high or low illumination are also removed since for those pixels chrominance may not be reliable. Then a rough segmentation for the foreground is done either based on K-means on chrominance of the foreground pixels or color distributions from the known model. In the initialization stage when the model hasn't been well established, or when we don't have a high confidence in the model, we need to start from the image itself and use a method such as K-means to find color distribution of the subject. However, when a model with a reliable color distribution is available, we can directly assign pixels to different segments based on the existing color distribution. The color distribution maintained by the model may not be accurate for all cameras, since in different cameras illumination may change. Also the subject's appearance may change due to the movement or lighting conditions. Therefore the color distribution of the model is only used for a rough segmentation in initialization of the segmentation scheme. Then an EM (expectation maximization) algorithm is used to refine the color distribution for the current image. The initial estimated color distribution plays an important role because it can prevent EM from being trapped in local minima.

Suppose the color distribution is a mixture of N Gaussian modes, with parameters $\Theta = \{\theta_1, \theta_2, \ldots, \theta_3\}$, where $\theta_l = \{\mu_l, \Sigma_l\}$ are the mean and covariance matrix of the modes. Mixing weights of different modes are $A = \{\alpha_1, \alpha_2, \ldots, \alpha_3\}$. The EM algorithms aims to find the probability of each pixel x_i belonging to a certain mode θ_l: $Pr(y_i = l|x_i)$.

However, the basic EM algorithm takes each pixel independently, without considering the fact that pixels belonging to the same mode are usually spatially close to each other. In [16] Perceptually Organized EM (POEM) is introduced. In POEM, influence of neighbors is incorporated by a weighting measure $w(x_i, x_j) = e^{-\frac{\|x_i - x_j\|}{\sigma_1^2} - \frac{\|s(x_i) - s(x_j)\|}{\sigma_2^2}}$. $s(x_i)$ is the spatial coordinate of x_i. Then "votes" for x_i from the neighborhood is given by

$$V_l(x_i) = \sum_{x_j} \alpha_l(x_j) w(x_i, x_j), \text{where } \alpha_l(x_j) = Pr(y_j = l|x_j) \tag{1}$$

Then modifications are made to EM steps. In the E step, $\alpha_l^{(k)}$ is changed to $\alpha_l^{(k)}(x_i)$, which means that for every pixel x_i, mixing weights for different modes are different. This is partially due to the influence of neighbors. In the M step, mixing weights are updated by

$$\alpha_l^{(k)}(x_i) = \frac{e^{\eta V_l^{(x_i)}}}{\sum_{k=1}^{N} e^{\eta V_k^{(x_i)}}} \tag{2}$$

η controls the "softness" of neighbors' votes. If η is as small as 0, then mixing weights are always uniform. If η approaches infinity, the mixing weight for the mode with the largest vote will be 1.

After refinement of the color distribution with POEM, we set pixels with high probability (e.g., bigger than 99.9%) that belong to a certain mode as markers for that mode. Then watershed segmentation algorithm is implemented to assign labels for undecided pixels. Finally for every segment an ellipse is fitted to it in order to obtain a concise parameterization for the segment.

3.4 Posture Estimation

Human posture estimation is essentially an optimization problem, in which we try to minimize the distance between the posture and ellipses from multi-view cameras. There can be several different ways to find the 3D skeleton model based on observations from multi-view images. One method is to directly solve for the unknown parameters through geometric calculation. In this method we need to first establish correspondence between points/segments in different cameras, which is itself a hard problem. Common observations for points are rare for human problems, and body parts may take on very different appearance from different views. Therefore it is difficult to resolve ambiguity in 3D space based on 2D observations. A second method would be to cast a standard optimization problem, in which we find optimal θ_i's and ϕ_i's to minimize an objective function (e.g., difference between projections due to a certain 3D model and the actual segments) based on properties of the objective function. However, if the problem is highly nonlinear or non-convex, it'll be very difficult or time consuming to solve. Therefore searching strategies which do not explicitly depend on the objective function formulation are desired.

Motivated by [17], Particle Swarm Optimization (PSO) is used as the optimization technique. The lower part of Fig. 6 shows the estimation process. Ellipses from local processing of single cameras are merged together to reconstruct the skeleton. Here we consider a simplified problem in which only arms change in position while other body parts are kept in the default location. Elevation angles (θ_i) and azimuth angles (ϕ_i) of the left/right upper/lower parts of the arms are specified as parameters. The assumption is that projection matrices from 3D skeleton to 2D image planes are known. This can be achieved either from locations of cameras and the subject, or it can be calculated from some known projective correspondences between the 3D subject and points in the images, without knowing exact locations of cameras or the subject.

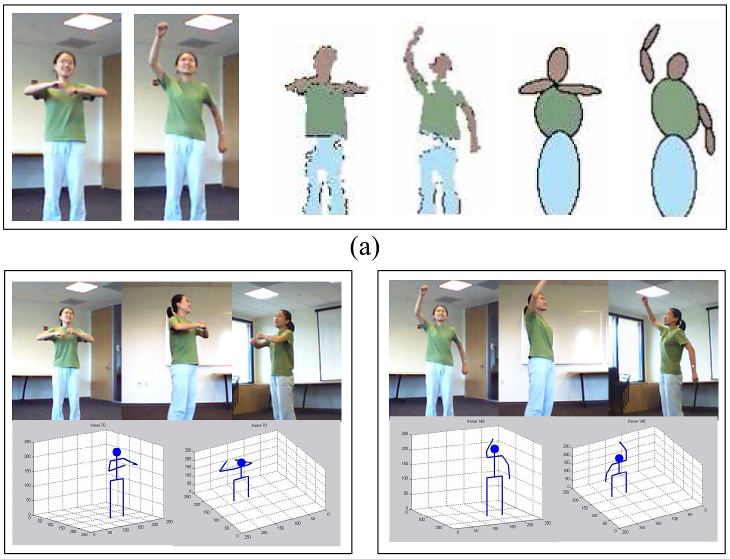


Fig. 7. Examples for gesture analysis in the vision network. (a) In-node segmentation results. (b) Skeleton model reconstruction by collaborative fusion.

PSO is suitable for posture estimation as an evolutionary optimization mechanism. It starts from a group of initial particles. During the evolution of the particles towards an optimal, they are directed to the good position while keep some randomness to explore the search space. Suppose there are N particles (test configurations) x_i, each is a vector of θ_i's and ϕ_i's. v_i is the velocity of x_i. The best position of x_i so far is $\hat{x}_i$, and the global best position of all x_i's so far is g. $f(\cdot)$ is the objective function that we wish to find the optimal position x to minimize $f(x)$. The PSO algorithm is as follows:

1. Initialize x_i and v_i. v_i is usually set to 0, and $\hat{x}_i = x_i$. Evaluate $f(x_i)$ and set $g = \text{argmin} f(x_i)$.
2. While the stop criterion is not satisfied, do for every x_i
   - $v_i \leftarrow \omega v_i + c_1 r_1(\hat{x}_i - x_i) + c_2 r_2 (g - x_i)$;
   - $x_i \leftarrow x_i + v_i$;
   - If $f(x_i) < f(\hat{x}_i)$, $\hat{x}_i = x_i$; If $f(x_i) < f(g)$, $g = x_i$.

