

Integrative Relational Spatial Representation and Reasoning

Hua Hua

A thesis submitted for the degree of
Doctor of Philosophy
The Australian National University

February 2022

© Hua Hua 2021

Except where otherwise indicated, this thesis is my own original work.

Hua Hua
26 February 2022

to my beloved mother.

Acknowledgments

Besides obstacles and struggles in research, I have also experienced the following during my PhD career (not a complete list): hail damage to my Toyota Camry¹; bush fire²; and the still ongoing COVID-19 pandemic. I would like to offer my sincere thanks to the following people without whom I cannot go through these physical and mental difficulties.

- My primary supervisor Prof. Jochen Renz. You treat me like a family member and a friend. You are the most supportive supervisor I have ever known and I will never forget the encouraging words you have said to me. I will also remember the academic skills learnt from you and the inspiring discussions between us.
- My colleagues. Special thanks to Dr. Xiaoyu Ge and Dr. Peng Zhang. You are my elder brothers in research. Thanks for the assistance and suggestions you gave when I was new to research.
- My friends in Canberra. Thank you Dongxu, Siqi, Quyu, Shidi, Ran, Chenchen, Weixuan, Yuli and Rui. I really enjoy those meals and research-relevant discussions we had together. Thank you my basketball friends. Those delightful hours with you in ANU Sports will always be in my memory.
- My roommates. Special thanks to Yintong Liu and his girlfriend. Thank you for keeping me company at home and letting me being the godfather of your kitty Monroe.
- My external collaborators. Special thanks to Prof. Stephan Winter for hosting the discovery project and offering me the opportunity to work together with the PhD students in the University of Melbourne. Special thanks to Prof. Anthony Cohn for the constructive comments you gave on my papers.
- My family members. Special thanks to my grandmother, my sister, and my two lovely nephews. Special thanks to my mother, who always has my back. Thank you my stepfather for bringing happiness to my mom.

This research is also in part supported by Australian Research Council Discovery Project 170100109, and a Google PhD Fellowship.

¹<https://bit.ly/3sLEO1X>

²<https://bit.ly/3sNind6>

Abstract

Qualitative Spatial Representation and Reasoning (QSRR) provides a qualitative and relational understanding of our spatial world, where each scenario is described by the set of qualitative spatial relations between existing objects. Spatial intelligence achieved by QSRR consists of at least two parts: the ability to represent spatial aspects that are indispensable for modeling a certain problem and the power of deriving new knowledge for problem solving based on information at hand. Qualitative spatial representation models consist of qualitative spatial relations in various forms. For example, qualitative topological relations between regions and qualitative direction relations between points. Qualitative spatial reasoning techniques enable converse reasoning, compositional reasoning, consistency checking, and neighborhood-based reasoning. The main advantage of QSRR is that it is close to the way humans describe spatial scenarios and reason about spatial problems. QSRR has been applied in domains like navigation and action description.

Previous research on QSRR pays more attention on theoretical analysis of representation models and reasoning techniques rather than practical issues when applying QSRR in solving real-world problems. To achieve better practical usefulness, qualitative spatial relations should be internally integrated with each other according to their implicit connections; and externally integrated with techniques from other relevant fields. In this thesis, we focus on solutions on four specific research problems to enlighten future research on achieving better practical usefulness of QSRR by integration. First, integrating qualitative direction relations that are described in the same or different frames of reference. We discuss and formally define three different categories of frames of reference; propose a set of rules that enables across-frames-of-reference converse reasoning and compositional reasoning; and define the constraint satisfaction problem in the context of direction relations across different frames of reference. Second, integrating spatial and non-spatial relations between landmarks to achieve human-like localization and navigation. We define qualitative maps to store qualitative relations between places; propose an interactive localization algorithm that is capable of locating agents based on eliminating unsatisfying landmarks; and design a reliable navigation system which takes number of landmarks and qualitative distance information into consideration. Third, integrating structured symbolic spatial representation with unstructured numeric learning methods. We propose an innovative neuro-symbolic approach that recognizes actions from videos with explanations. These explanations are in the form of qualitative spatial object relation chains and thus human-understandable. Forth, integrating various spatial representation models to achieve an efficient model selection strategy. We explore solutions for this problem and propose three different representation determination strategies.

