P–KAR 2012

PLACE-RELATED KNOWLEDGE ACQUISITION RESEARCH

Proceedings of the International Workshop

Monastery Seeon, Germany
August 31, 2012

Maria Vasardani, Stephan Winter, Kai-Florian Richter,
Krzysztof Janowicz & William Mackaness (Eds.)

Editors

Maria Vasardani, Stephan Winter & Kai-Florian Richter
Department of Infrastructure Engineering
The University of Melbourne
Parkville, VIC 3010
Australia
Email: {mvasardani, winter, krichter}@unimelb.edu.au

Krzysztof Janowicz
5806 Ellison Hall
Geography Department
University of California, Santa Barbara
Santa Barbara, CA 93106-4060
USA
Email: jano@geog. ucsb.edu

William Mackaness
School of GeoSciences
The University of Edinburgh
Geography Building, Drummond Street
Edinburgh EH89XP
UK
Email: William.mackaness@ed.ac.uk

Preface

This volume contains the proceedings of the *International Workshop on Place-related Knowledge Acquisition Research* (P-KAR 2012), held on August 31, 2012 in Monastery Seeon, Germany, in conjunction with the Spatial Cognition conference 2012.

Place has become a hot topic in GIScience: *place* is important in human cognition and communication, and hence, is a high priority for human-computer interaction. But *place* is also a challenging concept to model, reason with, and analyze in information systems, because of its fluency with context shifts, and its under-specification.

Place-related Knowledge Acquisition Research is using the concept of geographic place with the goal of building smarter services and integrating heterogeneous data. Thus, this workshop is built around the following challenge:

Achieve automatic estimation of the location of things or events based on verbal or graphical descriptions, or photographs, or a combination of them.

Imagine, for example, a geo-spatial service the interface of which allows for user interaction using place-based queries, references, or descriptions. Such a service may be able to deal with verbal descriptions, such as ‘the bar at the top end of Federation Square’, or with pictorial descriptions, where users sketch the location of said bar. For true interaction, the service needs to be able to *understand* these kinds of descriptions for both interpreting input, and for producing output. It should be able to estimate the location of a feature (in the real world) based on these descriptions, and to produce relevant descriptions of locations of features in response to user input.

The aim of the workshop was to bring together researchers from computational linguistics, data mining, artificial intelligence, geographic information science, and related disciplines, with a common interest in tackling this challenge. This volume contains their contributions in the form of peer-reviewed full papers or extended abstracts.

The P-KAR 2012 International Workshop featured a keynote talk, two presentation sessions of the accepted papers and abstracts, and a break out session where groups formed to discuss issues that emerged from the paper presentations, before conclusion. We would like to thank the Program Committee for their time, effort, and quality work in reviewing the papers, as well as all the authors for submitting their work for consideration. We are also grateful to our keynote speaker—Ross Purves (University of Zurich, Switzerland)—for accepting our invitation to participate in the workshop and address the participants, and everyone else that contributed to the workshop’s success.

August 2012

Maria Vasardani
Stephan Winter
Kai-Florian Richter
Krzysztof Janowicz
William Mackanes

Program Committee

Tim Baldwin	University of Melbourne, AU
Lawrence Cavedon	RMIT University, AU
Matt Duckham	University of Melbourne, AU
Allison Kealy	University of Melbourne, AU
Peter Kiefer	Institute of Cartography and Geoinformation, CH
Antonio Krueger	Saarland University, DE
Oliver Lemon	Heriot Watt University, UK
Ross Purves	University of Zurich, CH
Simon Scheider	University of Munster, DE
Lesley Stirling	University of Melbourne, AU
Stephan Winter	University of Melbourne, AU

Steering Committee

Stephan Winter	University of Melbourne, AU
William Mackaness	University of Edinburgh, UK
Krzysztof Janowicz	University of California, USA
Kai-Florian Richter	University of Melbourne, AU
Maria Vasardani	University of Melbourne, AU

Table of Contents

Keynote Talk

Abstract

As the call for this workshop stated: "Place has become a hot topic in GIScience". But what is place, and what are the implications of developing place-based methods? In my talk, I set out an agenda for place-based methods, and particularly for retrieval based on place. In doing so, I explore how unstructured text and tagged images can, firstly, inform methods which seek to take account of notions relating to place. Secondly, I illustrate how retrieval methods taking account of place can be developed using such insights, and discuss some successes (and failures!) in work developing such techniques.

Speaker's info:

Ross Purves is a lecturer in the Department of Geography at the University of Zurich in Switzerland. Previously he worked in the Department of Geography at the University of Edinburgh. His research focuses on two areas - environmental modeling and Geographic Information Retrieval. He has been involved, at various levels, in a variety of funded (and unfunded) projects in these areas. The TopIce project focuses on the influence of terrain representation on large-scale environmental modeling and its contribution to uncertainty of model results. The Swiss National Science Foundation funds the project. SPIRIT (Spatially-Aware Information Retrieval on the Internet) focused on the development of a spatially aware search engine, and was funded under the IST program of the European Commission. Project Tripod is also funded under the IST program of the European Commission, and will investigate the automatic captioning of images, based on their locations. While in Edinburgh, Ross was involved in a long-running program of research investigating paleo-climate change in Patagonia, through a variety of techniques. He is a strong believer in delivering high-quality teaching in universities, and was involved in the e-MapScholar project, which sought to develop innovative and customizable methods for the production of GIScience related online learning materials.

Identifying touristic places

Dominik Kremer and Christoph Schlieder

Research Group on Computing in the Cultural Sciences
University of Bamberg, Germany
{dominik.kremer,christoph.schlieder}@uni-bamberg.de

From the perspective of our research on geographic recommender systems, the goal of designing "systems and services that can understand 'place' in a way we humans do" seems little ambitious and potentially misleading. We argue that we should aim at designing services which are, in a specific sense, better than humans at understanding place. Every individual belongs to one or more social groups and, in general, the ability of humans to understand place conceptualizations from other social groups is rather limited. Geographic recommender systems should outperform humans in the handling of multiple group-specific place models (Matyas & Schlieder, 2009; Schlieder & Kremer, 2012, accepted).

One important consequence of looking at differences in conceptualizations has been pointed out by Schlieder and Henrich (2011). The classical membership problem of place research – does the point X belong to place P – transforms into more complex problems: does user U believe X belongs to P? Do users U and V share similar beliefs about X belonging to P? While this statement seems trivial according to our everyday life experience, its implications on geographic recommending are widely overlooked (Winter et al., 2009). Our talk illustrates the need for more complex place models by analyzing data about touristic conceptualizations of urban spaces derived from a recent GPS tracking study of touristic exploration behavior.

The data set is based on behavioral data from 17 first time visitors to the town of Bamberg, Germany, who came for a one-day-trip and volunteered to participate in the study. They were handed a camera equipped with GPS and a magnetic compass, as well as a second GPS receiver with better positional accuracy for recording the track data. The participants were told to explore the town in whatever way they liked and for how long as they pleased. A first analysis of the data is based on two dependent variables, the number of photographs taken in different regions of interest and the time the visitor spent there. We describe two marginal return models for place popularity, one based on photograph frequency, the other based on visit time. Individual differences in spatial choices are compared using these models.

A first group of findings relates to individual differences in the popularity of places while a second group of findings concerns the exploration strategies employed by the visitors. Within the former, we found that time-based and photograph-based measurements of interest in a place may significantly differ. We hold that this reflects a place's affordance for tourist activities. The popularity of some places is linked to their visual attractiveness while other places lend themselves for dwelling. In the analysis of the data, we use extended periods of zero motion speed as an indicator for dwelling behavior and the photographing

Vasardani M., Winter S., Richter K-F., Janowicz K. and Mackaness W. (eds.); P-KAR 2012, pp. 2-3

activity as an indicator for touristic attention. We found that tourists which move along nearly the same track, are likely to photograph different sights. Conversely, people that happen to catch the same glance will add semantics to it only according to their specific background, interests or even other places visited before. Based on our findings, we suggest the following (partial) answer to the workshop challenge of automatically detecting the location of things based on behavioral data:

– At least for most touristic places, there is no universally accepted location of the place that could be determined by machines (or by humans). Different place conceptualizations tend to coexist. The town of Bamberg, for instance, can be conceptualized as the beer capital of Bavaria or as the Unesco world heritage site of Baroque architectur or in many other ways.
– However, it is possible to analyze the spatial – and the thematic differences – of place conceptualizations by looking at data from close monitoring studies involving GPS tracks and photographs taken by visitors.
– We argue that the task of automatically detecting the location of a touristic place should not just map a place name onto a single geographic footprint, but rather on a set of footprints, different for different communities and different for different activities.

References

Matyas, C., & Schlieder, C. (2009). A spatial similarity measure for geographic recommender systems. In K. Janowicz, M. Raubal, & S. Levashkin (Eds.), *Geos-09, proc. geospatial semantics* (pp. 122–129). Berlin: Springer.

Schlieder, C., & Henrich, A. (2011). Spatial grounding with vague place models. *SIGSPATIAL Special Issue on Geographic Information Retrieval*, *3*(2), 20–23.

Schlieder, C., & Kremer, D. (2012, accepted). The cultural tourist's consumption of place: game or play? In *32nd international geographical congress, session on tourism mobilities and urban spaces.*

Winter, S., Kuhn, W., & Krüger, A. (2009). Guest editorial: does place have a place in geographic information science? *Spatial Cognition and Computation*, *9*(3), 171–173.

Reasoning About Large Places

Bernd Krieg-Brückner, Hui Shi

SFB/TR8 Spatial Cognition, Universität Bremen, Germany
{bkb,shi}@informatik.uni-bremen.de

Abstract. To support natural interactive way-finding tasks, computational formalisms of places are needed. In this extended abstract we present the idea of conceptual route graphs, which represent places using decision points, local route graphs and directional relations for modelling places in different way-finding situations. The application of formal spatial representations allows formal spatial reasoning about places, in particular deductions with qualitative calculi.

1 Small Places as Decision Points in Route Graphs

A framework for Conceptual Route Graphs has previously been presented, based on Route Graphs and the Double Cross Calculus [4]. In this approach, places and route segments are represented as nodes and directed edges of a route graph; routes are composed of consecutive route segments. Places act as decision points for the choice of a route segment, emanating from this place, to follow next in a route. When integrating several routes into a route graph, overlapping route segments are identified whenever their constituent places can be integrated. Thus *place integration* is a crucial notion: the orientation of outgoing route segments is computed relative to the *inherent* orientation of a place (analogously for incoming route segments); upon integration, this orientation is re-adjusted. The orientation of an outgoing from an incoming route segment is then computed as "the sum" of their orientations (relative to the inherent orientation of the place).

In practice, one would most likely define the inherent orientation of a place by reference to a salient landmark (e.g. "oriented towards the city hall"). When integrating the intermediate place of two consecutive route segments with a defined place in a route graph, the target/source orientation of the incoming/outgoing route segment is re-adjusted relative to the orientation of the defined place ("when incoming/outgoing, which orientation do I have w.r.t. the city hall?").

The (orientation at the) source and target places and the intermediate route segment can be taken as an element for the Double Cross Calculus (DCC) [1, 8], to reason about orientations of two consecutive route segments by composition in DCC.

2 Large Places with Local Route Graphs

Here, we extend the above notions for "small" places, suitable for reasoning about decision points in a network, e.g. in a corridor-like situation, with a framework

Vasardani M., Winter S., Richter K-F., Janowicz K. and Mackaness W. (eds.); P-KAR 2012, pp. 4-6

for “large” places that have a sizable extent, such that, to reach an outgoing from an incoming route segment, one has to “cross” the (supposedly vacuous) place, walk around its perimeter (as for a traffic circle), or do some even more complicated micro-navigation. First, we consider large places which cover a well-defined area (that constitutes a region in itself) with a border. We classify such large places according to their accessibility (and resulting navigability), e.g.:

open the area is an open navigable space, allowing cross-cuts;
closed the central area is closed, navigation is only possible around the perimeter (e.g. a traffic circle);
complex navigability is more complex (e.g. a traffic circle with an island, such that navigation around the perimeter permits occasional access to the island, which is a nested open space allowing cross-cuts).

In either case we may represent the large place by a local route graph that refines the place (as seen at a higher level of abstraction) into a route graph with a “higher resolution” for micro-navigation; incoming and outgoing route segments are connected to this local route graph such that it allows a transition from each incoming to every outgoing route segment (this condition has been specified for route graph refinements in [3]).

An open place is represented by a route graph that includes all possible connections of incoming to outgoing segments cutting across the open space; a closed place by a linear route graph representing the perimeter (mono- or bi-directional). The dynamic situation of a person being “in the middle” of an open space can be handled by a (dynamically moving) extra place representing the pose (position and orientation) of this person, with route segments connecting to (incoming and) outgoing segments.

3 Direction Relations for Large Places

As an alternative to route descriptions, human way-finding tasks often specify locations of salient landmarks in an open space. A recent experiment showed that such *scene descriptions* improve visualization, memorization and way-finding success, in an indoor environment [6]. Combining route descriptions with location specifications is required for navigation tasks in an environment containing large places where route graphs do not exist or are less explicit: from an incoming route ending at an exit location in the perimeter of a large place, navigation will be continued using a place description leading to an entry location of an outgoing route segment.