The stop criterion: after updating all N x_i's once, the increase in $f(g)$ falls below a threshold, then the algorithm exits. ω is the "inertial" coefficient, while c_1 and c_2 are the "social" coefficients. r_1 and r_2 are random vectors with each element uniformly distributed on [0,1]. Choice of ω, c_1 and c_2 controls the convergence process of the evolution. If ω is big, the particles have more inertia and tend to keep their own directions to explore the search space. This allows for more chance of finding the "true" global optimal if the group of particles are currently

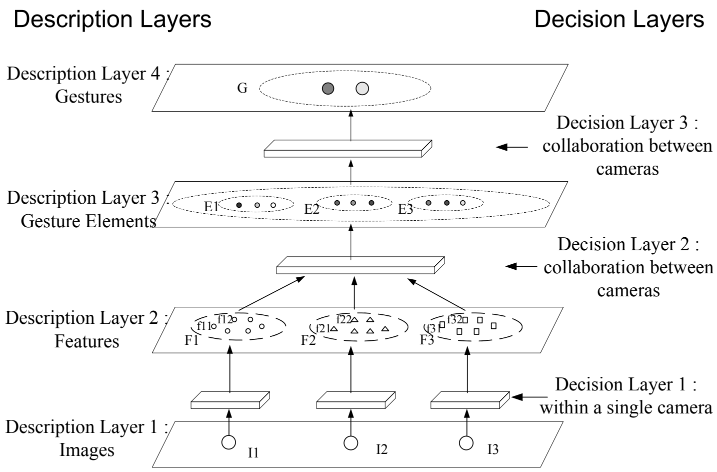


Fig. 8. The layered and collaborative architecture of the gesture analysis system. I_i stands for images taken from camera i; F_i is the feature set for I_i; E_i is the gesture element set in camera i; and G is the set of possible gestures.

around a local optimal. While if c_1 and c_2 are big, the particles are more "social" with the other particles and go quickly to the best positions known by the group. In our experiment, $N = 16$, $\omega = 0.3$ and $c_1 = c_2 = 1$.

Examples for in-node segmentation are shown in Fig. 7(a). Some examples showing images from 3 views and the posture estimates are in Fig. 7(b).

4 Towards Behavior Interpretation

An appropriate classification is essential towards a better understanding of the variety of passive gestures. Therefore, we propose a categorization of the gestures as follows:

- Static gestures, such as standing, sitting, lying;
- Dynamic gestures, such as waving arms, jumping;
- Interactions with other people, such as chatting;
- Interactions with the environment, such as dropping or picking up objects.

Fig. 8 illustrates the layered processing architecture defining collaboration stages between the cameras and the levels of vision-based processing from early vision towards discovery of the gesture elements.

To illustrate the process of achieving high-level reasoning using the collaborative vision-based architecture, we consider an application in assisted living, in which the posture of the user (which could be an elderly or a patient) is monitored during daily activities for detection of abnormal positions such as lying down on the ground. Each of the cameras in the network employs local vision processing on its acquired frames to extract the silhouette of the person. A second level of processing employs temporal smoothing combined with shape fitting

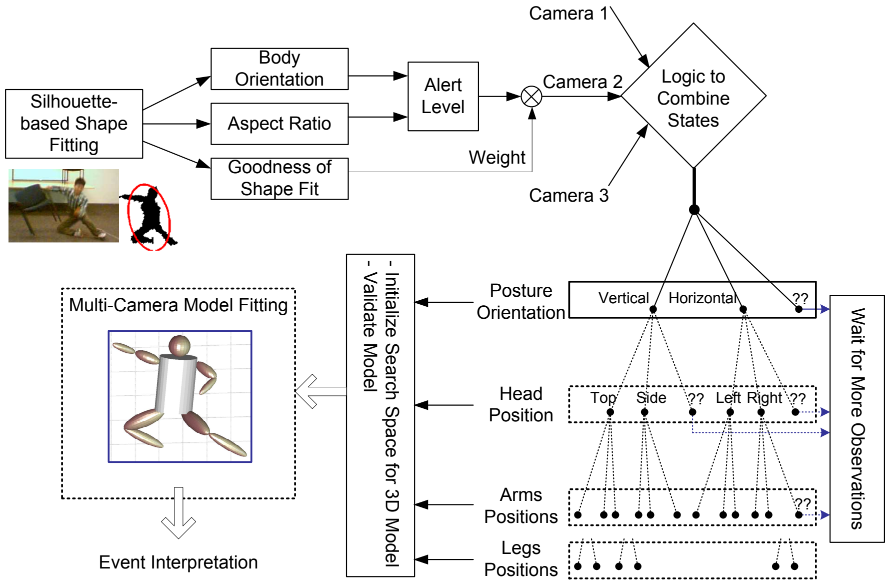


Fig. 9. A tree-based reasoning technique for fall detection. Qualitative descriptions can trace down the branches for specific event detection. The specific deductions can also be feedback for posture reconstruction.

to the silhouette and estimates the orientation and the aspect ratio of the fitted (e.g. elliptical) shape. The network's objective at this stage is to decide on one of the branches in the top level of a tree structure (see Fig. 9) between the possible posture values of *vertical*, *horizontal*, or *undetermined*. To this end, each camera uses the orientation angle and the aspect ratio of the fitted ellipse to produce an *alert level*, which ranges from -1 (for safe) to 1 (for danger). Combining the angle and the aspect ratio is based on the assumption that nearly vertical or nearly horizontal ellipses with aspect ratios away from one provide a better basis for choosing one of the *vertical* and *horizontal* branches in the decision tree than when the aspect ratio is close to one or when the ellipse has for example, a 45-degree orientation.

Fig. 10 illustrates an example of the alert level function combining the orientation and aspect ratio attributes in each camera. The camera broadcasts the value of this function for the collaborative decision making process. Along with the alert level, the camera also produces a figure of merit value for the shape fitted to the human silhouette. The figure of merit is used as a weighting parameter when the alert level values declared by the cameras are combined.

Fig. 11 presents cases in which the user is walking, falling and lying down. The posture detection outcome is superimposed on the silhouette of the person for each camera. The resulting alert levels and their respective weights are shared by the cameras, from which the overall alert level shown in the figure is obtained.

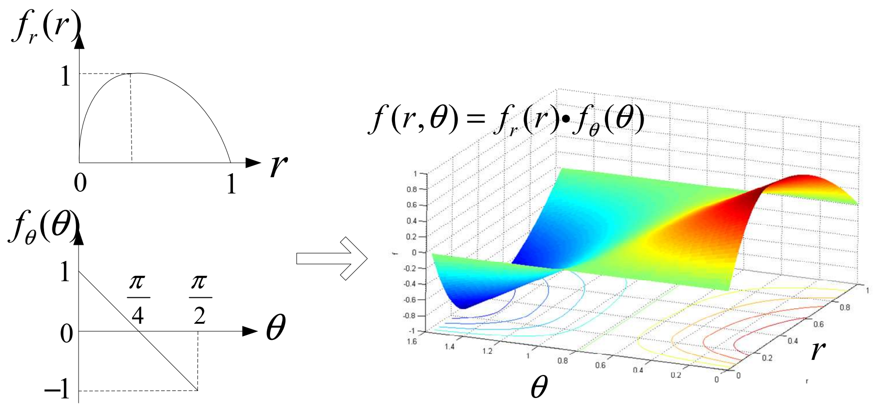


Fig. 10. The alert level functions based on the aspect ratio and the orientation angle of fitted ellipses.

5 Conclusions

In this paper we explore the interactive framework between vision and AI. While vision is helpful to derive reasoning building blocks for higher levels, there is more in the framework. We claim that the feedback between the vision module and the reasoning module is able to benefit both.

A framework of data fusion in distributed vision networks is proposed. Motivated by the concept of opportunistic use of available information across the different processing and interpretation levels, the proposed framework has been designed to incorporate interactions between the vision module and the high-level reasoning module. Such interactions allow the quantitative knowledge from the vision network to provide specific qualitative distinctions for AI-based problems, and in turn, allows the qualitative representations to offer clues to direct the vision network to adjust its processing operation according to the interpretation state. Two vision-based fusion algorithms were presented, one based on reconstructing the full-parameterized human model and the other based on a sequence of direct deductions about the posture elements in a fall detection application.

The current work includes incorporation of body part motion into the full-parameterized human body model allowing the model to carry the gesture elements in interactions between the vision network and the high-level reasoning module. Other extensions of interest include creating a link from the human model to the reduced qualitative description set for a specific application, and utilizing deductions made by the AI system as a basis for active vision in multi-camera settings.