List of Publications

The following is the complete list of publications during my PhD research. Part of them are the basis of this thesis.

Publications

- **Hua Hua**, Dongxu Li, Ruiqi Li, Peng Zhang, Jochen Renz and Anthony Cohn. “Towards Explainable Action Recognition by Salient Qualitative Spatial Object Relation Chains.” To appear in AAAI 2022. Principle author. Dongxu joined discussions on the learning part. Ruiqi, Peng, and Jochen helped with the qualitative spatial representation part. Anthony rectified several scientific mistakes and improved the overall presentation. (Chapter 5)
- **Hua Hua**, Peng Zhang, and Jochen Renz. “Qualitative Place Maps for Landmark-based Localization and Navigation in GPS-denied Environments.” In SIGSPATIAL 2019, *proceedings of the 27th International Conference on Advances in Geographic Information Systems*. Principle author. Peng joined the discussions and helped with the experiments. Jochen improved the overall motivation and presentation. (Chapter 4)
- **Hua Hua**, Jochen Renz, and Xiaoyu Ge. “Qualitative Representation and Reasoning over Direction Relations across Different Frames of Reference”. In KRR 2018, *proceedings of 16th International Conference on Principles of Knowledge Representation and Reasoning*. (Nominated as both Best Paper Award and Best Student Paper Award). (Chapter 3) Principal author. Jochen provided the original idea and joined the discussions. Xiaoyu improved the overall presentation.
- Ruiqi Li, **Hua Hua**, Patrik Haslum, and Jochen Renz. “Unsupervised Novelty Characterization in Physical Environments Using Qualitative Spatial Relations”. In KRR 2021, *proceedings of 18th International Conference on Principles of Knowledge Representation and Reasoning*. Second author. I contributed in discussions, formalizations, and writing.
- Xiaoyu Ge, Jochen Renz, and **Hua Hua**. “Towards Explainable Inference about Object Motion using QualitativeReasoning”. In KRR 2018, *proceedings of 16th International Conference on Principles of Knowledge Representation and Reasoning*. (Short Paper). Third author. I contributed in some of the experiments.
- Ehsan Hamzei, Hao Chen, **Hua Hua**, Maria Vasardani, Martin Tomko, and Stephan Winter. “Deriving place graphs from spatial databases”. In Research@

Locate 2018. Third author. I joined some of the discussions and participated in writing the paper.

- Haonan Li, Ehsan Hamzei, Ivan Majic, **Hua Hua**, Jochen Renz, Martin Tomko, Maria Vasardani, Stephan Winter, and Timothy Baldwin. “Neural Geospatial Question Answering.” *To appear in JOSIS*.

Preprints

- **Hua Hua**, Peng Zhang, and Jochen Renz. “Model Selection-based Automatic Qualitative Spatial-temporal Representation Determination.” *Under revision*. Principle author. Peng and Jochen joined some of the discussions. Jochen also helped with improving the overall motivation and presentation. (Chapter 6)