A place description comprises landmarks and directional relations. Since DDC is a model of relations between three points, or between a point and a directed route segment, and does not capture relations between objects with spatial extent, new models are needed for the formal representation of directional relations. For example, Goyal and Egenhofer’s direction-relation matrix [2], Skiadopoulos and Koubarakis’ projection based cardinal directional relations [7], or Kurata and Shi’s heterogeneous cardinal direction [5] are possible models for representing and deducing directional relations between spatial objects.

4 Conclusions

To model places in human or cooperative human-robot way-finding tasks, qualitative maps that integrate route graphs, point-based orientation models and directional models for representing places are needed. Depending on the level of abstraction, a place may be either, a decision point in a route graph, at which reorientation is needed to connect incoming and outgoing route segments; after refinement, a local route graph, which allows a transition from each incoming to every outgoing route segment; or a large open place, where scene descriptions are required to make the local connection.

Extending the conceptual route graph developed in [4] with directional models for place descriptions poses a new challenge. The focus will be on the connection of two route segments conjoined by a large place represented as a detailed conceptual route graph or modelled via a set of directional relations between spatial objects.

References

1. C. Freksa. Using orientation information for qualitative spatial reasoning. In A. U. Frank, I. Campari, and U. Formentini, editors, *Theories and methods of spatio-temporal reasoning in geographic space*, volume 639 of *LNCS*, pages 162–178. Springer, Berlin, 1992.
2. R. Goyal and M. Egenhofer. Consistent queries over cardinal directions across different levels of detail. In *11th International Workshop on Database and Expert Systems Applications*, pages 876–880, 2000.
3. B. Krieg-Brückner, U. Frese, K. Lüttich, C. Mandel, T. Mossakowski, and R. Ross. Specification of an Ontology for Route Graphs. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, and T. Barkowsky, editors, *Spatial Cognition IV: Reasoning, Action, Interaction. International Conference Spatial Cognition 2004, Frauenchiemsee, Germany, October 2004, Proceedings*, pages 390–412, Berlin, Heidelberg, 2005. Springer.
4. B. Krieg-Brückner and H. Shi. Orientation Calculi and Route Graphs: Towards Semantic Representations for Route Descriptions. In M. Raubal, H. Miller, A. Frank, and M. Goodchild, editors, *Geographic Information Science - Fourth International Conference, GIScience 2006*, number 4197 in Lecture Notes in Computer Science, pages 234–250. Springer, Berlin, 2006.
5. Y. Kurata and H. Shi. Toward heterogeneous cardinal direction calculus. In *Proceedings of the 32nd Annual German Conference on Artifical Intelligence*, volume 5803 of *Lecture Notes in Artifical Intelligence*, 2009.
6. V. Mast, C. Jian, and D. Zhekova. Elaborate descriptive information in indoor route instructions. In *Cognitive Science 2012*, to appear.
7. S. Skiadopoulos and M. Koubarakis. Composing cardinal direction relations. *Artificial Intelligence*, 152(2):143–171, February 2004.
8. K. Zimmermann and C. Freksa. Qualitative Spatial Reasoning Using Orientation, Distance, and Path Knowledge. *Applied Intelligence*, 6(1):49–58, 1996.

Component-wise Annotation and Analysis of Informal Place Descriptions

Igor Tytyk, Tim Baldwin

Department of Computing and Information Systems
The University of Melbourne
Melbourne VIC 3010, Australia
ihor.tytyk@gmail.com, tb@ldwin.net

Abstract. We analyse the strategies used in formulating situated informal location descriptions, by identifying geospatial expressions contained therein and annotating each for properties such as geospatial granularity and identifiability. Analysis of the annotations leads to insights such as the predominance of suburb-level expressions, and prevalence of vernacular expressions.

Key words: Informal place description, geospatial expression, named entity, vernacular geography, computational linguistics

1 Introduction

When informally describing one's whereabouts or giving directions, people make heavy use of place descriptions. In the descriptions they relate their location to the surrounding objects, or landmarks [1]. In order to make the instructions interpretable by the recipient, the description provider should use familiar landmarks and relate the location to them appropriately. Thus, for a human recipient this task is trivial. However, computational systems cannot easily interpret place descriptions expressed in natural language, or generate natural-sounding route or place descriptions.

Additionally, humans frequently make use of vernacular place descriptions, or refer to landmarks using non-standard renderings of their 'official' names, as a result, making it hard for computers to understand the description, and also humans unfamiliar with the locality being described. Wu and Winter state that placenames and spatial relations are main components of place descriptions, and in order to interpret the descriptions their components must be interpretable [2].

In this study we focus on analyzing placenames in the context of informal place descriptions, that is placenames that are elicited naturally and in situ, without any constraints or guidance. We manually identify geospatial expressions in a dataset of placename descriptions, and further annotate the granularity level, identifiability and normalised name of each such expression.

2 Dataset

Winter et al. collected situated place descriptions from players of the *Tell us where* location-based mobile game [3].[1] The game consisted of submitting textual descriptions of the location of smart phone users, along with their GPS location. The reasons we chose to use this data are many fold. First, the data was collected across a broad sample of users, ensuring the heterogeneity of the data and reducing sample bias. Second, the participants were asked to submit textual descriptions of their location from anywhere in the state of Victoria, Australia. This led to a diversity of locations, but within a restricted area of familiarity to our annotators and with the expectation of consistency in the strategies used by the participants to describe their location. Third, the users were given no guidelines for writing the descriptions, meaning that the data is rich in vernacular placename descriptions and the strategies used by users to describe their location are varied. Lastly, since the participants were using their mobile phones and basing the placename descriptions on their actual location. As a result, the descriptions are situated, spontaneous, and as natural as we could hope for.

A total of 2221 place descriptions were collected through the *Tell us where* game. However, the data contained duplicates. Since we are interested in qualitative rather than quantitative data, it was decided to eliminate all duplicates from the corpus. As a result, the final number of descriptions was 1858.

2.1 Annotation

We manually annotated the placename descriptions for *geospatial expressions*, in the form of: (1) geospatial named entities (*Federation Square, Swanston Street*); and (2) geospatial noun phrases (*school, a leafy park*). Named entities are proper names, and are generally subclassified according to the semantic class of the referent, e.g. into persons, locations and organisations. However, for the purposes of this research, we restrict our attention to geospatial named entities.

One of the broader goals of this work is the compositional semantic interpretation of place descriptions. It was thus decided that we should aim for maximum segmentation granularity in our annotation, while avoiding nested annotations. For example, if the place description were an address such as *Melbourne University Bookshop, in Parkville near the library*, we would segment it into the geospatial named entities *Melbourne University Bookshop* and *Parkville*, and the geospatial noun phrase *the library*. Note that we would not also identify *Melbourne University* as a geospatial named entity, as it is nested within *Melbourne University Bookshop*.

We expected many of the geospatial expressions in the dataset to be noun chunks. For example, *Queen Victoria Market* is a single noun chunk geospatial named entity, while *a tall building* is a single noun chunk geospatial noun phrase

[1] `http://telluswhere.net/`

Granularity level	Description
(1) Furniture	Location within a room, referring to furniture (*by my computer, in bed*)
(2) Room	Location within a building, or parts belonging to it (*in my room, third floor*), or medium-sized vehicles (*car, train*)
(3) Building	Location of a building, street no. or building name (*geomatics dpt, street corner/intersection,*)
(4) Street	Institution, public space or street level, larger than building and/or vaguer boundaries than building. For example, transport infrastructure (*railway, tramline, Ave, Circuit*), a public space (*school, cemetery, mall*), or a natural landmark (*lake, park*)
(5) District	Suburb, rural district or locality, or post code area (*carlton, North Melbourne, CBD*)
(6) City	Town or city level, and metropolitan areas (*Canberra, near Geelong*)
(7) Country	Everything beyond city level, including highways, freeways (*Princes Hwy*), islands (*French Island*), rivers (*Murray river*) and states (*WA*)

Table 1. Granularity level classification (Richter et al., 2012); all examples are taken from the actual dataset, and are presented using the original orthography

referring to a construction, which can be used as a reference point when describing a location. In the interests of expediting annotation, we first chunk-parsed the place descriptions, using the Stanford CoreNLP tools.

The annotation scheme we used is comprised of several layers. The first annotation layer contains information about whether a given segment is a geospatial named entity (*NE_NP*) or a geospatial noun phrase (*NP_NP*). The remaining layers apply to each geospatial expression.

The second layer of annotation is the *granularity level*, and captures the “zoom level” of each geospatial expression. The granularity level is judged on the scale from 1 to 7, based on the classification of Richter[4] as detailed in Table 1. In some instances, we diverge from Richter’s classification. For example, when a named entity is too big or too small for the bounding box of its default zoom level, we override the default to capture the zoom level which best matches the size of the bounding box. *Mountain Highway*, e.g., goes through only a few suburbs of Melbourne, so we override the *Country* granularity level for highways and assign it to the zoom level of *City* to better reflect its size. Similarly, when determining the granularity level of towns, it was decided to shift the small towns that do not have suburbs (e.g. *Warragul* and *Pakenham*) from *City* to *District*.

The third layer of annotation is *identifiability*. This captures whether a geospatial expression is unique within Victoria or there are multiple instances of it. There are three possible values for identifiability: *non-identifiable*, *identifiable ambiguous*, and *identifiable non-ambiguous*. All geospatial noun phrases (e.g. *school, park, monument*) are non-identifiable, since the set of these objects within Victoria is very large and it is not possible to geocode them without disambiguating information. Some geospatial named entities are considered to be non-identifiable due to their ubiquity and unavailability within standard gazetteers of an exhaustive listing of every instance within Victoria (e.g. *Mc-*

Donalds, 7-eleven). On the other hand, a geospatial named entity can refer to a small set of several places which are enumerated in a gazetteer, in which cause they are considered to be identifiable ambiguous. For example, there are four instances of *Canning Street* in Victoria, so every *Canning Street* in the corpus is annotated as identifiable and ambiguous. On the other hand, *Flemington Road* is identifiable non-ambiguous as there is only one instance in Victoria.

As with granularity, the determination of identifiability is inevitably subjective. To reduce the effects of subjectivity as much as possible, we base the judgement on two online gazetteers: OpenStreetMap[2] and Google Maps.[3]. Google Maps contains an extensive listing of named entities, but has poor coverage over non-standard or vernacular equivalents of less well-known named entities. Thus, while *melb uni* (standard = *The University of Melbourne*) and *fed square* (standard = *Federation Square*) can be found in Google Maps, it does not contain local vernacular such as *broady* (standard = *Broadmeadows*) or non-standard abbreviations such as *pi* for *Phillip Island* or *fg* for *Ferntree Gully*. Here, we elicited support from locals and the Google search engine to interpret the geospatial expression.

Names of cafes, restaurants, and other small businesses were the most difficult NEs to judge identifiability for. Even though OpenStreetMap lists a vast number of buildings, eating places, shops, many of them were missing.

The fourth and final level of annotation is the *placename normalisation*. Since the place descriptions were submitted by mobile phone, the dataset contains a lot of abbreviations, misspellings and vernacular names. The canonical name/spelling was provided in all such instances. For example, *melb uni* would be normalised to *The University of Melbourne*. We observed an inevitable dependency between identifiability and placename normalisation for geospatial named entities: if a geospatial named entity cannot be identified, it is not possible to determine its normalised rendering.

Some of the submitted place descriptions do not contain any information about the location (e.g. *this will be an everlasting love*) or are located outside of Victoria (e.g. *in Wagga Wagga*). All such descriptions were marked as irrelevant at the message level, using the *IRREL* label.

For the annotation we used brat,[4] a highly-configurable, easy-to-use web-based text annotation tool.

3 Analysis and Discussion

Having annotated the dataset, we extracted a feature vector for every annotated geospatial expression (excluding the irrelevant descriptions). Each feature vector contained a set of values: *id*, geospatial expression *type*, *granularity level*, *identifiability*, *original* spelling, and *canonic* (normalized) spelling. Then, all the vectors

[2] `http://www.openstreetmap.org`
[3] `http://maps.google.com.au/`
[4] `http://brat.nlplab.org/`

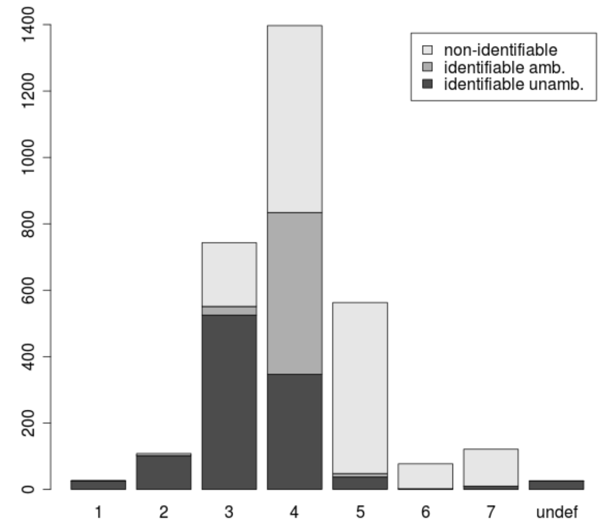


Fig. 1. Granularity level vs. identifiability in the dataset

were collated into a table and fed into the R statistical package[5] for analysis. In total, 3061 geospatial expressions were extracted, 2139 (70%) of which were geospatial named entities. That is, without any constraint on the description, about two thirds of geospatial expressions contained in place descriptions can potentially be found in gazetteers.