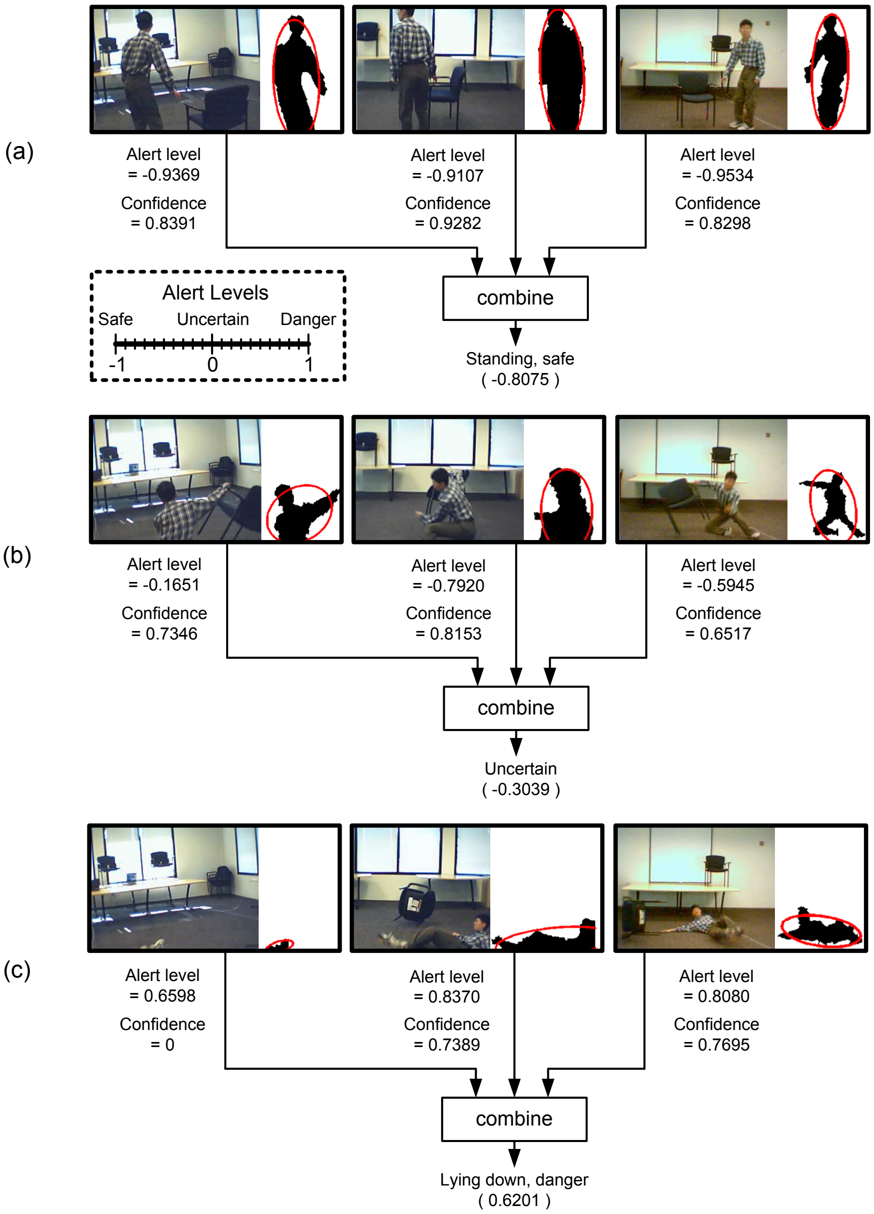


Fig. 11. Three sets of examples from three cameras of different views for fall detection. (a) standing; (b) falling; (c) lying on the ground. Alert levels and their confidence levels are shown. After combining observations from the three cameras a final score is given indicating whether the person is standing (safe) or lying (danger).

References

1. Kwolek, B.: Visual system for tracking and interpreting selected human actions. In: WSCG. (2003)
2. G. Ye, J. J. Corso, and G. D. Hager: 7: Visual Modeling of Dynamic Gestures Using 3D Appearance and Motion Features. In: Real-Time Vision for Human-Computer Interaction. Springer-Verlag (2005) 103–120
3. Aghajan, H., Augusto, J., Wu, C., McCullagh, P., , Walkden, J.: Distributed vision-based accident management for assisted living. In: ICOST 2007, Nara, Japan
4. Patil, R., Rybski, P.E., Kanade, T., Veloso, M.M.: People detection and tracking in high resolution panoramic video mosaic. In: Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS). Volume 1. (Oct. 2004) 1323–1328
5. Robertson, C., Trucco, E.: Human body posture via hierarchical evolutionary optimization. In: BMVC06. (2006) III:999
6. Rittscher, J., Blake, A., Roberts, S.: Towards the automatic analysis of complex human body motions. Image and Vision Computing (12) (2002) 905–916
7. Cucchiara, R., Prati, A., Vezzani, R.: Posture classification in a multi-camera indoor environment. In: ICIP05. (2005) I: 725–728
8. Björn Gottfried, Hans Werner Guesgen, and Sebastian Hübner: Spatiotemporal Reasoning for Smart Homes. In: Designing Smart Homes. Springer (2006) 16–34
9. Sidenbladh, H., Black, M.J., Sigal, L.: Implicit probabilistic models of human motion for synthesis and tracking. In: ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I, London, UK, Springer-Verlag (2002) 784–800
10. Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. (2000) II: 126–133
11. Cheung, K.M., Baker, S., Kanade, T.: Shape-from-silhouette across time: Part ii: Applications to human modeling and markerless motion tracking. International Journal of Computer Vision **63**(3) (August 2005) 225 – 245
12. Ménier, C., Boyer, E., Raffin, B.: 3d skeleton-based body pose recovery. In: Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization and Transmission, Chapel Hill (USA). (june 2006)
13. Mikic, I., Trivedi, M., Hunter, E., Cosman, P.: Human body model acquisition and tracking using voxel data. Int. J. Comput. Vision **53**(3) (2003) 199–223
14. Sidenbladh, H., Black, M.: Learning the statistics of people in images and video. **54**(1-3) (August 2003) 183–209
15. Wu, C., Aghajan, H.: Layered and collaborative gesture analysis in multi-camera networks. In: ICASSP. (Apr. 2007)
16. Weiss, Y., Adelson, E.: Perceptually organized em: A framework for motion segmentaiton that combines information about form and motion. Technical Report 315, M.I.T Media Lab (1995)
17. Ivecovic, S., Trucco, E.: Human body pose estimation with pso. In: IEEE Congress on Evolutionary Computation. (2006) 1256–1263

Division of Work During Behaviour Recognition - The SCENIC Approach

Kasim Terzić[1], Lothar Hotz[2], and Bernd Neumann[1]

[1] Cognitive Systems Laboratory, Department Informatik, Universität Hamburg
22527 Hamburg, Germany
{neumann|terzic}@informatik.uni-hamburg.de
[2] HITeC e.V. c/o Department Informatik, Universität Hamburg
22527 Hamburg, Germany
hotz@informatik.uni-hamburg.de

Abstract. Behaviour recognition in a video scene consists of several distinct sub-tasks: objects or object parts must be recognised, classified and tracked, qualitative spatial and temporal properties must be determined, behaviour of individual objects must be identified, and composite behaviours must be determined to obtain an interpretation of the scene as a whole. In this paper, we describe how these tasks can be distributed over three processing stages (low-level analysis, middle layer mediation and high-level interpretation) to obtain flexible and efficient bottom-up and top-down processing. The approach is implemented in the system SCENIC and currently applied to two domains: dynamic indoor scenes and static building scenes. We include details of an experiment where an ongoing table-laying scene is recognised.

1 Introduction

1.1 Application domains and requirements

Computer Vision in its most general form has been likened to silent-movie understanding [1], where people employ extensive common-sense knowledge about the physical world, typical situations, behaviour of people, and aspirations of individuals. On the first glance, behaviour recognition - which is addressed in the paper - appears to be a more restricted topic, with a focus on the recognition of very specific behaviours such as vandalism in a subway station [2], thefts at a telephone booth [3], filling up at a gas station [4], identifying activities at an airport [5] or placing dishes onto a table [6].