Contents

Acknowledgments	vii
Abstract	ix
Publications	xi
1 Introduction	1
1.1 Space, Intelligence, and Relations	1
1.2 Qualitative Spatial Representation and Reasoning	2
1.3 Research Problems	3
1.4 Thesis Outline	6
2 Background and Related Work	9
2.1 Qualitative Spatial Representation	9
2.1.1 Topology	9
2.1.2 Direction	10
2.1.3 Distance, Motion, and Size	12
2.1.4 Integrative Qualitative Spatial Representation	13
2.2 Qualitative Spatial Reasoning	14
2.2.1 Converse Reasoning	14
2.2.2 Compositional Reasoning	15
2.2.3 Consistency Checking	15
2.2.4 Neighborhood-based Reasoning	17
2.3 QSRR Applications	17
2.3.1 Navigation	18
2.3.2 Action Description	19
3 Qualitative Direction Relation Integration	23
3.1 Introduction	23
3.2 Related work	25
3.3 Representation	26
3.3.1 Spatial entities	26
3.3.2 Frames of reference	27
3.3.3 Space division and sectors	27
3.3.4 The spatial constraint language $\mathcal{DAF}_m$	28
3.4 Converse of Direction Relations	30
3.5 Composing Direction Relations	32

3.5.1	In-category composition	32
3.5.2	Across-category composition	35
3.6	Anchor Relation Operations	37
3.6.1	Operations over sets of basic relations	38
3.7	Towards Constraint-based Reasoning	38
3.8	Case Study and Applications	40
3.9	Conclusion	42
4	Integrative Qualitative Localization and Navigation	43
4.1	Introduction	43
4.2	Related Work	45
4.3	Qualitative Place Map	48
4.3.1	Direction Relations	49
4.3.2	Distance Relations	50
4.3.3	Extracting Qualitative Knowledge from Natural Language Descriptions	51
4.3.4	Map Construction	52
4.4	Interactive Localization	54
4.5	Reliable Navigation	56
4.5.1	Route Planning	56
4.5.2	Navigation System	58
4.6	Case Study	59
4.7	Experiments	60
4.7.1	Evaluation of Interactive Localization	60
4.7.2	Evaluation of Reliable Navigation	61
4.8	Conclusion and Future Work	63
5	Integrating Qualitative Spatial Relations with Learning-based Methods	65
5.1	Introduction	66
5.2	Related Work	67
5.2.1	Qualitative Spatial Action Description	67
5.2.2	Qualitative Action Recognition Explanation	68
5.2.3	Attention for Explanation	69
5.3	Problem Statement	69
5.4	Proposed Approach	70
5.4.1	Object Relation Chain Extraction	70
5.4.2	Neural Encoding of Object Relation Chains	71
5.4.3	Saliency Estimation and Action Recognition	71
5.4.4	Explanation Generation	73
5.4.5	Specialized Hyper-parameters	74
5.5	Experiments and Evaluation	75
5.5.1	Baseline Approaches	75
5.5.2	Datasets	76
5.5.3	Hyper-parameter Finetuning	76

5.5.4	Explanation Evaluation Metric	78
5.5.5	Experimental Results and Analysis	79
5.5.6	Large-scale Dataset	80
5.6	Conclusion and Future Work	81
6	Determining Integrative Relational Spatial Representation	83
6.1	Introduction	83
6.2	Related Work	84
6.2.1	Spatial Aspects	85
6.2.2	Object Forms and Relation arities	85
6.2.3	Granularities	86
6.3	Candidate Q-models	86
6.4	Model Selection Strategies	88
6.4.1	Naive	88
6.4.2	Random Climbing	89
6.4.3	Fast Selection	90
6.5	Q-representation Determination	92
6.6	Demonstrative Evaluation	92
6.6.1	Features of Determined Q-Representations	93
6.6.2	Effectiveness and Efficiency	93
6.7	Conclusion and Future Work	94
7	Conclusion	95
7.1	Contributions and Impact	95
7.2	Future Work	96
7.2.1	Integrative Qualitative Distance Relations	97
7.2.2	Human-like Localization and Navigation	97
7.2.3	Trustworthy AI Agents	97
7.2.4	Spatial Knowledge Graph	98
7.2.5	Automatic Qualitative Spatial Representation Determination	98
7.3	Summary	98