Figure 1 presents a distribution of geospatial expressions across zoom levels, broken down by identifiability. The mean granularity value is 4.01, with a standard deviation of 1.05. The most common granularity level is 4 (*Street*), with about 45% of all geospatial expressions. This means that when writing place descriptions, users tend to make heavy use of streets, parks, squares, universities and hospitals. Of the remainder, almost a quarter (24%) of the referents are of the *Building* granularity level (level 3), and about 18% are of the (*Suburb*) granularity level (level 5).

The correlation between the granularity level and the fraction of non-identifiable placenames is not very surprising: the bigger the spatial feature, the more likely it will be identifiable. On the other hand, the appreciable drop in non-identifiability at the *Suburb* level is proof of the salience and unambiguity of the placenames within this level. After dividing all the geospatial expressions by identifiability and filtering out from the non-identifiable ones the names of chain stores and eating places (e.g., *McDonald's, Subway, Coles*), it is possible to calculate how many of the named entities are not in the gazetteers (Open-

[5] `http://www.r-project.org/`

StreetMap and Google Maps). Out of 2139 named entities, 51 (2.4%) are not contained in the gazetteers. As a rule, among these placenames are names of restaurants, apartment blocks, and other small scale companies (e.g. *Pilkington Glass, Ching Chong Food, Yarra Crest Appartments*).

Another important category of geospatial expression is vernacular descriptions. We found a considerable number of entrenched vernacular equivalents of salient Victorian placenames, and common strategies for forming vernacular place names. Some of them are formed by simply dropping one of the constituent words (*Narre Warren* → *narre*), some by "clipping" the word (*Yackandandah* → *yack*, *Dandenong* → *dande*), and some are acronyms (*Phillip Island* → *pi*, *Ferntree Gully* → *fg*). However, the most productive pattern was "embellished clipping", shortening the expression to the first syllable and adding a diminutive suffix *-y, -ie*, (e.g. *Richmond* → *richy*, *Beaconsfield* → *beacy*, *South Gippsland Highway* → *south gippy*). The pattern is particularly peculiar to the Australian English. From the collected informal NEs, one can infer that only salient and unambiguous placenames undergo the process of vernacularization. Since suburb names in Victoria are unique and widely used for describing locations, they are most commonly substituted by their informal equivalents.

4 Conclusions and Future Work

In this paper, we have performed detailed component-wise analysis of informal place descriptions. From this study, we can conclude the following: (a) most geospatial expressions are streetnames, parks, buildings and suburbs; (b) the presence of a suburb-level placename in the description increases its identifiability; (c) vernacular place descriptions are commonly used, based on a small number of strategies; and (d) geospatial named entities which are mostly likely to not be contained in gazetteers are names of pubs, cafes, and small businesses.

This paper has considered placenames independently of the message-level interpretation. A logical next step is a compositional analysis of the place description based on the annotations we have done, and investigation of how spatial relational semantics (e.g, prepositions like *near, at, in*) impacts on message interpretability and the properties of its constituent placenames.

References

1. A. Klippel. *Wayfinding Choremes Conceptualizing Wayfinding and Route Direction Elements.* PhD thesis, Universitaet Bremen, 2003.
2. Y. Wu and S. Winter. *Interpreting Destination Descriptions in a Cognitive Way.* Schloss Dagstuhl, Dagstuhl, 2011.
3. S. Winter, K.-F. Richter, T. Baldwin, L. Cavedon, L. Stirling, M. Duckham, A. Kealy, and A. Rajabifard. Location-based mobile games for spatial knowledge acquisition. In Janowicz et al., editor, *Cognitive Engineering for Mobile GIS*, Belfast, Maine, USA, 2011.
4. D. Richter, M. Vasardani, L. Stirling, K.-F. Richter, and S. Winter. Zooming in zooming out hierarchies in place descriptions. Unpublished manuscript, 2012.

Classification of Localization Utterances using a Spatial Ontology

Mohammad Fazleh Elahi[1], Hui Shi[1], John A. Bateman[1],
Kathleen M. Eberhard[2], Matthias Scheutz[3]

[1] SFB/TR8 Spatial Cognition, Universität Bremen, Germany
{fazleh, shi, bateman}@informatik.uni-bremen.de
[2] Department of Psychology, University of Notre Dame, USA
eberhard.1@nd.edu
[3] Human-Robot Interaction Lab, Tufts University, USA
mscheutz@cs.tufts.edu

Abstract. Dialogue systems for spatially situated tasks need to provide referential descriptions of spatially located objects and understand such descriptions from users. To construct such dialogue systems, it is useful to investigate how humans describe object locations in their immediate environment and how they ask about object locations in remote environments. In this paper, we address the semantic classification of the localization utterances found in the CReST corpus, which is a dialogue corpus of humans performing a cooperative, remote, search task. The aim is to explore the relation between specific semantic configurations and the dialogically and situationally embedded linguistic forms employed. Specifically, we first extracted different types of localization utterances from the corpus and then paired these with semantic categories provided by the linguistically motivated spatial ontology GUM. The paper concludes with a discussion of the characteristics of different types of localization expressions on the basis of spatial concepts and descriptions employed.

1 Introduction

The localization of objects in an indoor environment poses a considerable challenge for dialogue-based intelligent systems, which are employed in a variety of tasks, such as mapping and localization Kruijff *et al.* [6], and indoor wayfinding Cuayahuitl *et al.* [3]. However, the support for object localization in intelligent systems is currently rather limited. For effective interaction, such dialogue systems should be able to deal with place-related knowledge acquired from verbal expressions, such as "now you're in the long hallway", in which the place "hallway" is described as "long". This requires them to support place-related information extraction, representation and reasoning. Our aim here is to explore the semantic classifications of localization utterances, a fundamental effort toward the construction of sophisticated place-aware spoken dialogue systems.

Vasardani M., Winter S., Richter K-F., Janowicz K. and Mackaness W. (eds.); P-KAR 2012, pp. 13-18

Spatial localization has two main aspects: location description and location query. The recently collected human-human dialogue corpus CReST (cooperative, remote, search task) of Eberhard *et al.* [4] covers both communicative aspects and, therefore, provides useful data. In addition, research investigating the relation between linguistic expressions involving space and semantic representations is of relevance (e.g., broad-coverage characterizations of semantics for naturally occuring texts and dialogues [8]). Hence, we will focus on the relationship of natural linguistic expressions with a broader range of more formal spatial characterizations capable of covering more of the semantics of naturally occuring, task-based expressions, combining a data-driven and a semantic classification approach.

The Generalized Upper Model (GUM) is a general task and domain independent "linguistically motivated ontology" that provides linguistic semantics for spatial expressions. It serves as an intermediate "interface ontology" mediating between linguistic forms and contextualized interpretations. In previous work, Bateman *et al.* [2] employed GUM to classify spatial relations in three spatial language corpora: the *Trains* 93 Dialogues (Heeman & Allen [5]), the *HCRC Map Task* (Anderson *et al.* [1]), and the *IBL Corpus* (Lauria *et al.* [7]), and showed that the GUM spatial ontology provided a characterization of the semantics that was of immediate use for several natural language processing tasks. Here we follow this line of research and present the semantics of a set of localization expressions from the CReST corpus using GUM, with focus on static spatial configurations and their elements.

2 Corpus Analysis

The CReST corpus [4] of natural language dialogues was obtained from an experiment involving humans performing a cooperative, remote, search task. The experiment required individuals in a dyad to coordinate their actions via remote audio communication in order to accomplish several tasks with target objects ("colored boxes") that were scattered throughout an indoor search environment (see Figure 1). Neither individual was familiar with the environment before the experiment. One individual was designated as the director (D), and the other as the searcher (S).

The CReST corpus is particularly relevant for the localization problem because the experiment directly involves several distinct localization subtasks. In one subtask, the searcher was to report the locations of eight green boxes in the environment and the director was to mark their locations on the map. Hence, the director needed to learn about the boxes' locations through dialogue. In a second subtask, the director was to direct the searcher to the cardboard box at the furthest point in the search environment, while the searcher was to collect blocks from blue boxes and put them into the cardboard box. Here, the director needed to track the searcher in the environment (from dialogue only, with no visual feedback) and then give further instructions to direct the searcher to the blue boxes.

In order to extract probable relatum and locatum from the dialogues, three wordlists were created. The first wordlist (BP) contains the words related to the places in the environment (room, cubicle, office, hallway, doorway, etc), the second (FO) words related to the fixed objects (door, wall, steps/stairs, stage, booth, etc), and the third (MO) words related to the movable objects (computer, chair, table, shelf, cabinet, etc). The localization-related utterances of the directors were first extracted on a word-by-word basis through comparing each word with those in the wordlists. These utterances were then processed using a set of tools (e.g., tokenizer, part-of-speech tagger, and regular expression based NP chunker) developed in NLTK toolkit[1], in order to find the meaningful noun-phrase(s) (NP) of individual utterances. For instance, NP [the/DT little/JJ tiny/JJ room/NN] is the meaningful phrase of the utterance "is it in the little tiny room?".

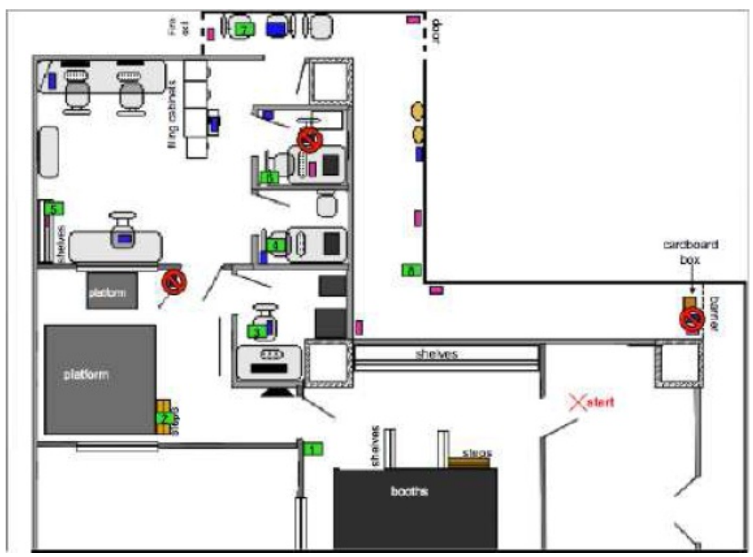


Fig. 1. The figure shows the actual locations of the eight green boxes in the environment which the director was to indicate on the map (CReST; Eberhard *et al.*: [4]). The map of the search environment displayed to the director on a computer screen is as same as this map but without green boxes. As can be seen from the figure, the environment consisted of three big rooms, three small rooms and a surrounding hallway.

Analyzing the number of utterances of each group, we find that rooms of the wordlist BP and doors of the wordlist FO are most frequently used. Moreover, we found that "small", "big","tiny", "first", etc. (see Table 1) are often used to distinguish rooms.

[1] http://nltk.org/

Table 1. Some example utterances from different groups

Group	Examples	Descriptions from the corpus
BP	are you still in the smaller room? okay now you're in the long hallway?	little, small, big, long, large, tiny, initial, first, third, etc
FO	there's a open door in front of you right? is there a door with a single door?	open, close, single, double, small, two, etc
MO	are there filing cabinets you see? are there two chairs at the desk?	filing cabinets, two chairs, etc

3 Semantic Classification of Localization Expressions

The GUM ontology [2] is a *linguistically motivated* ontology based on grammatical evidence from a broad range of linguistic spatial expressions which provides a unified account of spatial concepts and their relations. We use the GUM concept SpatialLocating (SL) to define and classify the semantics of the localization utterances in CReST for two reasons. First, GUM's SpatialModality (SM) covers distance, direction and relationship between the locatum (L) and relatum (R) which are crucial for localization expression. Second, GUM provides spatial semantics to localization expressions regardless of contextual interpretations. As can be seen from table 2, in GUM, "you were in the little cubicle, right?" and "are there like a desk in front of the computer chair?" receive spatial semantics SpatialLocating, where SpatialLocating is the concept that specifies the place (as a placement relation) where an entity (as a locatum) is being positioned. A placement relation can be a GeneralizedLocation (e.g., "on the table") or a GeneralizedRoute (e.g., "through the hallway"). The GeneralizedLocation binds together a relatum and a spatial relationship hasSpatialModality within a single structured entity that may stand in a placement relation within a spatial configuration. The utterance "There is an open door in front of you", for example, is bound with the following GUM semantics:

Configuration:	SpatialLocating	
Locatum:	an open door	
Placement:	GeneralizedLocation	
	hasSpatialModality:	FrontProjectionExternal
	relatum:	you(person)

Exploring the different forms of localization expressions, we found several utterances (e.g, 19.5% utterances of SpatialLocating), in which the locatum (L) is unknown and the relatum (R) is either an object or a person (e.g, the searcher). As can be seen from table 2, the locatum of the sentence "what do you see immediately to your left?" (Configuration: SpatialLocating, Locatum: unknown, Relatum: person, SpatialModality: LeftProjectionExternal) is unknown. From the analysis of spatial relations between a locatum and a relatum, we have found that there are many utterances (e.g, 33.8% utterances of SpatialLocating) related to the GUM concept Containment of a person (e.g., "Are you in a hallway or small room?") or an object (e.g., "Is there like a room that has a bunch of desks with tables and chairs?") in a spatial place (i.e., BP). This is because locating the person first and then the entities that can be seen from that place is essential for solving localization problems in a search task. The GUM concepts ProjectionRelation (e.g., "on the left"), MultipleDirectional (e.g., "it's in the bottom right or bottom left?"), and GeneralDirectional (e.g., "on the opposite side of the filing cabinets") are frequently used for directional relations in the corpus, together with clauses containing words such as "face" or "look" for orientation; for example: "Okay three steps and now when you look to your right there's another open door." and "if you look outside the cubicle there should be a door to your right is that correct?".