But at the moment one attempts to find a generic framework for behaviour recognition, one faces most of the challenges of silent movie understanding. So what are the challenges of generic behaviour recognition? In the following we propose eleven requirements which go beyond traditional single-object recognition and must be met by a system for behaviour recognition. The requirements pertain to a framework for *model-based* behaviour recognition, i.e. behaviour recognition based on explicit representations of behaviour concepts and the necessary procedures for recognising instances of such models in a concrete scene.

R1 Behaviours describe a scene at an abstraction level above the level of single-object trajectories, requiring qualitative and symbolic representations.

R2 Behaviours are typically embedded in a compositional hierarchy with increasing abstraction towards higher levels.

R3 Behaviours are often defined in terms of qualitative spatial relations between objects. These relations must be evaluated efficiently to support behaviour recognition.

R4 Similarly, behaviours may be defined in terms of temporal relations between parts which also must be evaluated efficiently.

R5 With behaviour recognition, we often face the task of interpreting scenes incrementally and in real-time along the temporal dimension.

R6 Behaviour recognition often involves part-whole reasoning, in particular guessing future behaviour from past observations.

R7 Part-whole reasoning and guessing the future means hypothesising interpretations and hence entails uncertainty management and the need for hypothesis revisions.

R8 Expectations generated by behaviour recognition provide a focus of attention and induce top-down guidance for further processing steps.

R9 Behaviour recognition may require that contextual information from other sources than the visual sensors be exploited.

R10 For behaviour recognition it may be necessary to resort to commonsense knowledge, beyond the knowledge about visual phenomena.

R11 Representation and interpretation facilities of the behaviour recognition framework must be domain independent and adaptable to specific application domains by declarative specifications.

Let us consider an example of the traffic domain to illustrate these requirements. A driver assistance system equipped with a front-view camera is supposed to warn the driver when a person is likely to enter the lane in front of the car. "Entering the lane in front up the car" is a qualitative concept (R1). It may be part of more complex behaviours, such as a pedestrian crossing a street or a child running after a ball (R2). "person on lane in front of car" as well as the behaviour described by "enter" involve qualitative spatial relations (R3). To recognise dangerous situations, the temporal relation between the expected car position and a person on the lane must be determined (R4). This must happen in real-time, keeping up with the evolving scene (R5). If a ball is observed running into the lane, this may be part of a possible event "child running after ball" and should cause a warning (R6). Depending on further circumstances (e.g. a clear view of the curb area), the hypothesis of "child running after ball" may be discarded (R7). A verification of this hypothesis may require focussed image analysis in an area where the child would be expected (R8). Context information, e.g. communicated from another car, may be available and must be considered (R9). The example "child running after ball" also illustrates a simple case of using common sense (R10). A more sophisticated warning system would, for example, also consider a possible fencing which would prohibit a child to enter the lane (R10). Finally, the same framework should be utilisable for - say - behaviour recognition in an elderly-care scenario (R11).

We suggest that a computer vision system for behaviour recognition should be designed to support these requirements as far as possible, and that claims regarding generality should be measured against these requirements. Of course, for specific tasks, it may be appropriate to devise special approaches. But in the interest of economical application developments there is a premium on reusable frameworks meeting all of these requirements.

In this paper we describe our approach towards generic behaviour recognition. In agreement with other existing system frameworks [4, 7–9] and conceptual studies [10, 11], our system consists of three major blocks as shown in Figure 1.

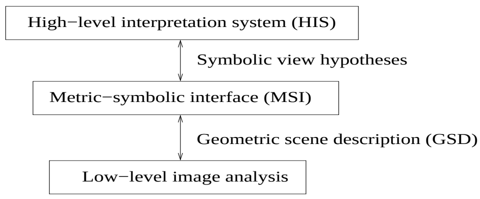


Fig. 1. Basic system structure for behaviour recognition

Low-level image analysis encompasses diverse image processing modules (IPMs) which compute a geometric scene description (GSD) in terms of segments, blobs or regions of interest (ROIs) tracked through the image sequence. The output is represented in terms of evidence objects which possess both a symbolic identity and a quantitative description. IPMs may be focussed or parametrised by top-down information.

The middle layer is called Metric-Symbolic Interface (MSI), but it has more than a mere interface function. One novel task arising from hypothesis generation in the high-level interpretation unit (R7) is to match top-down hypotheses with bottom-up evidence. This task differs from conventional bottom-up interpretation as (uncertain) hypotheses must be mapped into available evidence or may even trigger IPMs to provide further evidence. Another novel task of the MSI is related to the computation of spatial relations which play a significant role in high-level interpretations (R3). Qualitative spatial relations such as "touch" or "on" are natural constituents of symbolic high-level concepts, but they are grounded in the quantitative metrics of the GSD and can be computed much more efficiently using a map-based representation rather than the descriptions of symbolic objects. The same is true for temporal relations (R4) such as "approach" which also benefit from grounding in a metric representation. A dedicated data structure supporting the representation and computation of spatio-temporal relations has been postulated earlier [12, 13]. In our approach this data structure has a natural place in the MSI.

The high-level interpretation system HIS consists of a conceptual knowledge base and interpretation mechanisms. The conceptual knowledge base describes concepts for object categories, occurrences, behaviours and meaningful object configurations using aggregates as generic structures (R2). The underlying idea is that all high-level structures in a scene can be described in a homogeneous way as composite entities with spatially and temporally related parts. This approach differs from scenarios [9], situation-graph trees [14] or other structures which employ different representations at different abstraction levels, e.g. state-transition networks or Markov Chains for action sequences [3]. We believe that a generic framework (R11) should be able to represent arbitrary temporal relations between the sub-parts of a behaviour (and not only state transition sequences), as can be expressed, for example, by Allen's interval relations [15]. In our implemented system SCENIC (see Section 4) we use quantitative temporal constraints which realise a convex subset of Allen's interval relations [11]. The same formalism has also been proposed by [16].

Another advantage of our homogeneous object-oriented knowledge representation is the possibility to integrate behaviour knowledge with other commonsense knowledge (R10). The viability of this perspective was shown in [17], where description logics were investigated as a knowledge representation framework for scene interpretation. Description logics are known to provide the theoretical basis for knowledge representation with the Semantic Web language OWL[3].

The interpretation mechanism provided by our HIS is designed to deal with incomplete evidence - which is natural in evolving temporal scenes (R6) - as well as additional context information which may be available from other sources than low-level image analysis (R9). This flexibility is achieved by abstaining from an inbuilt interpretation strategy and allowing interpretation steps depending on the information on hand, for example conventional bottom-up steps for interpreting evidence as well as top-down steps for predicting future parts of ongoing behaviour or consequences of context information, for hypothesising occluded objects, for computing spatial and temporal relations in the MSI, or even for triggering focussed image analysis (R8). This general use of top-down steps is novel in existing systems. However, the same idea underlies the temporal prediction mechanism in [14].

In the following sections, we will concentrate on those aspects of our approach which we deem most interesting for the behaviour recognition community. In Section 2, we describe the MSI mediating between symbolic and metric representations. Section 3 presents the knowledge representation and interpretation facilities of the HIS. We use examples from the table-laying domain where the task is to recognise actions such as placing a cover on a table as part of various table-laying behaviours. In Section 4 we present a concrete experiment with our scene interpretation system SCENIC. Section 5, finally, concludes the paper with a summary and an outlook on future research.

2 Middle-level Processing

The metric-symbolic interface (MSI) connects the low-level scene analysis (tracking and primitive object classification) with the reasoning system. It takes input

[3] Web Ontology Language, $www.w3.org/TR/owl-ref/$

from both the low-level process (in terms of a GSD) and from the reasoning layer (in terms of hypotheses and requests). It has two important tasks: performing a spatiotemporal analysis, which turns the GSD into a set of high-level objects and occurrences such as *moves, touches* and *approaches*, and acting as an interface between the low-level image processing modules (IPMs) and the high-level reasoning system. As a part of this interface work, the MSI creates instances of high-level concepts from evidence, matches hypotheses to existing evidence, and passes information between the low-level and reasoning stages, e.g. initiating a focussed image analysis.