List of Figures

1.1	A simple spatial intelligence test: which of the following objects fits into the object at the top?	1
1.2	Three RCC-8 relations: DC (disconnected), EC (externally connected), and PO (partially overlapping) [Randell et al., 1992b].	2
1.3	Three robots in different colours: red, blue, and orange. They stand in a row and are in certain direction relations.	3
1.4	Internal integrations between qualitative spatial relations; external integrations between qualitative spatial relations and other artificial intelligence representations and techniques.	4
1.5	(a) Given the direction relation between place A and B, we may reach A from B by turning left. (b) However, place A is not reachable from B if there is an obstacle in-between.	5
1.6	Categories of typical QSRR models and their relationships [Dylla et al., 2017]. Some of them will be introduced in detail in Chapter 2.	6
1.7	An overview of integrations in Chapter 3-6.	6
2.1	Visualization of RCC-8 relations [Randell et al., 1992b].	10
2.2	Deriving RCC-5 from RCC-8 [Randell et al., 1992b; Chen et al., 2015].	10
2.3	Cone-based and projection-based cardinal direction relations [Frank, 1991]. (a) 4 cones. (b) 8 cones. (c) One projection: north/south division. (d) One projection: west/east division. (e) Two projections: combining the north/south division with the west/east division. (f) Two projections with neural zone (U).	11
2.4	(a) In the Flipflop calculus, the space is divided into 6 parts with AB as the baseline. (b) DCC divides the space into 15 sectors by 3 lines. (c) In $STAR_4$, the 2D space is divided into 16 sectors by 4 lines. (d) $OPRA_m$: Oriented point A is in relation of (3, 13) with oriented point B because A is in Sector 3 of the space division centered with B and B is in Sector 13 of that with A. (e) In TPCC, the frame of reference for direction relation is decided by the line from the relatum to the origin. Distance relation is also considered. For example, “cfl” is short for “close-front-left” while “dfl” is for “distant-front-left”. (f) In CDR, the space is divided into 9 parts, where the central tile is decided by the shape of the relatum object.	12

-
- 2.5 (a) One possible configuration of QDC_4 , where the space is divided into 4 sections based on the distance from the center point. (b) One possible configuration of QDC_3 , where the space is divided into 3 sections. (c) The arrows starting from a or b show the moving directions of a or b. Longer arrow indicates greater speed. (top) The first – means a is moving towards b; the second – indicates b is also moving towards a; the last + means a's speed is greater than that of b. (bottom) The first – means a is moving towards b; the second + shows b is moving away from a; the last 0 means they move in the same speed. . 13
- 2.6 (a) Converse table of Point Algebra (PA). $r^\sim$ means the converse of r. (b) Composition table of PA. Source: [Dylla et al., 2017] 14
- 2.7 (a) A basic version of the correspondence between the Double Cross Calculus relations and their natural language interpretations. (b) An extended version of such correspondence. Source: [Krieg-Brückner and Shi, 2006]. 18
- 2.8 (a) $OPRA_2$ is applied to described two moving boats. (b) $OPRA_2$ is extended so that it is possible to express two boats are moving closer (+), moving in parallel (P), or moving away from each other (-). Source: [Dylla, 2009]. 19
- 2.9 Two predefined sets of qualitative spatial relations to describe moving vehicle relations [Fernyhough et al., 2000]. (a) Qualitative relative position relations. (b) Qualitative relative motion direction relations. . . 19
- 2.10 The qualitative action descriptions in [Sridhar et al., 2011a] consider five sets of qualitative spatial relations: topological relations; direction relations; speed relations; size relations; and moving trajectory relations. 20
- 3.1 (a) Start at the intersection and then go south until the library is located by relating to the bus station.(b) Absolute FoR: based on north, the direction "you" go is south. (c) Intrinsic FoR: based on the direction "you" are facing to the bus station is to the left of "you". (d) Relative FoR: based on the direction from the bus station to "you" the library is to the right of the station. 24
- 3.2 (a) "b is to the north of a and c is to the north of b", which is inconsistent with (b) "from a's perspective, c is in front of b", because based on the former, b is between a and c while based on latter b is not. . . . 25