Many utterances (e.g, 35.1% utterances of SpatialLocating) in CReST contain what in GUM terms are called ComplexConfigurations, in which the director describes two or more spatial locations in a single utterance. The specifications of such utterances usually combine several SpatialLocating configurations with conjunction ("SLCjSL") or disjunction ("SLDjSL") as shown in table 2. For example, by saying "there's a cubicle on your right hand side and then straight in front of you there's also a door?", the director required the searcher to confirm a spatial setting which only matches with the room in the upper left of the setup shown in Figure 1. Table 2 shows that the complex sentence can be divided into two sentences: "there's a cubicle on your right hand side" and "there is also a door straight in front of you" respectively.

Table 2. Some examples of different kinds of utterances of *SpatialLocating (SL)*

	Examples	L	R	SM	%
SL	what do you see immediately to your left?	-	per	LeftProEx	19.5
	to the left of the filing cabinets what do you see?	-	MO	LeftProEx	
	you were in the little cubicle, right?	per	BP	Contain	33.8
	is there a door anywhere near you?	FO	per	Proximal	16.9
	there's a filing cabinet in front of you, right?	MO	per	FrontProEx	
	is there like a desk in front of the computer chair?	MO	MO	FrontProEx	6.5
SL-CjSL	there's a cubicle on your right hand side and	BP	per	RightProEx	35.1
	then straight in front of you there's also a door?	FO	per	FrontProEx	
SL-DjSL	are you in a hallway or	per	BP	Contain	
	(are you in a) small room?	per	BP	Contain	

4 Conclusions

In this paper, we analyzed location descriptions and location queries in the CreST corpus which naturally includes different forms of localization utterances in a dialogue context. First, to analyze the probable locatum and relatum, we extracted localization utterances which are more frequent in the corpus by classifying the environment into places (BP), fixtures (FO) and objects (MO). Second, we extracted descriptions of each group in order to explore the levels of descriptions attached with locatum and relatum. Finally, we paired these utterances with semantic categories provided by the GUM ontology to explore the specific semantic configurations and linguistic forms employed in localization query. The findings include (1) the adjectives and clauses are used to describe the locatum and relatum, (2) the entities of the locatum and relatum are often found unknown or variable, (3) the spatial relations such as Containment and Projectional are found frequently used, and (4) the query which describes two or more spatial locations are used to uniquely identify a spatial setting in the environment. Since linguistic semantics of localization expressions formalized in GUM is domain-independent, the semantic classifications and findings of this research offer a general conceptualization for relating place-sensitive natural expressions to their spatial semantic interpretation and therefore, provide a sophisticated foundation for the contextualization and generation of place-aware natural expressions in situated dialogue systems, like DAISIE (see Ross & Bateman [9]).

References

1. A. H. Anderson, M. Bader, E. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert. The HCRC Map task Corpus. *Language and Speech*, 34(4):351–366, 1991.
2. J. A. Bateman, J. Hois, R. J. Ross, and T. Tenbrink. A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174(14):1027–1071, September 2010.
3. H. Cuayáhuitl, N. Dethlefs, K.-F. Richter, T. Tenbrink, and J. Bateman. A dialogue system for indoor wayfinding using text-based natural language. In *Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2010): posters and short presentations*, Iasi, Romania, 2010. March 21-27.
4. K. Eberhard, H. Nicholson, S. Kbler, S. Gunderson, and M. Scheutz. The indiana "cooperative remote search task"(crest) corpus. In *Proceedings of the Seventh Language Resources and Evaluation Conference (LREC 10)*, 2010.
5. P. A. Heeman and J. Allen. The Trains 93 Dialogues. Trains Technical Note 94-2, Computer Science Dept., University of Rochester, mar 1995.
6. G.-J. M. Kruijff, H. Zender, P. Jensfelt, and H. I. Christensen. Clarification dialogues in human-augmented mapping. In *Proceedings of the 1st Annual Conference on Human-Robot Interaction (HRI'06)*, Salt Lake City, UT, March 2006.
7. S. Lauria, G. Bugmann, T. Kyriacou, J. Bos, and E. Klein. Training Personal Robots via Natural-Language Instructions. *IEEE Intelligent Systems*, pages 38–45, 2001.
8. I. Mani, C. Doran, D. Harris, J. Hitzeman, R. Quimby, J. Richer, B. Wellner, S. Mardis, and S. Clancy. SpatialML: annotation scheme, resources, and evaluation. *Language Resources and Evaluation*, 44(3):263–280, 2010. Special Issue LREC 2008: Selected papers, Edited by: N. Ide and N. Calzolari.
9. R. J. Ross and J. A. Bateman. Daisie: Information state dialogues for situated systems. In V. Matoušek and P. Mautner, editors, *Text, Speech and Dialogue*, volume 5729 of *Lecture Notes in Computer Science*, pages 379–386. Springer, Berlin / Heidelberg, 2009. 12th International Conference, TSD 2009, Pilsen, Czech Republic, September 13-17, 2009. Proceedings.

Representing Vague Places : Determining a Suitable Method

Mohammed Imaduddin Humayun and Angela Schwering

Institute for Geoinformatics, University of Muenster, Germany
humayun@uni-muenster.de, schwering@uni-muenster.de

Abstract. The representation of places with vague or ill-defined boundaries continues being an issue for information systems. Despite the presence of multiple representation methods, it is still unclear how to determine which approach is best suited for a particular task. This paper proposes a set of characteristics based on the application domain, conceptual, and logical levels and differentiates the approaches according to these characteristics. We demonstrate how they are matched with the task requirements and influence the choice of representation method.

1 Introduction

Representing 'place' poses challenges, more so when the extents or boundaries cannot be well-defined. Although humans are capable of interpreting what is being referenced in such cases, handling these in information systems is more complex. Some of the associated problems are discussed in literature under discipline of *spatial vagueness* and *uncertainty*, especially their philosophical and representational aspects. The former aspects address whether vagueness is intrinsic to the real world or just a feature of language[1], the different kinds of vagueness [2] and how to handle imperfection in geographic information [3]. The latter suggest models and theories to handle spatial vagueness, each with its distinct assumptions and properties. This has resulted in the development of various representation methods such as *probabilistic* [4,5], *fuzzy-set based* [6,7], *egg-yolk* model [8], *rough-sets* [9], and *supervaluation*[10], among others. We propose a methodology to distinguish between different representation methods based on their characteristics, which may then be matched with the application requirements in order to determine a suitable method.

No single representation can claim to be applicable for all cases. The methods differ in the way assumptions are made about space, the underlying formal models, applicability of data models and the kinds of reasoning they allow. Selecting the right one for a given task is a matter of fitness for purpose and requires that the method's capabilities are matched to the requirements. Requirements vary and can be specified in numerous ways. A consistent way of specifying these requirements is needed. Our approach is to specify these in terms of the model characteristics. The characteristics themselves may be defined at different levels similar to the levels of data abstraction in an information system [11].

Vasardani M., Winter S., Richter K-F., Janowicz K. and Mackaness W. (eds.); P-KAR 2012, pp. 19-25

1. At the *application domain* level, a subset of the reality to be represented is chosen with respect to a particular domain. We also specify what kind of reasoning is to be performed on a representation. It is also important to decide here whether vagueness is perceived to be intrinsic to the entity, or if different possible interpretations should be supported.
2. The *conceptual* level is the next and deals with how the vague place is conceptualized in an implementation independent fashion. Important concerns here are, how the vague referent can be *individuated*? (e.g. through use of objectifiable parameters), how it is *demarcated*? and its *identity* (e.g. temporal changes).
3. The *logical* level is the next and deals with more detailed specifics such as the data model of the data sources, or how the extents should be modelled.

In this paper, we identify a criteria set to determine suitable representation methods for vague places. Section 2 analyzes the requirements of an application task. Based on these requirements, section 3 develops a criteria set and differentiates vague representations in terms of these criteria. Section 4 gives an example how to choose the right representation method for a given task based on using our criteria set.

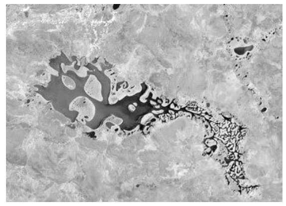
2 Analyzing Requirements for a Use Case

Lake Carnegie in Australia is ephemeral. Depending on the amount of precipitation the lake may or may not be filled with water (Fig. 1). Though a lake in vernacular terms, in dry seasons it is reduced to a muddy marsh[1]. This presents a problem, since it is now unclear where exactly the boundaries of the lake lie.

Suppose a user needs a representation of Lake Carnegie. We examine a few questions that need to be answered to arrive at a clear understanding of what needs to be represented.

1. *What is a 'lake'?* First, the semantics of the term 'lake' need to be clear. Is it a single contiguous body of water or does it include smaller scattered pools in the vicinity as well? Do the requirements dictate that water be present in the lake all year round? This is treated as the first step towards arriving at a solution.
2. *What is the purpose of representation?* Requirements for an ecologist differ from that of a cartographer. An ecologist is likely to be interested in the variation of the lake over time; a fuzzy spatial extent rather than precisely defined boundaries being of importance. A cartographer is more interested in the lake as a crisp object.
3. *What data sources are available?* The choice of representation is influenced by the data sources. A different method is needed for a representation built up from satellite imagery, than another which uses water level observations from sensors scattered through the lake.

[1] `http://www.nasa.gov/multimedia/imagegallery/image_feature_817.html`



(a) Apr 2011

(b) Sep 2011

Fig. 1: Landsat images of Lake Carnegie, Australia
(*courtesy: U.S. Geological Survey*)

As one can infer, varying user needs and requirements must be met with a suitable method to represent the same place. From the myriad of possibilities, one needs a way to identify the right representation approach. Next, we briefly propose certain characteristics and explain how these characteristics can be used to distinguish among representation methods as the first step to this end.

3 Methods of Representation

From the different levels of abstraction, we identify characteristics which will serve as the criteria to differentiate between methods. Some commonly used methods for representing vague places are then briefly analyzed based on these.

3.1 Criteria for Differentiation

Starting from the different levels of abstraction, we propose the following characteristics for use in deciding upon the correct method to employ for representation of vague entities.

1. *Conceptualization of space* - The adopted perspective of vagueness (whether it is linguistic or ontic) affects the choice of the representational and semantic framework [12]. How the phenomenon is treated by the method forms a useful basis for differentiation. This also has implications on the kind of boundary of the phenomenon (crisp, graduated, indeterminate etc.).
2. *Formal model* - This differentiates between the methods based on whether the underlying model is *stochastic, fuzzy set based, three-valued logic* or other.
3. *Data model* - Certain methods handle only regions (egg-yolk and supervaluation) whereas others are well suited for points or grid based data structures (fuzzy sets for instance). Since sources of data differ according to the data model they use (raster versus vector data), it is important to consider how a method behaves with respect to it. This also has implications on the kinds of boundaries that can be defined, e.g. how is a crisp boundary generated in the *raster* data model?

4. *Reasoning* - This characteristic determines what kinds of reasoning can be performed with the representation methods. Reasoning covers metric, directional and topological operations performed on vague places. This is particularly important from the perspective of a task, since it limits what kind of analysis can be done on the vague place. Some representations provide a well-defined framework for reasoning, whereas others do not.

3.2 Analysis of Representation Methods

Base representations - We coin this term to refer to those methods which abstract or crisp the vague place. Possible ways are to define the feature *a priori* according to some metric, or reduce it to a simple feature type (point, line or polygon), a minimum bounding rectangle (MBR) which covers the entire extent of the space where the entity is located, or through tessellations of space. These are usually in the form of vector data. Examples may be seen in VGI where a real world feature is outlined by contributors (from GPS tracks or tracing from aerial imagery), or in gazetteers where a feature is simply located by a representative point. Here vagueness is not preserved, and they are generally not classified under methods for vagueness representation. They are included here for the sake of completeness since they are often applied and prove adequate in some cases. The methods themselves do not provide any theory for reasoning.