2.1 Low-level input

Although low-level video analysis lies outside of the scope of this paper, we will briefly describe the output of the low-level stage needed for the middle layer. This functionality was implemented as a part of a complete interpretation system for the table-laying scenario (see Section 4).

The low-level stage of video interpretation consists of two main steps: tracking of the objects in the scene and their classification. The tracking stage identifies all moving objects in the scene and assigns each primitive object a unique ID which is kept throughout the interpretation process. The image sequence is sampled at (usually regular) time intervals. The position (oriented bounding box) of each primitive object in motion is recorded for each time instant. The result is a quantitative description of the trajectories of all objects in the scene. Depending on the complexity of the domain, these objects may be blobs, regions of interest (ROIs) or at best regions corresponding to complete physical objects.

The appearance of objects carries important clues about possible classifications and primitive objects are pre-classified using one of many low-level classification algorithms. This is only as reliable as the algorithms used, may be ambiguous, and can be rejected by the high-level stage if it conflicts with other information. Nevertheless, classification is important for the initialisation of the high-level interpretation process.

The result of the low-level analysis is a quantitative description of the scene at each observed time point, consisting of a list of all primitive objects present, each described by: object ID, object class detected by a low-level classifier, position (centre of gravity), orientation, the oriented bounding box, and colour.

2.2 Spatiotemporal analysis

The spatiotemporal analysis within the MSI consists of three steps: calculating *perceptual primitives*, computing *qualitative primitives*, and detecting *occurrences* within the scene.

Perceptual primitives In the first step of the analysis, a set of functions is applied to the object properties from the GSD to obtain quantitative measurements for spatial and temporal relations. The results of these functions are called *perceptual primitives*. The intuition behind this step is to derive location- and timepoint-invariant descriptors. Typical perceptual primitives include:

– the rate of change of the position of an object (its velocity),
– the distance between the centres of two objects,
– the rate of change of this distance,
– the horizontal and vertical distance between axis-parallel bounding boxes of two objects, etc.

These primitives are still quantitative, but they are important for the detection of *primitive occurrences* such as a move (change in position), approach (change in distance) and touch/overlap (intersection of bounding boxes).

Qualitative primitives The second step involves a qualitative evaluation of the perceptual primitives. This is done by applying *predicates* to the perceptual primitives which compute a "qualitative constancy" for each time point, for example containment in a specific value range, being below a certain threshold or being approximately zero. Applying these predicates results in *qualitative primitives*, corresponding to notions like near, far, touching, approaching, moving away, stationary, etc. This process is illustrated with five common qualitative primitives:

– *Moving.* If the difference in the position since the last measurement is approximately zero, the object is *stationary.* Otherwise, it is *moving.*
– *Speed.* The speed of the movement can be qualitatively described by applying a threshold predicate on the rate of change of position of the object. The movement can then be described as *slow*, *fast* or other predicates.
– *Orientation.* By dividing the full circle into several intervals, orientation predicates can be defined relative to the image axes to describe whether the object points *forward*, *backward*, *left* or *right.* These predicates can also be applied to other reference axes, like the direction of movement.
– *Touching.* If the bounding boxes of two objects overlap, the objects are assumed to *touch.*
– *Nearing.* If the distance between two objects is decreasing, the two objects are *nearing* each other.

Spatiotemporal occurrences In the third step, *primitive occurrences* are built by combining qualitative primitives into units extending over time intervals of maximal length, and by creating more complex models. Primitive occurrences, such as move, approach or touch occurrences, form the basis for high-level reasoning.

A primitive occurrence is a concept which encompasses one or more qualitative primitives and a maximal time interval during which the qualitative primitives and possibly a certain set of constraints are always fulfilled. A primitive occurrence is defined by a start and end time, by the objects involved and the qualitative primitives which have to be true during this time period.

– *Move* is an occurrence where an object fulfils the *moving* qualitative primitive throughout a time interval. An interval between two successive move occurrences is a *stay* occurrence.

– *Approach* is an occurrence where the qualitative primitive *nearing* holds between two objects throughout a time interval.
– *Pair move* is a move involving two objects moving at the same speed into the same direction at each instant of an interval. An example is a cup on a saucer moving together.
– *Touch* is defined as an interval during which two objects *touch*, as defined in the previous section.
– *Touching move* is a pair move during which both objects *touch*.
– *Transport* is a touching move consisting of an object which can move by itself (e.g. a person or a hand) and an object which can be moved (e.g. a cup or a saucer).

These processing steps turn a quantitative GSD into a set of qualitative occurrences which can be represented symbolically and correspond to notions used in human perception, thus providing a basis for meaningful high-level concepts. Interesting events in many domains can be described using primitive occurrences of this kind, for example a person purchasing a ticket in a subway station: a person approaches the ticket machine, the person touches the ticket machine, the person moves away from the ticket machine, the person approaches a train.

2.3 Spatial and temporal indexing

The calculation of spatial relationships profits from a map-based representation. Looking for a left neighbour of a primitive object, for example, is simply a matter of traversing a corridor in a map containing all primitive objects, instead of performing an expensive comparison with all objects in the scene. The matching of hypotheses to evidence also profits from this type of representation, as the search can be concentrated on the part of the image confined by the hypothesis.

The map-based representation used in SCENIC is a grid dividing the image into rectangular fields. There is a map representing evidence (computed by low-level image analysis) and one representing views (representing hypotheses of the scene interpretation). Each field contains references to all evidence or view objects whose spatial extent intersects with it. Correspondingly, each evidence item and each view has a list of all fields that it covers in the evidence or view map, respectively. Thus, searching evidence for a hypothesis is turned into a simple lookup operation. The fields covered by a hypothesis in the view map are identified, and the corresponding fields in the evidence map contain references to all applicable evidence items that can be matched to the given hypothesis.

Since the GSD enters the middle layer in terms of data based on image frames, temporal indexing from each time point to objects of the GSD and to qualitative primitives is already available. Primitive occurrences, however, extend over intervals and are not naturally included in a frame-based representation. A top-down request asking for an occurrence in a specific time interval would require checking all primitive occurrences and comparing their begin and end times. Because of this, temporal indexing is extended so that each time point also contains a list of the references to all primitive occurrences taking place at this time point.

2.4 Evidence-view mapping

In addition to the spatiotemporal analysis of the GSD, the central role of the middle layer in SCENIC is matching real-world evidence to instances of object views in the high-level interpretation system. In an interpretation process, this matching may occur in two directions: bottom-up by assigning evidence to a view of a high-level object, or top-down by checking a view hypothesis against the available evidence.

The bottom-up case is a classification step which assigns existing evidence to the view class tied to an object class of the conceptual knowledge base. This step is ambiguous in general, as is well-known from single-object classification, and probabilistic guidance may be required for efficiency. As a result of the classification, a view instance (or in short: view) is created. This bottom-up step is typical for initialising the interpretation process.

In a top-down step, the middle layer receives a view hypothesis created by the interpretation system and has the task of confirming or refuting it. To do so, the middle layer can either match the hypothesised views to known evidence (already identified by the low-level system), or start a new low-level process to look for more evidence at the position indicated by the hypothesis. If a hypothesised view is matched to evidence, the hypothesis is *confirmed* and the evidence is linked to the hypothesis. Otherwise, the hypothesis is *refuted*. The reasoning system can take this new information into account.

In both cases, matches between evidence and views are recorded. If a particular match results in a conflict in the interpretation process, it can be withdrawn. Failed matches are also recorded to avoid repeating them in the future.

Due to the amount of raw data involved in the interpretation of even simple scenes, efficient indexing of information is extremely important when trying to match hypotheses to evidence. The spatial and temporal indexing introduced in this chapter significantly reduces the matching complexity by providing fast access to all evidence in individual space and time segments.