-
- 3.3 (a) The value of an intrinsic anchor ϕ_a or a relative anchor ϕ_{ab} is decided by the value of the anticlockwise angle deviated from the absolute anchor ϕ_0 (α for ϕ_a , β for ϕ_{ab}). (b) In an absolute FoR, the 2-D plane is divided into 16 sectors with the relatum a as the centroid and the absolute anchor ϕ_0 decides the direction of Sector 0. The absolute direction relation between a and b is decided by the value of the angle that is formed by vector **ab** anticlockwise deviating from ϕ_0 . (c) In an intrinsic FoR, the 2-D plane is divided into 16 sectors with the relatum a as the centroid and the intrinsic anchor ϕ_a (i.e. a 's orientation) decides the direction of Sector 0. The intrinsic direction relation between a and b is decided by the value of the angle that is formed by vector **ab** anticlockwise deviating from ϕ_a . (d) In a relative FoR, the 2-D plane is divided into 16 sectors with the relatum b as the centroid and the relative anchor ϕ_{ab} (i.e. the orientation from the observer a to the relatum b) decides the direction of Sector 0. The relative direction relation between a , b and c is decided by the value of the angle that is formed by vector **bc** anticlockwise deviating from ϕ_{ab} 26
- 3.4 (a) $\beta = \alpha + \pi$ where $\alpha = \phi_{ab}$ and $\beta = \phi_{ba}$ (b) $\gamma = \beta + \pi - \alpha$ where $\alpha = \phi_b - \phi_a$, $\beta = \phi_{ab} - \phi_a$, and $\gamma = \phi_{ba} - \phi_b$ 30
- 3.5 (a) when R_{abc}^i and $0 < i < 2m$, $0 < \beta < \alpha$ where $\alpha = \phi_{bc} - \phi_{ab}$ and $\beta = \phi_{ac} - \phi_{ba} - \pi$. (b) when R_{abc}^i and $2m < i < 4m$, $0 < \beta < \alpha$ where $\alpha = 2\pi - (\phi_{bc} - \phi_{ab})$ and $\beta = \pi - (\phi_{ac} - \phi_{ba})$ 32
- 3.6 (a) when A_{ab}^i, A_{bc}^j and $0 < j - i < 2m$, $0 < \beta < \alpha$ where $\alpha = \phi_{bc} - \phi_{ab}$ and $\beta = \phi_{ac} - \phi_{ab}$. (b) when A_{ab}^i, A_{bc}^j and $2m < j - i < 4m$, $0 < \beta < \alpha$ where $\alpha = 2\pi - (\phi_{bc} - \phi_{ab})$ and $\beta = 2\pi - (\phi_{ac} - \phi_{ab})$ 33
- 3.7 (a) $\alpha = \beta - \gamma$ where $\alpha = \phi_{ab} - \phi_a$, $\beta = \phi_{ab} - \phi_0$ and $\gamma = \phi_a - \phi_0$. (b) $\gamma = \alpha + \beta$ where $\alpha = \phi_{bc} - \phi_b$, $\beta = \phi_b - \phi_a$ and $\gamma = \phi_{bc} - \phi_a$ 34
- 3.8 (a) when R_{bcd}^j and $0 < j < 2m$, $0 < \gamma - \alpha < \beta$ where $\alpha = \phi_{bc} - \phi_{ab}$, $\beta = \phi_{cd} - \phi_{bc}$ and $\gamma = \phi_{bd} - \phi_{ab}$. (b) when R_{bcd}^j and $2m < j < 4m$, $0 < \alpha - \gamma < 2\pi - \beta$ where $\alpha = \phi_{bc} - \phi_{ab}$, $\beta = \phi_{cd} - \phi_{bc}$ and $\gamma = \phi_{bd} - \phi_{ab}$. 36
- 3.9 (a) $\alpha + \beta = 0$ where $\alpha = \phi_b - \phi_a$ and $\beta = \phi_a - \phi_b$. (b) $\gamma = \alpha + \beta$ where $\alpha = \phi_b - \phi_a$, $\beta = \phi_c - \phi_b$ and $\gamma = \phi_c - \phi_a$ 37
- 3.10 (a) An absolute direction relation A_{ab}^i can be transformed into an intrinsic direction relation $I_{a'b}^i$ where $S_{a'a}^2$ and $\Phi_{a'0}^0$. (b) An intrinsic direction relation I_{ab}^i can be transformed into a relative direction relation $R_{a'ab}^i$ where $\phi_a = \phi_{a'a}$ 40
- 3.11 (a) $\alpha + \beta = 0$ and γ_1, γ_2 and γ_3 can form the triangle abc . (b) Relative positions of the hunters (**a** and **b**), the dog (**d**) and the hare (**h**). ϕ_0 indicates the north and ϕ_b is the orientation of **b**. 41
- 4.1 (a) By using the qualitative direction relation calculus $\mathcal{RDR}_4$, the direction relation between A , B and C is 1. (b) The direction relation labelling function that assigns the 16 relations with the labels "front, left, back and right". 49