Probabilistic methods - These methods derive the membership value of an individual in a set through a statistically defined probability function. These are used mainly to handle *uncertainty.* The underlying stochastic model assumes that phenomena are crisp and knowable, with the result that no measurable way for metrics such as *precision* in the case of vagueness exist [13]. These methods are best suited for phenomena with measurable objective properties such as flow, temperature, or water level. Probabilistic interpretations have also been employed to determine where city centres lie, based on probability of sample points computed from trials using participant studies [5]. These are generally suited for point or field based data and allow for a variety of statistical reasoning techniques to be performed.

Fuzzy-set methods - These are based on Fuzzy-set theory and ideal for modelling objects which have graduating or indeterminate boundaries [6,7,14]. The membership value α ($0 \leq \alpha \leq 1$), of a point in the region is highest at the *core* of the region and decreases gradually as the *boundary* is approached. Determination of membership value itself is subjective and may not relate directly to the phenomenon itself. The model also allows for obtaining a crisp boundary by means of α-cuts which are a way of obtaining crisp sets from a fuzzy set. This method is applicable in the case of both raster and vector data, where a feature or cell may be assigned a membership. Reasoning using fuzzy set operations such as intersection, union, complement etc. can also be performed [6].

Egg-yolk method - A vague region is considered analogous to an egg – the *yolk* corresponds to the minimal extent, the *white* being the indeterminate region and its maximal extent. Any acceptable crisping must lie between these inner and outer subregions. This method allows performing qualitative spatial reasoning between vague regions or between a vague region and a crisp region under the framework of the regional connection calculus (RCC-5) [8]. Reasoning is possible on different possible configurations of two regions represented this way. However, the theory itself does not make any assertion as to how the crisp regions are obtained. Egg-yolk models are ideal when topological reasoning on vector data in the form of regions is to be performed.

Rough sets - The basis for rough sets is the *indiscernibility relation* – where a collection of elements is indiscernible from another. Rough sets use a three-valued logic ($true, false, maybe$) to determine the membership of a point to a region as opposed to the binary notion of membership ($true, false$) in classical set theory. Similar to the egg-yolk model, a region may be represented by its determinate *lower approximation* and an indeterminate *upper approximation* [9,3]. Rough sets are ideal for reasoning on multi-resolution raster data, where a change in resolution results in indiscernibility.

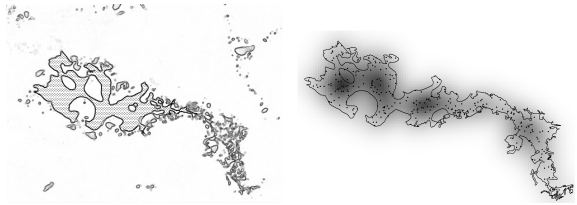
Supervaluation - The idea behind *supervaluation* is to account for the different possible interpretations of a vague predicate when multiple interpretations for a vague region exist. The *positive extension* is where all interpretations are true. Its inverse the *negative extension* is the region where no interpretation is true. The remaining regions constitute the *penumbra* [10]. Supervaluation enables use of classical logic to reason about vagueness, but computational applications are hampered by the fact that all admissible interpretations must be explicitly specified, which is difficult in practice [12]. These are applicable in data models where regions are primitives and allow for reasoning on vague regions where several boundaries may be associated with an object.

4 Determining a Suitable Representation - An Example

We take the example of Lake Carnegie and consider two different sets of requirements for representations.

- The cartographer imagines the lake to be a single contiguous body of water with a crisp boundary, though the reality is different. Satellite imagery is used as source of data. No reasoning needs to be performed.
- The ecologist views the lake as a non-crisp object defined by level of water. Water level observation data from sensors is available. The need is to generate a surface exhibiting water presence over a period of time.

In the first case, the space is conceptualized as a crisp body. Data from satellite imagery is a raster, from which the boundary needs to be derived. One obvious solution is to simply trace the outline from the image, resulting in a base



(a) (b)

Fig. 2: (a) Fuzzy set and (b) probabilistic representations of Lake Carnegie

representation. The suggested approach in this case is however to use fuzzy-set modelling with α-cuts. By varying α, different boundaries (each of which is crisp) can be obtained. One such representation by using fuzzy membership of pixel values and obtaining a crisp boundary is seen in Fig. 2a.

In the second case, the lake is conceptualized depending on its objective property (water level). Spatial distribution of the data source, sensors which provide observations, can be thought of as consisting of points (vector). Since the user needs to obtain interpolated values in order to obtain a lake surface, application of probabilistic methods is suitable here. A probabilistic representation simulated from randomly distributed sensors with arbitrary observations is shown in Fig. 2b, with outline of the lake from OpenStreetMap [2] for reference.

This is a trivial example, but the same principles apply in other cases as well. For example, in the Tell Us Where[3] project dataset, it would be possible to obtain a representation of places with noncrisp boundaries. This however has not been attempted here owing to sampled locations in the current dataset being insufficient in number to demonstrate our cause.

5 Conclusion

Various representation methods have been proposed in literature for the representation of places with vague boundaries. It is important to enable decision makers to adopt the right method based on their needs. The approach taken here uses multiple levels of abstraction to specify the requirements in a consistent manner. The levels allow identification of characteristics with which different methods can be analyzed. Differing requirements in the modelling of a vague region such as a lake can lead to different possible solutions as presented in the lake use case.

6 Acknowledgements

This research was carried out under the International Research Training Group on Semantic Integration of Geospatial Information (IRTG-SIGI) and is funded by the DFG (German Research Foundation), GRK 1498.

[2] http://www.openstreetmap.org
[3] http://telluswhere.net

References

1. Varzi, A.C.: Vagueness in geography. Philosophy & Geography **4**(1) (2001) 49–65
2. Bennett, B.: Modes of concept definition and varieties of vagueness. Applied Ontology **1**(1) (2005) 17–26
3. Duckham, M., Mason, K., Stell, J., Worboys, M.: A formal approach to imperfection in geographic information. Computers, Environment and Urban Systems **25**(1) (2001) 89 – 103
4. Leung, Y., Yan, J.: A locational error model for spatial features. International Journal of Geographical Information Science **12**(6) (1998) 607–620
5. Montello, D.R., Goodchild, M.F., Gottsegen, J., Fohl, P.: Where's downtown?: Behavioral methods for determining referents of vague spatial queries. Spatial Cognition & Computation **3**(2) (2003) 185–204
6. Usery, E.L.: A conceptual framework and fuzzy set implementation for geographic features. In Burrough, P.A., Frank, A.U., eds.: Geographic Objects with Indeterminate Boundaries. Taylor and Francis, London (1996)
7. Wang, F., Hall, G.B.: Fuzzy representation of geographical boundaries in gis. International journal of geographical information systems **10**(5) (1996) 573–590
8. Cohn, A., Gotts, N.: The 'egg-yolk' representation of regions with indeterminate boundaries. Geographic Objects With Indeterminate Boundaries **2** (1996) 171–187
9. Worboys, M.: Imprecision in finite resolution spatial data. GeoInformatica **2**(3) (1998) 257–279
10. Kulik, L.: A geometric theory of vague boundaries based on supervaluation. In Montello, D., ed.: Spatial Information Theory. Volume 2205 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2001) 44–59
11. Worboys, M., Duckham, M.: GIS: A Computing Perspective. 2nd edn. CRC Press (2004)
12. Bennett, B.: Spatial vagueness. In Jeansoulin, R., Papini, O., Prade, H., Schockaert, S., eds.: Methods for Handling Imperfect Spatial Information. Springer (2011) 15–47
13. Duckham, M., Sharp, J.: Uncertainty and geographic information: Computational and critical convergence. In: Re-presenting GIS. John Wiley & Sons Ltd. (2005) 113–124
14. Fisher, P., Wood, J., Cheng, T.: Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. Transactions of the Institute of British Geographers **29**(1) (2004) 106–128

Intuitive and Natural Interfaces for Geospatial Data Classification

Falko Schmid[1], Oliver Kutz[2], Lutz Frommberger[3], Till Mossakowski[4], Tomi Kauppinen[5], and Chunyuan Cai[6]

[1] Cognitive Systems, University of Bremen, Germany, schmid@informatik.uni-bremen.de
[2] Research Center Spatial Cognition (SFB/TR 8), University of Bremen, Germany, okutz@informatik.uni-bremen.de
[3] Capacity Lab, University of Bremen, Germany, lutz@capacitylab.org
[4] DFKI Lab Bremen, Germany, Till.Mossakowski@dfki.de
[5] MUSIL, University of Muenster, Germany, tomi.kauppinen@uni-muenster.de
[6] GeoCrowd, University of Bremen, Germany, cai@informatik.uni-bremen.de

Abstract. Increasing availability of GPS-enabled devices technically enables a broad variety of people to participate in the volunteered geographic information (VGI) movement and to collect and share information about places and spatial entities. But in order to be useful, geo-data has to be correctly classified, and inexperienced users need assistance to be able to provide correctly classified information, because the classification system is complex and not always intuitive. In this paper, we propose a natural classification approach for spatial entities based on speech recognition and ontological reasoning to allow users to contribute geo-data with as little barriers as possible.

1 Data for Everyone

In the last decade, volunteered and participatory initiatives to create repositories of geo-spatial information gained overwhelming success. The most prominent and successful example of *volunteered geographic information* (VGI) is OpenStreetMap[1] (OSM). OSM offers the opportunity to collect data where no commercial data sets are available for lack of (commercial) interest, such as for example rural areas of developing countries.

The great advantage of OSM data is the collection and provision by interested users. This method supports the collection not only of rather traditional data such as streets, buildings, or natural features. OSM contains a large variety of particular data like, e.g., barriers or surface properties, thus providing information essential for creating assistance for, e.g., disabled persons or athletes. This is a great advantage compared to official data sets: OSM contributors collect and share the information relevant to them and other users with similar interests. Such possibilities add enormous value to the freely available data, as it does not only map the street network, but potentially every spatial asset and facet of a place which is of interest to someone.

[1] www.openstreetmap.org

Vasardani M., Winter S., Richter K-F., Janowicz K. and Mackaness W. (eds.); P-KAR 2012, pp. 26-32

2 Interfaces for Everyone

To enable systems to make correct use of the collected data, it has to be classified correctly. For example, cartographic renderers can only draw and label objects with correct style if the entities follow a certain specification. The classification of geo-data is complex and often ambiguous. For example, the type of a street or the function of some grass covered ground may remain unclear to the contributor. Trained contributors know how to apply a classification system correctly; for non-experts or casual contributors, the lack of this knowledge marks a barrier: most of the tools to collect, contribute, and classify geo-spatial data are complex systems requiring high technical affinity and skills. Moreover, even for experts, repeated classification of objects can become tiresome, leading to the danger of incompletely specified data.

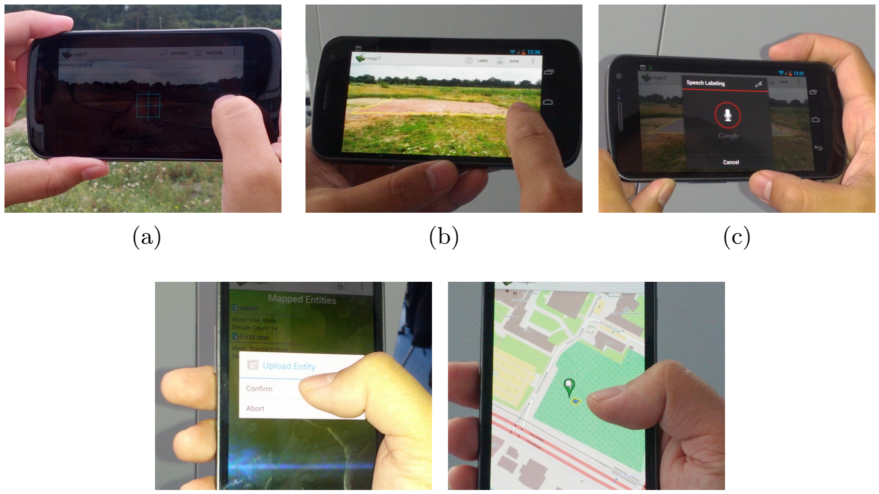
Places have different facets for different people. Namely, the same place can have very different functional roles depending on who is looking at it [9]. For example, the entrance area of our Bremen office building is frequently used by skateboarders in the late afternoons. So what is an entrance for the people working there is an urban sports facility for others. Thus, the place can be classified differently depending on the reporter. But, at a certain level of abstraction, all views on the place will be the same; in the end, the entrance area is a paved spot. Another example is a fish pond: for some, it is just a recreational decoration, for others a food supply; but in any case it is a (artificial) water body and in OSM terms "water". In this paper, we focus on the latter: a natural classification system for VGI applications that allows the collection of geo-data for untrained contributors.