3 Reasoning Level

In our framework we view scene interpretation as a compositional task where the observed spatial and temporal occurrences in a video have to be composed into *aggregates* with increasing level of abstraction until a *scene interpretation* according to a given goal is reached. This composition is based on a declarative representation of the knowledge in a conceptual knowledge base, a *conceptual model* of a domain. In principle, this knowledge generically represents all scenes which may occur in a domain. In the table-laying domain, the conceptual model represents scenes about table laying actions that may occur in such a video.

This conceptual model provides the logical basis for the scene interpretations. In the following, we will shortly describe the knowledge representation language (Subsection 3.1), common aggregates for behaviour recognition (Subsection 3.2), knowledge and reasoning about aggregates for representing compositional (Subsection 3.3), spatial and temporal occurrences (Subsection 3.4) and the process of merging objects (Subsection 3.5).

3.1 Knowledge representation language

The knowledge representation language consists of the following facilities:

Concept Hierarchies. Object classes (*concepts*) are described using a highly expressive object description language, and embedded in taxonomic and compositional hierarchies. Object properties are specified by parameters with restricted value ranges or sets of values. A compositional hierarchy is induced by the special structural relation `part-of`. All concepts are compositional structures called *aggregates* except concepts without parts, which are called *primitive aggregates*. Objects selected for a concrete scene interpretation are instantiations of these concepts.

Constraints. Constraints pertaining to properties (relations or parameters) of more than one object are administered by a constraint net. Conceptual constraints are formulated as part of the conceptual knowledge base and instantiated as corresponding objects are instantiated. Constraints are multi-directional, i.e. propagated regardless of the order in which constraint variables are instantiated. At any given time, the remaining possible values of a constraint variable are given as ranges or value sets.

Task Description. A task is specified in terms of an aggregate which must be constructed (the *goal*) and possibly additional restrictions such as choices of parts, prescribed properties, etc. Typically, the goal is the root node of the compositional hierarchy governing the concepts which are relevant for the task.

Control Knowledge. Strategies for controlling the inference process can be specified in a declarative manner. For example, it is possible to prescribe phases of bottom-up or top-down processing conditioned on certain features of the evolving scene interpretation. As mentioned earlier, there is no inherent interpretation strategy built into the system.

This knowledge representation language is logic-based, general and thus, domain-independent. It is used to model knowledge 1) specific for behaviour recognition in general by specifying an appropriate *upper model* and 2) specific for a certain domain, like table-laying scenarios.

3.2 Upper model

The upper model enables the distinction between occurrences and parts of occurrences representing real world entities (i.e. 3D-objects and their behaviour). Views of primitive occurrences are identified by the middle layer and passed to the high-level system. Conceptually, views are instances of the concept `view` (or its specialisations) of the upper model (see Figure 2). The upper model also contains view classes for all distinct primitive occurrences that can be identified by the middle layer (e.g. of the type move, stay, touch, or approach).

3D-objects are instances of the concept `real-world-entity` or its specialisations. A 3D-object instance may be related to a corresponding view object linked to evidence in the scene, or may be hypothesised without evidence [18]. Further upper-model concepts that are specific to behaviour recognition, such as `transport`, `action`, `sub-action`, are discussed in the following subsections.

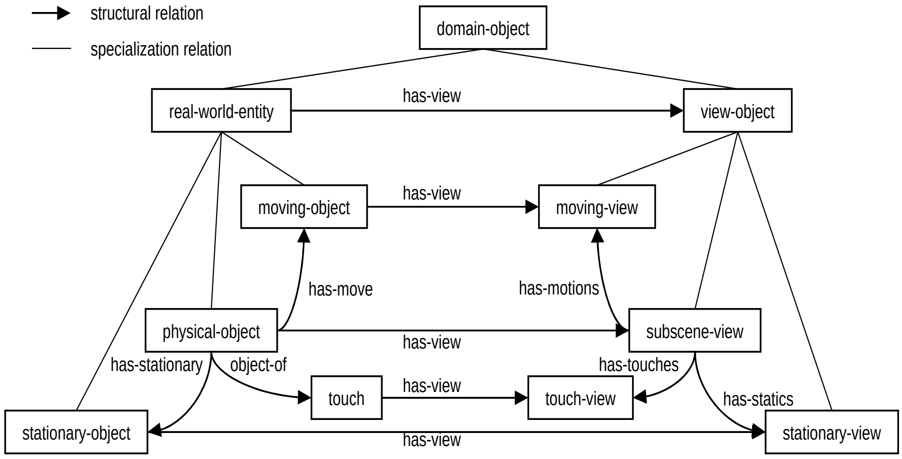


Fig. 2. Upper-Model for interpreting behaviours.

3.3 Reasoning with aggregates - integrating perceived parts and making hypotheses

A scene interpretation may be described as an aggregate composed of behaviours of constituent objects which in turn may be aggregates with constituent parts, etc. In the table-laying scenario, laying a complete cover (`create-cover-action`) may consist of laying a cup cover (`create-cupcover-action`) and laying of a plate cover (`create-basecover-action`). Laying a cup cover consists of transport occurrences of the involved objects (e.g. `hand-cup-transport` and `hand-saucer-transport`), a stationary occurrence representing the steady state of a laid cup cover (`cupcover`) and optionally an approach occurrence prescribing the decrease of distances between the object cup and the object saucer which are part of creating a cup cover (`cup-saucer-approach`).

Parts of an aggregate may be mandatory, optional or number-restricted. An example represented in our knowledge-representation language is given below:

```
(define-concept :name create-cupcover-action
  :super create-action
  :relations
   ((has-cup-transport       (:set (:some (a hand-cup-transport)    :min 1 :max 1)))
    (has-saucer-transport    (:set (:some (a hand-saucer-transport) :min 1 :max 1)))
    (has-spoon-transport     (:set (:some (a hand-spoon-transport)  :min 1 :max 1)))
    (has-cup-saucer-approach (:set (:some (a cup-saucer-approach)   :min 0 :max 1)))
    (has-cupcover            (:set (:some (a cupcover) :min 1 :max 1 )))
    (subaction-of (a create-cover-action))))

(define-concept :name create-cover-action
  :super create-action
  :relations
   ((has-subactions (:set (:some (a create-action) :min 2 :max 2)
                   :specializations
                        (:some (a create-basecover-action) 1 1)
                        (:some (a create-cupcover-action) 1 1)))
    (has-cover (:set (:some (a cover) :min 1 :max 1 )))
    (action-of (:or (a dinner-for-two-si) (a single-dinner-si)))))
```

```
(define-concept :name dinner-for-two-si
  :super cover-interpretation
  :relations
   ((has-actions (:set (:some (a create-cover-action) :min 2 :max 2)))))
```

Such concept descriptions are used to reason about the compositional structure of a scene in a top-down or bottom-up manner. For example, if a `create-cover-action` was instantiated for some reason, the appropriate parts are instantiated top-down (i.e. hypothesised objects are created). If a `hand-cup-transport` was instantiated, it is recognised as part of a `create-cupcover-action` and the corresponding aggregate is instantiated bottom-up. If variability occurs (for example `create-cover-action` can be part of `dinner-for-two-si` as well as `single-dinner-si`), a mechanism is needed for selecting one interpretation and evaluating it. In our approach, we currently use backtracking search, but probabilistic methods are also being developed.

3.4 Spatial and temporal reasoning

Conceptual descriptions of a scene involve spatial and temporal properties of occurrences and spatial and temporal relations between occurrences to a significant extent. The SCENIC approach supports this by providing appropriate concept parameters (like `tp-end`, `tp-start` for time intervals and `bb-left-upper-x` etc. for bounding boxes) and constraints related to these parameters.