4.2	(a) The sector model used by Klippel et al. f: front, hl: half left, l: left, sl: sharp left, b: back, sr: sharp right; r: right; hr: half right. (b) An equivalent sector model with $m = 16$ and a dedicated direction relation labelling function.	50
4.3	(a) The default qualitative distance calculus QDR_5 . vc: very_close, c: close, n: neural, f: far, vf: very_far. (b) Another qualitative distance calculus QDR_2 . c: close, f: far.	50
4.4	Assuming v_4 is a landmark and is visible from v_3 . When moving from v_1 to v_3 via v_2 , v_4 is helpful as a landmark that might be used to locate v_3 . However, v_4 can also increase the unreliability because it is the turn of (v_1, v_2, v_3) is instruction equivalent to the turn of (v_1, v_2, v_4)	58
4.5	A qualitative place map constructed from two route descriptions.	59
5.1	(a) Essential spatial relations when performing the action of <i>car parking</i> : at first, the car (grey rectangle) is disconnected from the parking space (dotted rectangle) and then they overlap; in the end, the car is included in the parking space. (b) An accidental interaction between the car and the parking space.	66
5.2	An overview of our approach: the input is a set of video clips and the outputs are a recognized action and its explanations. Our approach consists of three components: input, saliency estimation, and outputs. Solid rectangles are modules and dotted rectangles are data in different forms. Our innovative modules in each component are object relation chain extractor (blue), neural saliency estimator (yellow), and explanation generator (green).	71
5.3	Our neural saliency estimator will be applied to each object relation chain vector in M^v : each k-hot encoded x is mapped into a saliency score vector s by our pairwise co-occurrence saliency mapping matrix M^{cs} . Then s is normalized by the softmax function. The normalized saliency score vector s' is used to constitute the normalized saliency score matrix M^{os} and produce the saliency weighted object relation chain vector x^{sw} by elementwise multiplied with x . The saliency weighted object relation chain matrix M^{sw} consists of x^{sw} from each video clip and it is the input of our action classifier.	72
5.4	Changes in action recognition accuracy and top1 rates as the number of epochs increases. Results are from training our approach on CAD120++ with different number of epochs. The red and blue lines show the changes in action recognition accuracy and top1 rate respectively.	80
5.5	A correctly recognized action (top) and a wrongly recognized action (bottom). Right: top 5 salient object relation chains in the explanation. Ground truth object relation chains are in bold. Objects in red are involved in ground truth object relation chains while irrelevant objects are in yellow.	81