3 MapIt: Intuitive and Natural Interfaces for VGIying

Research on VGI and Human-Computer Interaction (HCI) is increasingly addressing the technological gap between potential contributors and the existing data collection applications (e.g., [3]). The MapIt system [8] offers an intuitive interface for collecting spatial entities and is targeted at casual contributors with low technical affinity. It only requires basic smartphone usage knowledge: the user just has to make a photo, outline the entity on a map, classify it using natural language, and finally upload it to a server (see Fig. 1).

3.1 Ontological reasoning for spatial classification

When we allow users to annotate spatial entities by means of natural language rather than by using a predefined catalogue, we have to expect a significant mismatch between what users think the entity is and what the classification system allows to describe. In [9], the authors demonstrated that natural descriptions of the same places are highly heterogeneous between individual users. To solve the mismatch between natural expressions and a catalogue based classification, we



(a) (b) (c)

(d) (e)

Fig. 1: The mapping process: Taking a photo (a), outlining the entity (b), annotating via speech (c), uploading to a geo-server (d), checking the entity on map (e).

propose an ontological reasoning system to identify the best matching classifier for an entity.

Consider the following situation: some member of a development project wants to contribute data about the distribution of small backyard fish ponds which have been installed to minimize the lack of protein supply in poor areas of developing countries. This user is not educated to use a geographic classification system and is not aware of the proper term within a system like CityGML[2], OSM, ATKIS[3], or the OS MasterMap[4].

If the user now labels the backyard fish ponds with the term "fish pond", none of the above mentioned systems will recognize it as a valid entity. Without a proper classification, however, the data remains useless as it cannot be rendered or addressed by other algorithms.

To be able to match natural concepts of spatial entities with spatial classification systems, we propose a reasoning system as illustrated in Figure 2. The goal of the proposed reasoner is to identify the closest conceptual match in the classification system with the naturally spoken term. The term should not just be replaced, but the link between the spoken term and the linked term in the classification system is kept for further refinement of both the classification system

[2] http://www.citygml.org/
[3] http://www.adv-online.de
[4] http://www.ordnancesurvey.co.uk/oswebsite/products/os-mastermap/index.html

and the reasoner's capabilities. A main ingredient to make this re-classification possible is an abstraction layer on top of existing GIS classifications, namely the meta-ontology GeoMO sketched in the next section.

3.2 The meta-ontology GeoMO and the OntoHub repository

OntoHub. Existing ontology repositories such as BioPortal[5] lack the ability to host heterogeneous ontologies in the sense of being formulated in ontology languages other than $\mathcal{OWL}$. As not all relevant ontologies will be $\mathcal{OWL}$ ontologies (Dolce, e.g., is formulated in first-order logic) we host our ontologies at OntoHub[6]. Users of OntoHub can upload, browse, search and annotate basic ontologies written in various standard ontology languages via a web frontend (see [7] for more information on OntoHub). Beyond basic ontologies, OntoHub supports linking ontologies across ontology languages, and creating distributed ontologies as sets of basic ontologies and links among them. An important difference to the mapping facilities of, e.g., BioPortal is that links in OntoHub have formal semantics, and therefore enable new reasoning and interoperability scenarios between ontologies, features that are essential for the automated classification scenario described in this paper.

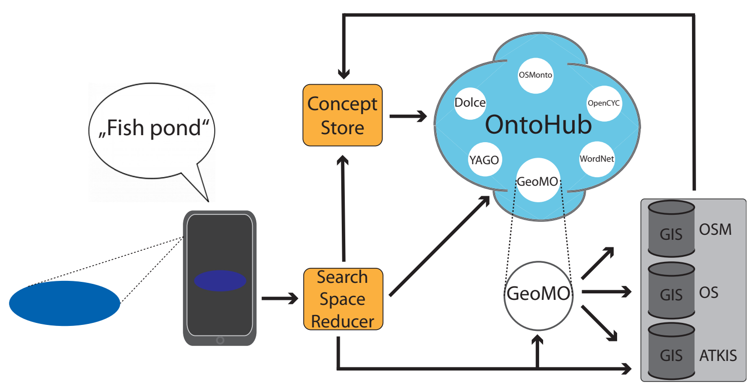


Fig. 2: Conceptual overview of the reasoning architecture of MapIt.

GeoMO. The role of the meta-ontology GeoMO is twofold: first, the *mediation* between human everyday concepts of space and spatial entities that should be matched against existing geo-spatial classifications, and secondly, to *translate* between different classification systems such as OSM, ATKIS, OS MasterMap, CityGML, etc. For OSM, we have already designed OSMOnto, an automatically

[5] See http://bioportal.bioontology.org/
[6] See http://ontohub.org/

generated ontology of OSM tags [2, 1].[7] In contrast to GeoMO, OntoHub is a collection of different ontologies with GeoMO being a part of it. The role of OntoHub is the provision of different sources of concepts of different domains and relations between them. We propose to use DBPedia[8], OpenCYC[9], YAGO [11], Dolce [4] and WordNet[10] as ontologies to mediate between everyday concepts and classification systems. DBPedia is an ontology extracted from Wikipedia entries, OpenCYC a collection of commonsense knowledge, whilst WordNet provides, e.g., *synsets*, i.e. sets of terms that are considered synonymous in natural language.

The GeoMO ontology, on a technical level, results from a colimit operation on the ontologies reflecting the classification systems of the participating GISs (we mentioned the OSMOnto ontology above, being one component), together with knowledge (i.e. term mapping, subsumptions between terms, etc.) about their relationship. Such mappings are part of the OntoHub infrastructure.

Here is a simple example illustrating the functionality of GeoMO. The OS MasterMap might contain the category s (i.e. 'water structure — manmade'), whilst OSM might use the term t (i.e. 'water body'). GeoMO establishes the subsumption $s \sqsubseteq t$, i.e. the term t is *more general* than s. If the user now expresses the term 'Fish pond' with spoken, natural language, the term is translated by available speech recognition into a proccessable term. The Concept Store uses this term for a lookup in WordNet and identifies the synonym w. Moreover, OpenCYC will tell us that this synonym w is in fact a special case of s, an official category in the OS MasterMap classification scheme. Finally, GeoMO can infer that t can be used as a more general category for labeling 'Fish pond', without any user interaction.

3.3 A sketch of the Architecture of MapIt

The reasoner depicted in Figure 2 will work as follows: the smartphone translates the spoken term "fish pond" via a standard speech recognition module into parsable text. The detected term "fish pond" is then send to the Search Space Reducer (SSR). The function of the SSR is to cut down the search space in a context-sensitive way: as we are in a geographic domain, we only want to query ontologies or parts of ontologies dealing with spatial objects and activities related to them. This situation allows the SSR to ignore a significant amount of entries, like facts about artists, movies, books, vehicles, etc.

After checking for the existence of the term in the target classification (in this case OSM) and GeoMO. If both do not contain a direct correspondence, the reasoner looks up the *Concept Store*. A core component of the *Concept Store* is illustrated in Fig. 3. It illustrates the implementation of a workflow, previously developed in [6], for aligning sets of ontologies and checking for consistency of

[7] See also http://wiki.openstreetmap.org/wiki/OSMonto
[8] www.dbpedia.org
[9] www.opencyc.org
[10] www.wordnet.princeton.edu

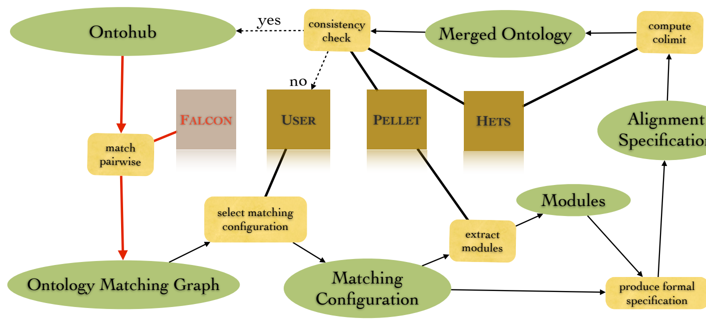


Fig. 3: A basic workflow of the concept store.

their combination. This workflow is in particular essential for the construction of GeoMO, as the compatibility of mappings between the terms used in the various GIS ontologies has to be verified. We here briefly introduce these tools.

The ontologies to be matched and aligned are taken from OntoHub. As matching system we use Falcon [5] which matches $\mathcal{OWL}$ ontologies by means of linguistic and structural analysis. For module extraction as well as consistency checks we use Pellet [10] which in particular makes use of the $\mathcal{OWL}$-API [11]. Finally, we use Hets[12] for the computation of colimits (i.e. 'realized' alignments).

4 Conclusion and Outlook

The MapIt architecture carefully integrates existing ontologies and reasoning systems and aims at an enhanced classification technology for geo-data. We expect that MapIt, once realized as a system, has the potential to lower the barrier of contribution of VGI tag data to OpenStreetMap or any other geographic classification catalogues. Currently, tagging in OpenStreetMap mostly happens at geographical level, and much less at a higher ontological level, e.g., concerning activities or individual perception, or place usage of users. This situation could greatly improve using MapIt.

Acknowledgements

Support by the German Research Foundation (DFG) is gratefully acknowledged.

References

1. Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., and Rau, R. DO-ROAM: Activity-Oriented Search and Navigation with OpenStreetMap. In

[11] See http://owlapi.sourceforge.net
[12] See www.informatik.uni-bremen.de/cofi/hets

Fourth International Conference on GeoSpatial Semantics (GeoS-11) (2011), M. B. C. Claramunt, S. Levashkin, Ed., vol. 6631 of *Lecture Notes in Computer Science*, Springer, pp. 88—107.

2. Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., and Rau, R. OSMonto—An Ontology of OpenStreetMap Tags. In *In State of the map Europe (SOTM-EU)* (2011).
3. Frommberger, L., Schmid, F., Cai, C., Freksa, C., and Haddawy, P. Barrier-free mapping for development and poverty reduction. In *Role of Volunteered Geographic Information in Advancing Science: Quality and Credibility, Workshop at GIScience 2012* (to appear), B. Bhaduri, R. Devillers, M. Haklay, and M. Goodchild, Eds.
4. Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A. Sweetening WordNet with DOLCE. *AI Magazine 24*, 3 (2003), 13–24.
5. Hu, W., and Qu, Y. Falcon-AO: A practical ontology matching system. In *Proc. of WWW-07* (2008), pp. 237–239.
6. Kutz, O., Normann, I., Mossakowski, T., and Walther, D. Chinese Whispers and Connected Alignments. In *Proc. of the 5th International Workshop on Ontology Matching (OM-2010)* (9th International Semantic Web Conference ISWC-2010, November 7, 2010, Shanghai, China., 2010).
7. Lange, C., Mossakowski, T., Kutz, O., Galinski, C., Grüninger, M., and Vale, D. C. The Distributed Ontology Language (DOL): Use Cases, Syntax, and Extensibility. In *Proc. of the 10th Terminology and Knowledge Engineering Conference (TKE 2012)* (Madrid, Spain, 2012).
8. Schmid, F., Cai, C., and Frommberger, L. A new micro-mapping method for rapid VGI-ing of small geographic features. In *Geographic Information Science: 7th International Conference (GIScience 2012)* (to appear).
9. Schmid, F., and Kuntzsch, C. In-situ communication and labeling of places. In *Proceedings of the 6th International Symposium on LBS and TeleCartography* (2009), M. Jackson, S. Anand, and G. Gartner, Eds., University of Nottingham.
10. Sirin, E., Parsia, B., Cuenca Grau, B., Kalyanpur, A., and Katz, Y. Pellet: A practical OWL DL reasoner. *Journal of Web Semantics 5*, 2 (2007), 51–53.
11. Suchanek, F. M., Kasneci, G., and Weikum, G. YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. In *Proceedings of the 16th international Conference on World Wide Web (WWW-07)* (New York, NY, USA, 2007), ACM, pp. 697–706.

Conversational Natural Language interaction for Place-related Knowledge Acquisition

Srinivasan Janarthanam[1], Oliver Lemon[1], Xingkun Liu[1], Phil Bartie[2], William Mackaness[2], Tiphaine Dalmas[3] and Jana Goetze[4]

[1]Interaction Lab, Heriot-Watt University, Edinburgh
[2] School of GeoSciences, University of Edinburgh
[3]School of Informatics, University of Edinburgh
[4]KTH Royal Institute of Technology, Stockholm, Sweden
sc445,o.lemon,x.liu@hw.ac.uk, philbartie@gmail.com,
william.mackaness@ed.ac.uk, tiphaine.dalmas@aethys.com, jagoetze@kth.se

Abstract. We focus on the problems of using Natural Language interaction to support pedestrians in their place-related knowledge acquisition. Our case study for this discussion is a smartphone-based Natural Language interface that allows users to acquire spatial and cultural knowledge of a city. The framework consists of a spoken dialogue-based information system and a smartphone client. The system is novel in combining geographic information system (GIS) modules such as a visibility engine with a question-answering (QA) system. Users can use the smartphone client to engage in a variety of interleaved conversations such as navigating from A to B, using the QA functionality to learn more about points of interest (PoI) nearby, and searching for amenities and tourist attractions. This system explores a variety of research questions involving Natural Language interaction for acquisition of knowledge about space and place.