When an occurrence is inferred for a video scene, the corresponding concept is instantiated with the spatial and temporal parameters as described above. They are initially set to specific values provided by the middle layer (if created bottom-up) or to intervals provided by the concept definition (which may be the open range of $[0 \ldots inf]$) in the case of top-down hypotheses. By processing the related spatial and temporal constraints in subsequent processing steps, the value ranges of parameters are further reduced, leading to final uncertainty intervals or conflicts causing backtracking.

We distinguish between *conceptual constraints* and *constraint relations.* Constraint relations are equations or inequalities about spatial and temporal parameters. Conceptual constraints describe a structural situation which is a precondition for evaluating certain constraint relations. For example:

```
(define-conceptual-constraint
  :name cup-before-saucer
  :structural-situation
   ((?cupcover-act :name create-cupcover-action)
    (?cup-tp       :name hand-cup-transport
                   :relations ((cup-transport-of ?cupcover-act)))
    (?saucer-tp    :name hand-saucer-transport
                   :relations ((saucer-transport-of ?cupcover-act))))
  :facts ((>= (?cup-tp tp-start) (?saucer-tp tp-end))))
```

specifies that, if a `create-cupcover-action` has a `hand-cup-transport` and a `hand-saucer-transport` (described with the structural situation), then the time point `tp-start` of the `hand-cup-transport` should be after (>=) the `tp-end` of the `hand-saucer-transport`. Conceptual constraints are specified for concepts and hold for every instance of these concepts (here for every `create-cupcover-action`).

Constraints can also be defined domain-independently for concepts of the upper model, e.g. for computing a bounding-box of a real-world entity.

3.5 Merging objects

A further reasoning service is needed when two objects were created independently but can be treated as the same object. In this case both object instances should be *merged* as this is will provide a simpler and hence preferable scene description. The need for a merge may occur, for example, when an object has been hypothesised top-down (say, a laying-a-dinner-for-two-action) and bottom-up processing of evidence has come up with the same hypothesis. Merging implies that the two objects are unified with all their properties and relations. In SCENIC, merging is accomplished by a conceptual constraint specifying all conditions which must be fulfilled by the two merging candidates.

4 System and Experiments

4.1 Architecture

The system SCENIC consists of five system components connected via remote procedure calls and file transfer[4]. This enables us to plug in different low-level algorithms and allows for distributed processing on several computers in a network. In the following we give details about each of the system components.

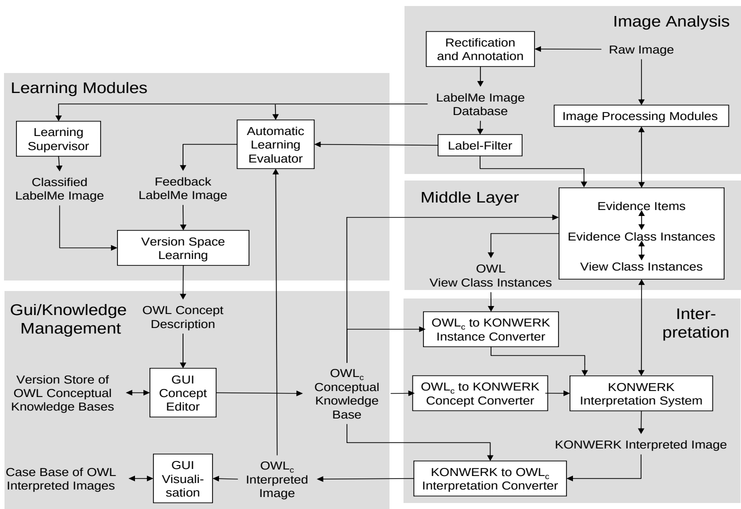


Fig. 3. Overview of SCENIC's modules.

[4] We are currently migrating from a stream-based list format (see below) to an XML-RPC interface (*www.xmlrpc.com*).

GUI/Knowledge Management The GUI has the following tasks: Interactive control of the three processing levels low-level image analysis, middle layer, high-level interpretation; presentation and depiction of results; management of distinct versions of the knowledge base. The knowledge base is implemented as an OWL knowledge base and augmented with constraints (depicted with OWL_c in Figure 3). The constraints are represented with a proprietary constraint language [19] enabling n-ary constraints at the concept level.

Image and Video Analysis Image and video analysis can be performed with distinct types of image processing moduls (IPMs) or manually by annotating images. For behaviour recognition we use a simple tracking unit and a model-based object recognition algorithm (see Section 2). The results of video analysis are represented using a proprietary format for the Geometric Scene Description (GSD), see Section 4.2.

Middle Layer The middle layer has access to the conceptual knowledge base in order to map IPM output to views which are instances of view concepts defined in the conceptual knowledge base. For behaviour recognition the middle layer mainly identifies trajectories in the GSD and recognizes move and touch occurrences.

Interpretation The interpretation module converts the OWL knowledge base and the input received from the middle layer into internal representations of the structure-based configuration system KONWERK [19–23], which is reused here for scene interpretation. KONWERK features an expressive concept language, a declarative control language, and inference capabilities based on specialisation relations and a powerful constraint system.

Learning The learning module is a separate module not relevant for the topics of this paper (see [24]). It provides aggregate concepts in the form of augmented OWL concepts.

4.2 Experiments

We have executed several experiments with SCENIC in the dynamic table-laying domain [6, 18] and the static building domain [25]. In this paper we focus on the interplay between high-level interpretation and middle layer in a dynamic scene. As input, we use a video where two human agents, sometimes acting in parallel, place dishes and other objects onto a table, for example, create covers as customary for a dinner-for-two. The tracking system identifies primitive objects in each frame, e.g.:

```
(FR 188 (ID 1  (PV TYPE SAUCER)(PV CENTER (435 191))(PV BOX (404 160 467 224))
               (PV SM(20 10 17 0 3 98)))
        (ID 2  (PV TYPE PLATE)(PV CENTER (110 274))(PV BOX (64 228 158 322))
               (PV SM(2 0 3 98 0 5)))
        (ID 3  (PV TYPE UNKNOWN)(PV CENTER (427 379))(PV BOX (411 369 445 387))
               (PV SM(0 0 0 0 0 0))))
        ...
(FR 216 (ID 1  (PV TYPE SAUCER)(PV CENTER (435 191))(PV BOX (404 160 467 224))
```

```
       (PV SM(20 10 17 0 3 98)))
(ID 2  (PV TYPE PLATE)(PV CENTER (110 274))(PV BOX (64 228 158 322))
       (PV SM(2 0 3 98 0 5)))
(ID 4  (PV TYPE SAUCER)(PV CENTER (209 311))(PV BOX (178 281 241 344))
       (PV SM(13 5 11 0 1 97)))
(ID 3  (PV TYPE SAUCER)(PV CENTER (437 199))(PV BOX (404 159 467 250))
       (PV SM(12 14 20 9 7 75)))
(ID 54 (PV TYPE HAND)(PV CENTER (211 362))(PV BOX (199 337 224 385))
       (PV SM(0 0 0 100 0 0)))
(ID 53 (PV TYPE HAND)(PV CENTER (477 272))(PV BOX (454 242 500 303))
       (PV SM(0 0 0 100 0 0))))
```

The MSI identifies movements, stationary occurrences and touches for all objects and stores the results in so called *motion frames* (see below). However, not all possible spatial data are initially created. For example, distance changes (i.e. approach occurrences) between all objects are not computed for combinatorial reasons.[5]

```
MOTION-FRAME:
object: #3 object-types: (SAUCER-VIEW UNKNOWN-VIEW CUP-VIEW)
type: GENERAL-MOTION  start: (-1000000000000 188) end: (228 230)
trajectory: ((-1000000000000 (427 379)) (190 (429 371)) (192 (430 358))
             (194 (431 345)) (196 (432 332)) (198 (433 319))
             (200 (434 307)) (202 (435 295)) (204 (436 284))
             (206 (437 274)) ...)
```

The high-level unit receives the move, stay and touch occurrences in form of instances of `moving-view`, `stationary-view` and `touch-view` as input. The interpretation process uses the conceptual model (see Section 3) as basis for interpreting the scene. In Figure 4 left, an intermediate scene interpretation is illustrated. Besides others, the system has recognized a `create-cupcover-action`. As defined in the model for `create-cupcover-action`, a `cup-saucer-approach` has to be present. The high-level system therefore creates a hypothesis for such an approach object with the appropriate time and spatial parameters, inferred from the transport objects (see Figure 4 right). This approach object is passed to the middle layer as feedback from high-level interpretation. The middle layer computes all approach objects in the given temporal and spatial region of interest and matches the given hypothesis against the computed evidence. It confirms the hypothesis and thus, supports the hypothesized interpretation.