6.1	(a) Two examples of QTC_B relations: (left) the first $-$ means a is moving towards b (or it is moving to the region that is right to Line l_1); the second $-$ indicates b is also moving towards a (or it is moving to the region that is left to Line l_2); the last $+$ means a 's speed is greater than that of b ; (right) the first $-$ means a is moving towards b ; the second $+$ shows b is moving away from a (or it is moving to the region that is right to Line l_2); the last 0 means they have the same speed. (b) $QAMD_1$ (left): QTC_B is integrated with $STAR_1$. The moving direction of an object can be described with 0, 1, 2 or 3; in contrast, $QAMD_2$ (right) is finer-grained and describes moving directions with 8 discrete values..	85
6.2	(a) One possible definition of $QASz_3$: rectangles are divided into three distinct groups according to their sizes (or areas) and their qualitative sizes will be denoted as 0^{sz} , 1^{sz} , and 2^{sz} respectively. $QASz_3$ relations are in the forms of $i^{sz}j^{sz}$ where $i, j \in \{0, 1, 2\}$. (b) $QASp_4$: qualitative speeds of moving objects (arrow orientation indicates the moving direction and arrow length indicates magnitude of moving speed) are described with 4 discrete values 0^{sp} , 1^{sp} , 2^{sp} and 3^{sp} regardless of their moving directions. $QASp_4$ relations are in the forms of $i^{sp}j^{sp}$ where $i, j \in \{0, 1, 2, 3\}$	87
6.3	The default hierarchical tree of candidate q-models. Models marked with $*$ are first formally proposed in this chapter.	89
7.1	Summary of key contributions and impact of Chapter 3-6.	95

List of Tables

4.1	Evaluation results of interactive localization. ‘v’ represents the number of vertices in the place graph; ‘e’ represents the number of edges in the place graph; ‘r’ is the number of requeries; ‘ r_{ave} ’ is the average number of requeries made per guidance.	61
4.2	Evaluation results for different cost functions on a place graph with 200 vertices and 200 edges. “#instructions” is the number of instructions provided by the navigator to the agent. “#wrong turns” is the number of incorrect turns made by the agent because of ambiguous turns. “distance” is the travelling distance (in units of length) of the agent including wrong movements and movements by the interactive localization mechanism. All the data in the following 4 tables are averaged per trip. “SMR” is short for the shortest most reliable path algorithm and “ILRN” is short for our interactive localization and reliable navigation algorithm.	61
4.3	Evaluation results for different weight configurations on a place graph with 200 vertices and 300 edges.	62
4.4	Evaluation results for different weight configurations on a place graph with 500 vertices and 500 edges.	62
4.5	Evaluation results for different weight configurations on a place graph with 500 vertices and 750 edges.	62
5.1	CAD120 and CAD120++ datasets. #videos: number of videos. #acts: number of action categories. #objs: number of object categories. #frames: number of frames per video. g.t.: whether ground truth action definitions are provided.	76
5.2	Action definitions in CAD120++. No definition for <i>null</i> ; <i>close</i> and <i>open</i> are defined as attribute transitions of a specific object; and the other seven actions are defined as relation transitions between a pair of subject and object. To save space, m-box is short for “medicine-box” and micr is short for “microwave”.	77
5.3	Best action recognition accuracies given different qualitative spatial calculi are applied to describe actions in CAD120++ videos. Both the default number of splits and the default maximum length of object relation chains are set as 2.	78
5.4	Action recognition accuracies given various number of clips a video is divided into. The default choice of qualitative spatial calculi is RCC-4 and the default maximum length of object relation chains is 2.	78