1 Motivation

Although spatial information is available to users through graphical/visual smartphone apps, there are still important problems such as how such information is delivered safely and proactively, and without cognitively overloading the user. [1] suggested that cognitive load of information presented in textual and speech-based interfaces is medium and low respectively when compared to more complicated visual interfaces. An interesting research objective, therefore, is to build a hands-free and eyes-free system that engages pedestrian users by presenting all information and receiving user requests through Natural Language.

In addition, and in contrast to other mobile applications, we focus on *conversational* interaction – meaning that the system accumulates information over time, and that the interface plans its utterances to achieve long-term communicative goals. Note that recent speech interfaces such as *Siri* use Natural Language input but they are not truly conversational because they mostly handle single user inputs one at a time, without maintaining and updating models of the dialogue context.

Vasardani M., Winter S., Richter K-F., Janowicz K. and Mackaness W. (eds.); P-KAR 2012, pp. 33-38

Our Natural Language interface integrates with a city model and a visibility engine [2] to identify points of interest and visible landmarks for presentation, a pedestrian tracker to improve the GPS positioning of the user, and a question-answering (QA) system to enable users to explore information about the city more freely than with a graphical interface.

Table 1 presents an example dialogue interaction with the interface showing the use of visibility information and Question-Answering.

User: I want to go to a museum.
System: How about the National Museum of Scotland? It is a five star rated attraction.
User: Okay.
System: Okay. Directions to the National Museum of Scotland.
System: Turn right on to South Bridge and walk towards the tower in front of you.
...
User: I am looking for an Indian restaurant.
System: Okay. There is an Indian restaurant named Kushis Diner on West Nicholson Street.
User: Okay. Take me there.
...
System: Near you is the famous statue of David Hume.
User: Tell me more about David Hume.
System: David Hume is a Scottish philosopher who ....

Table 1. An example interaction with the interface

2 Related work

There are several mobile apps such as *Triposo, Tripwolf*, and *Guidepal* that provide point of interest information, and apps such as *Google Navigation* that provide navigation instructions to users. However, they demand the user's visual attention because they predominantly present information on a small screen of a mobile device. In contrast, we are developing a speech-only interface in order to keep the user's cognitive load low and avoid users from being distracted (perhaps dangerously so) from their primary task.

Previously, generating navigation instructions in the real world for pedestrians has been an interesting research problem in both computational linguistics and geo-informatics [3, 4]. For example, *CORAL* is an NLG system that generates navigation instructions incrementally by keeping track of the user's location, but the user has to ask for the next instruction when he reaches a junction [3]. *DeepMap* is a system that interacts with the user to improve positioning [5]. It asks users whether they can see certain landmarks, and based on their answers improves the user's GPS position estimate. However, in many such current systems, interactions happen through the use of GUI elements such as drop-down lists and buttons, and not by using speech interaction. The *Edinburgh Augmented*

Reality System (EARS) was a prototype system that presented point of interest information to users based on visibility [6].

In contrast to these earlier systems our objective is to present navigational, point-of-interest and amenity information in an integrated way using Natural Language dialogue, with users interacting eyes-free and hands-free through a headset connected to a smartphone.

3 Architecture

The architecture of the current system is shown in figure 1. The server side consists of a dialogue interface (parser, interaction manager, and generator), a City Model, a Visibility Engine, a QA server and a Pedestrian tracker.

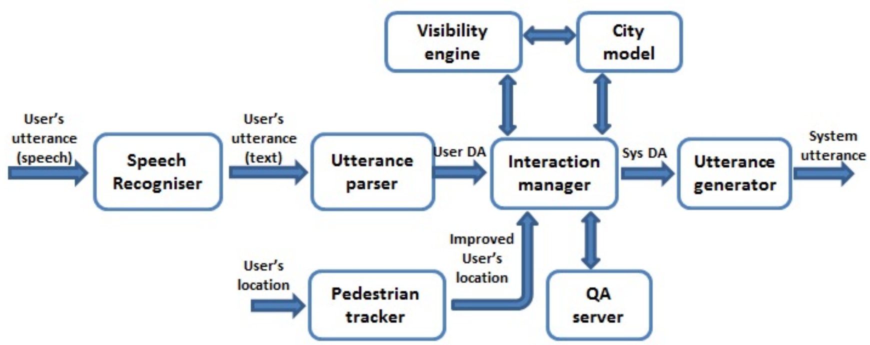


Fig. 1. System Architecture

3.1 Dialogue interface

The dialogue interface consists of a speech recogniser, an utterance parser, an Interaction Manager and an utterance generator. The speech recognition module recognises the user's utterance from the user's speech input. The utterance parser translates user utterances in to meaning representations called *dialogue acts*. The Interaction Manager is the central component of this architecture, which provides the user navigational instructions and interesting PoI information. It receives the user's input in the form of a dialogue act and the user's location in the form of latitude and longitude information. Based on these inputs and the dialogue context, it responds with system output dialogue act (DA), based on a dialogue policy. The utterance generator is a natural language generation module that translates the system DA into surface text, using the Open CCG toolkit [7].

3.2 Pedestrian tracker

Global Navigation Satellite Systems (GNSS) (e.g. GPS, GLONASS) provide a useful positioning solution with minimal user side setup costs, for location aware applications. However urban environments can be challenging with limited sky views, and hence limited line of sight to the satellites, in deep urban corridors. There is therefore significant uncertainty about the user's true location reported by GNSS sensors on smartphones [8]. This module improves on the reported user position by combining smartphone sensor data (e.g. accelerometer) with map matching techniques, to determine the most likely location of the pedestrian [2].

The output includes a robust street centreline location, and a candidate space showing the probability of the user's more exact position (e.g. pavement location). This module ensures that any GNSS-reported location placing the user at a rooftop location would be corrected to the most likely ground level location, taking into consideration user trajectory history and map matching techniques. User orientation is inferred from their trajectory.

3.3 City Model

The City Model is a spatial database containing information about thousands of entities in the city of Edinburgh. These data have been collected from a variety of existing resources such as Ordnance Survey, OpenStreetMap and the Gazetteer for Scotland. It includes the location, use class, name, street address, and where relevant other properties such as build date. The model also includes a pedestrian network (streets, pavements, tracks, steps, open spaces) which can be used to calculate minimal cost routes, such as the shortest path.

3.4 Visibility Engine

This module identifies the entities that are in the user's *vista space* [9]. To do this it accesses a *digital surface model*, sourced from LiDAR, which is a 2.5D representation of the city including buildings, vegetation, and land surface elevation. The visibility engine uses this dataset to offer a number of services, such as determining the line of sight from the observer to nominated points (e.g. which junctions are visible), and determining which entities within the city model are visible. A range of visual metrics are available to describe the visibility of entities, such as the field of view occupied, vertical extent visible, and the facade area in view. These metrics can be then used by the interaction manager to generate effective Natural Language navigation instructions. E.g. "Walk towards the castle", "Can you see the tower in front of you?", "Turn left after the large building on your left after the junction" and so on.

3.5 Question-Answering server

The QA server currently answers a range of Natural Language *definition* questions. E.g., "Tell me more about the Scottish Parliament", "Who was David

Hume?", etc. QA identifies the entity focused on in the question using machine-learning techniques [10], and then proceeds to a textual search on texts from the Gazetteer of Scotland and Wikipedia, and definitions from WordNet glosses. Candidates are reranked using a trained confidence score with the top candidate used as the final answer. These are usually long, descriptive answers and are provided in spoken output as a flow of sentence chunks that the user can interrupt. This information can also be offered by the system when a salient entity appears in the user's viewshed.

4 User interface

Users can interact with the system using a smartphone client that communicates with the system via the 3G network. The client is an Android app running on the user's mobile phone. It consists of two parts: the user's position tracker and the interaction module. The position tracker module senses user's position (latitude and longitude) and accelerometer readings. This information is sent to the system. The interaction module captures the user's speech input and relays it to the system. It also receives the system's utterances, which then is converted in to speech using the Android text-to-speech service.

We also built a web-based user interface to support the development of the system modules. It allows web-users to interact with our system from their desktops. It uses Google Street View to allow users to simulate pedestrian walking. An interaction panel lets the user interact with the system using Natural Language text or speech input. The system's utterances are synthesized using the Cereproc text-to-speech engine and presented to the user. For a detailed description of this component, please refer to [11]. A demonstration of this system will be presented at [12].

5 Future work

There are many remaining challenges in this research area for discussion, for instance:

- interleaving question-answering and navigation dialogue in a coherent manner;
- optimising the action selection of the dialogue interface (i.e. what to say next in the conversation), using machine learning techniques similar to [13–15];
- robustly handling the uncertainty generated by GPS sensors, speech recognition, and ambiguity of Natural Language interaction itself;
- generating useful referring expressions (e.g. the church on your left with the spire) which combine spatial and visual information;
- evaluating this system with real pedestrian users (this phase of the project is imminent).

Acknowledgments

The research has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 270019 (SPACEBOOK project http://www.spacebook-project.eu/).

References

1. Kray, C., Laakso, K., Elting, C., Coors, V.: Presenting route instructions on mobile devices. In: Proceedings of IUI 03, Florida. (2003)
2. Bartie, P., Mackaness, W.: D3.4 pedestrian position tracker. Technical report, The SPACEBOOK Project (FP7/2011-2014 grant agreement no. 270019) (2012)
3. Dale, R., Geldof, S., Prost, J.: CORAL : Using Natural Language Generation for Navigational Assistance. In: Proceedings of ACSC2003, South Australia. (2003)
4. Richter, K., Duckham, M.: Simplest instructions: Finding easy-to-describe routes for navigation. In: Proceedings of the 5th Intl. Conference on Geographic Information Science. (2008)
5. Malaka, R., Zipf, A.: Deep Map - challenging IT research in the framework of a tourist information system. In: Information and Communication Technologies in Tourism 2000, Springer (2000) 15–27
6. Bartie, P., Mackaness, W.: Development of a speech-based augmented reality system to support exploration of cityscape. Transactions in GIS **10** (2006) 63–86
7. White, M., Rajkumar, R., Martin, S.: Towards Broad Coverage Surface Realization with CCG. In: Proc. of the UCNLG+MT workshop. (2007)
8. Zandbergen, P.A., Barbeau, S.J.: Positional accuracy of assisted gps data from high-sensitivity gps-enabled mobile phones. Journal of Navigation **64(3)** (2011) 381–399
9. Montello, D.: Scale and multiple psychologies of space. In Frank, A.U., Campari, I., eds.: Spatial information theory: A theoretical basis for GIS. (1993)
10. Mikhailian, A., Dalmas, T., Pinchuk, R.: Learning foci for question answering over topic maps. In: Proceedings of ACL 2009. (2009)
11. Janarthanam, S., Lemon, O., Liu, X.: A web-based evaluation framework for spatial instruction-giving systems. In: Proc. of ACL 2012, South Korea. (2012)
12. Janarthanam, S., Lemon, O., Liu, X., Bartie, P., Mackaness, W., Dalmas, T., Goetze, J.: Integrating location, visibility, and question-answering in a spoken dialogue system for pedestrian city exploration. In: Proc. of SIGDIAL. (2012)
13. Janarthanam, S., Lemon, O.: Learning Adaptive Referring Expression Generation Policies for Spoken Dialogue Systems. In: Empirical Methods in Natural Language Generation. Springer (2010)
14. Janarthanam, S., Lemon, O.: Learning to adapt to unknown users: referring expression generation in spoken dialogue systems. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics (2010) 69–78
15. Rieser, V., Lemon, O.: Reinforcement Learning for Adaptive Dialogue Systems: a Data-driven Methodology for Dialogue Management and Natural Language Generation. Theory and Applications of Natural Language Processing. Springer (2011)

From Pattern Recognition to Place Identification

Sven Eberhardt, Tobias Kluth, Christoph Zetzsche, and Kerstin Schill

Cognitive Neuroinformatics, University of Bremen, Enrique-Schmidt-Straße 5, 28359 Bremen, Germany,
sven2@uni-bremen.de,
homepage: http://www.informatik.uni-bremen.de/cog_neuroinf/en

Abstract. What are the ingredients required for vision-based place recognition? Pattern recognition models for localization must fulfill invariance requirements different from those of object recognition. We propose a method to evaluate the suitability of existing image processing techniques by testing their outputs against these invariances. The method is applied to several holistic and one local model. We generalize our findings and identify model properties of locality, spatial configuration and generalization as key factors for applicability to localization tasks.