5 Conclusions

In this paper, we have presented the SCENIC approach to video interpretation. This approach features a flexible mix of bottom-up and top-down processing steps and a division of tasks distributed over (i) a low-level stage for image analysis and tracking, (ii) a middle layer for matching evidence with primitive occurrences, and (iii) a high-level interpretation system for composing the scene description. The middle layer has several novel features: It supports selective computation of spatiotemporal relations using top-down guidance and

[5] One might argue that in a case as simple as our experimental table-laying scene, the combinatorial explosion of binary spatial object relations may be ignored. However, we aim at a system architecture which can be scaled up to more complex scenes.

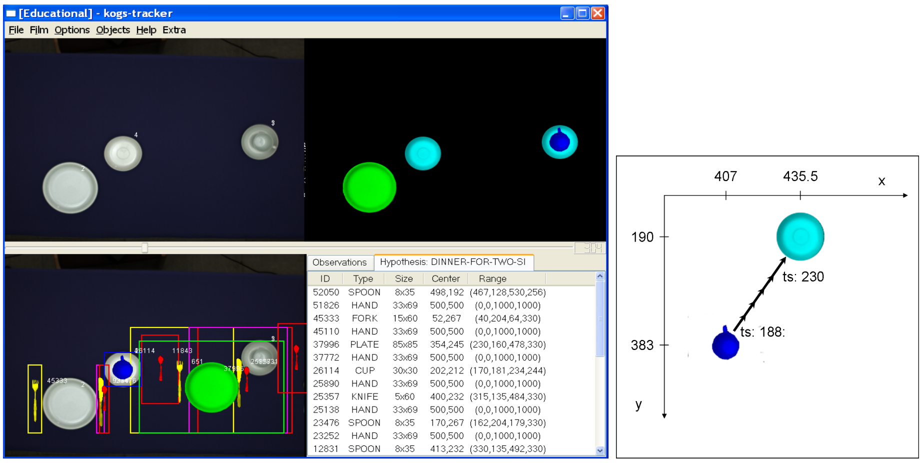


Fig. 4. Left: Intermediate scene interpretation as an instance of lay-dinner-for-two. Objects in natural colours are supported by evidence, objects in artificial colours are hypotheses based on high-level conceptual knowledge. Hypotheses are shown at the center of boxes, which represent possible locations. The low-level result is presented at top right; the original video at top left. Right: Hypothesized approach object for cup and saucer.

exploiting its map-based representations, and it evaluates top-down hypotheses by matching a hypothesis against available evidence or even initiating low-level image analysis processes. An experiment has been presented which illustrates feedback in form of a hypothesis from the high-level to the middle layer. Future work will include learnt concepts about scenes and a probabilistic guidance for the selection of interpretation steps.

Acknowledgement

This research has been supported by the European Community under the grant IST 027113, eTRIMS - eTraining for Interpreting Images of Man-Made Scenes.

References

1. Neumann, B., Novak, H.J.: Event Models for Recognition and Natural Language Description of Events in Real-World Image Sequences. In: Proc. of Fifth Int. Joint Conf. on AI IJCAI-83. (1983) 724–726
2. Vu, V.T., Brémond, F., Thonnat, M.: Automatic video interpretation: A novel algorithm for temporal scenario recognition. IJCAI (2003) 1295–1302
3. Hongeng, S., Bremond, F., Nevatia, R.: Representation and Optimal Recognition of Human Activities. In: IEEE Proceedings of Computer Vision and Pattern Recognition, South Carolina, USA (2000)
4. Nagel, H.H.: From image sequences towards conceptual descriptions. Image Vision Comput. **6**(2) (1988) 59–74

5. Thirde, D., Borg, M., Ferryman, J.M., Fusier, F., Valentin, V., Brémond, F., Thonnat, M.: A real-time scene understanding system for airport apron monitoring. In: ICVS '06: Proceedings of the Fourth IEEE International Conference on Computer Vision Systems, Washington, DC, USA, IEEE Computer Society (2006) 26
6. Hotz, L., Neumann, B.: Scene Interpretation as a Configuration Task. Künstliche Intelligenz **3** (2005) 59–65
7. Bauckhage, C., Hanheide, M., Wrede, S., Sagerer, G.: A cognitive vision system for action recognition in office environments. In: CVPR (2). (2004) 827–833
8. Heintz, F., Doherty, P.: DyKnow: A framework for processing dynamic knowledge and object structures in autonomous systems. In: Proceedings of the International Workshop on Monitoring, Security, and Rescue Techniques in Multiagent Systems (MSRAS). (2004)
9. Brémond, F., Thonnat, M., Zuniga, M.: Video understanding framework for automatic behavior recognition. Behavior Research Methods **3**(38) (2006) 416–426
10. Aisbett, J., Gibbon, G.: A General Formulation of Conceptual Spaces as a Meso Level Representation. Artificial Intelligence **133**(1-2) (2001) 189–232
11. Neumann, B.: Natural language description of time-varying scenes. In Erlbaum, L., ed.: Semantic Structures, D. Waltz (1989) 167–206
12. Mohnhaupt, M., Neumann, B.: On the Use of Motion Concepts for Top-Down Control in Traffic Scenes. In: Proc. ECCV-90, Springer (1990) 598–600
13. Gärdenfors, B.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge, MA, USA (2000)
14. Arens, M., Ottlik, A., Nagel, H.H.: Using Behavioral Knowledge for Situated Prediction of Movements. In: Proc. 27th German Conference on Artificial Intelligence (KI-2004). Volume LNAI 3238., Springer (September 2004) 141–155
15. Allen, J.F.: Maintaining knowledge about temporal intervals. Commun. ACM **26**(11) (1983) 832–843
16. Ghallab, M.: On chronicles: Representation, on-line recognition and learning. In: KR. (1996) 597–606
17. Neumann, B., Möller, R.: On Scene Interpretation with Description Logics. In: Cognitive Vision Systems. Volume LNCS 3948., Springer (2006) 247–275
18. Hotz, L.: Configuration Configuration from Observed Parts. In: Proc. of 17th European Conference on Artificial Intelligence (Configuration Workshop), Riva del Garda, Italy (2006)
19. Günter, A.: Wissensbasiertes Konfigurieren. Infix, St. Augustin (1995)
20. Soininen, T., Tiihonen, J., Männistö, T., Sulonen, R.: Towards a General Ontology of Configuration. Artificial Intelligence for Engineering Design, Analysis and Manufacturing (1998), 12 (1998) 357–372
21. Hotz, L., Wolter, K., Krebs, T., Deelstra, S., Sinnema, M., Nijhuis, J., MacGregor, J.: Configuration in Industrial Product Families - The ConIPF Methodology. IOS Press, Berlin (2006)
22. Cunis, R., Günter, A., Strecker (Hrsg.), H.: Das PLAKON-Buch. Springer Verlag Berlin Heidelberg (1991)
23. Günter, A., Hotz, L.: KONWERK - A Domain Independent Configuration Tool. Configuration Papers from the AAAI Workshop (July 19 1999) 10–19
24. Hartz, J., Neumann, B.: Version Space Learning of Spatial Structures for High-Level Scene Interpretation. eTRIMS EU-Project, Deliverable D2.4 (2007)
25. Hotz, L., Neumann, B., Terzić, K., Šochman, J.: Feedback between Low-Level and High-Level Image Processing. In: submitted. (2007)

Author Index