Keywords: visual, model, pattern recognition, localization

1 Introduction

Although the concept of place is essential to the way humans represent and interact with spatial environments, many of its determinants are not yet completely understood. One important question is what kind of information and what computations can be used to determine a specific place. Among the different types of input suitable for this purpose pictorial information has a particularly high potential. In biological terms, the investigation of place cells, for example, indicates the importance of visual cues for the robust localization of rodents.[1]

However, the exact processing mechanisms that can enable a successful vision-based localization are still unclear. In particular, it has to be understood how the classical determinants of pattern recognition systems, invariance and generalization properties, relate to the problem of localization. Invariance properties seem to play a crucial role, since for example the activation of a place cell is primarily determined by the animal's location, whereas it is independent of the orientation and other conditions like illumination. These are typical invariance properties. It may thus be assumed that the classic invariance principles attributed to human vision, and the corresponding computer vision approaches, can also be applied to the problem of localization (or place recognition). In this paper, we will argue that this is not necessarily the case, and that successful localization requires specific properties that can be in direct in opposition to those underlying other basic visual capabilities, like for example object recognition. For this, we will first introduce a basic framework that enables the description and differentiation of image processing techniques with respect to their applicability for localization

Vasardani M., Winter S., Richter K-F., Janowicz K. and Mackaness W. (eds.); P-KAR 2012, pp. 39-44

as compared to, e.g., object recognition. We will then discuss how some established image processing techniques can be described in terms of the suggested framework. This will then motivate an investigation of the suitability of some of these techniques for the specific problem of localization, or place recognition. In particular, we will investigate whether one of the most successful models of visual object recognition, the HMAX model[2], can also be used for the task of vision based localization.

1.1 Invariance in Place Recognition

One of the difficulties in place recognition from visual input is that even minor changes in observer's orientation or location, as well as unrelated changes such as variations in illumination, can cause vast changes of retinal input. Successful models for place identification should provide output that is invariant to such small changes in the observer's view. Although this is a requirement which is shared with object recognition models, there are some fundamental differences in which kind of invariance is desired.

While changes in scale, position and occlusion of elements in a scene are often irrelevant in the context of object recognition, they correspond to movement of the observer and should elicit changes in the output for place recognition models. On the other hand, spatial shifting of a scene as a whole corresponds to rotation of the observer. Similarly, rotations within the viewing plane correspond to tilting of the viewer's head. A place detector that mimics the behavior of place cells should be invariant to such rotations.

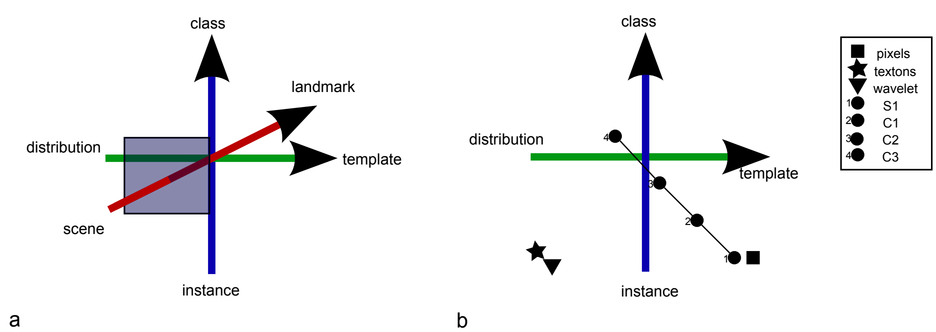


Fig. 1. a: Conceptual space for classifying pattern recognition models. b: Position of analyzed models in two dimension of our conceptual space.

Given these fundamental differences, can models for object recognition be used for place identification at all? The large amount of existing pattern recognition algorithms makes testing this hypothesis a tedious task. We therefore suggest to categorize algorithms into a conceptual space with three dimensions [3] and seek to find a systemic correspondence between the placement of models within these dimensions and their applicability to place identification.

The *first dimension* is locality. A local approach processes image data from selected image regions, whereas a global approach always takes the whole image into account. Naturally, local approaches need a detection mechanism to determine regions of interests (ROI). Such mechanisms may rely on low-level image data such as curvature[4], local brightness extrema[5] or generalized features[6]. Ideally, the detection mechanism picks out informative image regions containing objects or landmarks useful to solve the given task.

The *second dimension* measures the invariance to changes in spatial configuration. Algorithms that are sensitive to spatial layout match templates of stored objects against the input image, but fail to generalize if object components are rearranged or scrambled. On the other hand, the largest invariance to spatial layout is provided by models relying on image statistics [7] or bags-of-features like [8]. The class of HMAX models by [2] follow an intermediate approach where invariance to feature locations is increased step-by-step in a multi-layer hierarchy.

The *third dimension* describes how well a model generalizes among several instances of a class. Most local descriptor-based algorithms such as [5, 6] only store patterns specific to the particular instance and view of an object, so multiple patterns are required to describe a class. Usually, category-level generalization can be achieved by clustering specific descriptors into broader categories [9].

These dimensions describe key attributes required for a model to be suitable for place identification. The first dimension, locality, is certainly useful to determine place. If each detected feature is attributed to a position, the relation of these positions provides valuable information in determining the position of an observer[10]. For the second dimension, spatial configuration, the requirements are not so clear. On the one hand, changes in spatial configuration result from changes in position of an observer and invariance to such changes is not desired. On the other hand, invariance to small changes in configuration increase robustness of the detection of features, and could improve detection when scenes are presented under slightly different conditions. The third dimension, generalization properties, are probably required to some extent to generalize different views from the same place onto the same class. Too much generalization is not desirable, because it might project locations that look similar onto the same place.

In the following study, we investigated the invariance properties of models that vary within the second and third dimension. In particular, we varied the two parameters of location and orientation. We judged algorithms based on how well they stayed invariant to changes in orientation compared to their variation induced by changes in location. We tested two holistic models, wavelet-like histograms[7] and texture descriptors called ‘textons’[11]. In comparison, we chose the HMAX model as a hierarchical model of which we analyzed each model step separately. Finally, performance on raw pixel values has been checked as a baseline.

2 Methods

We developed a test setup to evaluate the applicability of pattern recognition methods for place identification. We recorded input images x_α^L at $n_L = 10$ different locations L, and $n_\alpha = 181$ different observer rotation angles α spanning 180 degrees of rotation. If a model S is applied to two input images, the dissimilarity of output vectors can be written as their euclidean distance d^S.

$$d^S(x_{\alpha_1}^{L_1}, x_{\alpha_2}^{L_2}) := \left\| S\left(x_{\alpha_1}^{L_1}\right) - S\left(x_{\alpha_2}^{L_2}\right) \right\|_2 \tag{1}$$

We measure the invariance to rotation $\tilde{I}_{rot}^S(\alpha)$ of a model by averaging the dissimilarity to a midpoint rotation over all locations.

$$\tilde{I}_{rot}^S(\alpha) := \frac{1}{n_L} \sum_L d^S(x_\alpha^L, x_0^L) \tag{2}$$

A low value of $\tilde{I}_{rot}^S(\alpha)$ means the output of the model is highly invariant to the given rotation α. In order to measure usability for place identification, we need to put this value in relation to variations of model outputs achieved by changing the place. We define a relative orientation invariance measure $I_{rot}^S(\alpha)$ as:

$$I_{rot}^S(\alpha) := \frac{1}{n_L} \sum_L \frac{d^S(x_\alpha^L, x_0^L)}{\frac{1}{n_\alpha} \sum_{\alpha'} \min_{L' \neq L} d^S(x_{\alpha'}^L, x_{\alpha'}^{L'})} \tag{3}$$

Values larger than 1 for $\tilde{I}_{rot}^S$ for a given angle mean that the model produces more dissimilar outputs under rotation by that angle than it would by switching the place. We therefore define the maximum angle of invariance α_I^S as the largest value under which this condition is met:

$$\alpha_I^S := \max\left\{ |\alpha| \;\middle|\; I_{rot}^S(\alpha) < 1 \right\} \tag{4}$$

Large values of α_I^S stand for good invariance to rotation compared to changes in place, which attributes the model as suitable for place recognition.

We applied this method to the raw input pixels, as well as outputs from the texton algorithm, wavelet descriptors and the HMAX model at various stages. For the HMAX model, we were particularly interested in how the rotational invariance properties vary with increasing layers. We extracted values at the gabor filter layer (S1), as well as the fist and second local invariance layers (C1, C2) and the final, global invariance layer (C3) At each layer, a maximum of 500 features was extracted. For non-global layers, a random sub sampling over features and locations was done. The same features at the same locations were subtracted for all images.

3 Results

We find that, in accordance with our predictions, pattern recognition models display vastly different performances when investigated for their applicability in

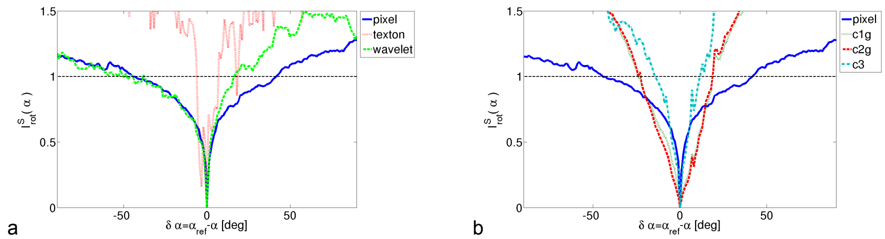


Fig. 2. Relative orientation invariance measures $I^S_{rot}(\alpha)$ for (a) raw pixels, textons and wavelets and (b) different layers of the HMAX model.

place identification. Relative orientation invariance measures $I^S_{rot}(\alpha)$ for holistic models (textons and wavelets) as well as raw pixels are shown in fig. 2a. The maximum angle of invariance for texton outputs ($\alpha_I^{\text{Texton}} = 7°$) is actually lower than for raw pixels ($\alpha_I^{\text{Pix}} = 44°$), which shows that these models are more invariant to changes in location than to changes in rotation compared to raw pixels. Invariance to wavelet transformation is only slightly lower than pixels ($\alpha_I^{\text{Wavelets}} = 38°$).

For HMAX, performance for each layer is shown in figure fig. 2b. Again, α_I sinks below performance on raw pixel down to ($\alpha_I^{\text{Wavelets}} = 24°$) for the successive layers C1 and C2 and further down to ($\alpha_I^{\text{Wavelets}} = 14°$) for the final layer C3. This decay of performance in higher stages of the model show that invariance to place increases faster than invariance to orientation.

4 Discussion

We have investigated the question of how a place can be characterized in terms of visual properties. In particular, we have investigated which invariances are required to uniquely determine a place and how these are related to the invariance properties commonly attributed to visual processing. We have evaluated different models asking how they are able to generalize across all possible views of a place while still being selective enough to guarantee a unique localization.

We have shown that the invariance requirements for place recognition are not necessarily met by models popular for object recognition, such as texton outputs or HMAX. Further, we found that higher layers in the hierarchy of the model, which correspond to more complex features and higher levels of invariance to spatial configuration, lead to a reduced level of invariance to rotation. This yields the hypothesis that invariance to spatial layout, i.e. the second dimension of our conceptual space in fig. 1a, is a detrimental ingredient for invariant place recognition in general. However, since we have explored only a small part of the space of approaches, a more comprehensive study needs to be done.

How much generalization is needed to perform localization? Being able to generalize across different views of the same location is certainly helpful. However, if generalization leads to higher invariance across different locations, as happens in the higher stages of the HMAX model in our case, reliable place identification performance decreases.

Interestingly, [12] proposes a hierarchical model architecture for place cells very similar to that of the HMAX model. In his model, cells are repeated across locations and pooled over increasingly receptive fields in higher stages. The main difference to HMAX lies in that features are trained explicitly to be invariant to rotations using slow feature analysis. This shows that the invariance properties wired into a model greatly affect its suitability for localization, as long as the learning stage is tuned generalize across views, but not across places.

These results suggest that a universal vision system for both object recognition and localization methods is unfeasible. While some of the processing mechanisms may be shared between architectures for the two tasks, specific mechanisms are required to uniquely determine a place.

Acknowledgement This work was supported by DFG, SFB/TR8 Spatial Cognition, project A5-[ActionSpace].

References

1. O'Keefe, J., Dostrovsky, J.: The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain research **34**(1) (1971) 171
2. Serre, T., Oliva, A., Poggio, T.: A feedforward architecture accounts for rapid categorization. PNAS **104**(15) (2007) 6424–6429
3. Wolter, J., Reineking, T., Zetzsche, C., Schill, K.: From visual perception to place. Cognitive Processing **10** (2009) 351–354
4. Zetzsche, C., Barth, E.: Fundamental limits of linear filters in the visual processing of two-dimensional signals. Vision Research **30** (1990) 1111–1117
5. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV **60**(2) (2004) 91–110
6. Kim, S., Kweon, I.: Biologically motivated perceptual feature: Generalized robust invariant feature. In Narayanan, P., Nayar, S., Shum, H.Y., eds.: Computer Vision ACCV 2006. Volume 3852 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2006) 305–314
7. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV **42**(3) (2001) 145–175
8. Grauman, K., Darrell, T.: Efficient image matching with distributions of local invariant features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005. Volume 2. (2005)
9. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology (2007)
10. Se, S., Lowe, D., Little, J.: Vision-based global localization and mapping for mobile robots. IEEE Transactions on Robotics **21**(3) (2005) 364–375
11. Renninger, L., Malik, J.: When is scene identification just texture recognition? Vision Research **44**(19) (2004) 2301–2311
12. Franzius, M., Sprekeler, H., Wiskott, L.: Slowness and sparseness lead to place, head-direction, and spatial-view cells. PLoS Comput Biol **3**(8) (2007) e166