5.5	Action recognition accuracies given different maximum length of object relation chains. The default choice of qualitative spatial calculi is RCC-4 and the default number of clips is 2 for CAD120++, and 5 for CAD120.	79
5.6	Results of experiments on CAD120++ and CAD120. No.: experiment indexes. approach: the approach to action recognition and explanation generation. intui_thres is short for intuitive thresholding; intui_max is short for intuitive maximum; and ours stands for our approach. accuracy is the metric for action recognition while top1/2/5 rates is to evaluate generated explanations. Results in accuracy and top1/2/5 rates are in percentage (%) and best results are highlighted in bold. . . .	80
6.1	Scores achieved by different model selection strategies (Naive, Rdm, FastC, FastF, and FastR) given different number of positive (+) and negative (−) examples (as in Column 1). Rdm is short for Random Climbing. FastC means Fast Selection with the coarsest q-models from each aspect as the initial q-models; FastF is Fast Selection with the finest initial q-models; and FastR means Fast Selection with random initial q-models. The average numbers of iterations are in brackets for the three Fast Selection strategies.	94

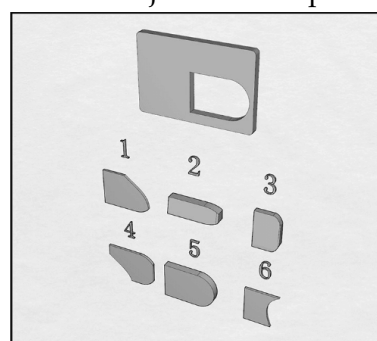
Introduction

“The essence of man is no abstraction inherent in each single individual. In reality, it is the ensemble of the social relations.” – Karl Marx¹

1.1 Space, Intelligence, and Relations

Space is a fundamental component of our living world [Renz and Nebel, 2007]. Spatial intelligence is the wisdom and smartness that tackle problems about space. An example of a spatial intelligence test is shown in Figure 1.1². In this spatial problem, we are asked to determine which object (indexed from 1 to 6) have the same shape as the hole in the rectangle object. As an essential part of intelligence, the mechanism of human spatial intelligence is an important research topic in psychology and pedagogy. In contrast, artificial intelligence researchers are more interested at building AI agents with human-like spatial intelligence.

Figure 1.1: A simple spatial intelligence test: which of the following objects fits into the object at the top?



Relations are ubiquitous in our world. In other words, it is impossible to find an object that is absolutely independent from other objects. Object relations can be space-relevant or not. To solve the problem in Figure 1.1, we are required to analyze the “same shape” relations between objects.

¹In the Sixth of the Theses on Feuerbach, 1845.

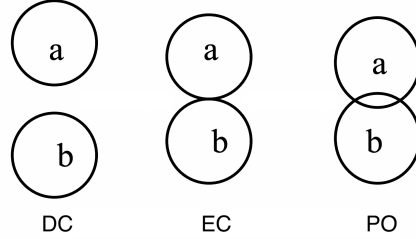
²Source: <https://www.arealme.com/spatial-intelligence-test/en/>

1.2 Qualitative Spatial Representation and Reasoning

In this thesis, we will focus on spatial intelligence that is based on Qualitative Spatial Representation and Reasoning (QSRR) [Renz and Nebel, 2007; Cohn and Renz, 2008], which is a sub-field of Knowledge Representation and Reasoning (KRR), Artificial Intelligence. Early works in this field were inspired by [Allen, 1990], where Interval Algebra is introduced to represent and reason about the relations between time intervals. There are 13 qualitative temporal relations in Interval Algebra: “before”, “equal”, “meets”, etc. For example, $X\text{-before-}Y$ indicates interval X happens before interval Y . Or, mathematically, given $X = (s_X, e_X)$, $s_X < e_X$, and $Y = (s_Y, e_Y)$, $s_Y < e_Y$, if $e_X < s_Y$, then $X\text{-before-}Y$ is true.

In the following several decades, many innovative spatial algebra or calculi have been designed [Chen et al., 2015; Dylla et al., 2017]. These models have covered most of the fundamental spatial (or spatial-temporal) aspects of the world: topology, direction, size, distance, velocity, shape, etc. One of the most influential spatial models is the Region Connection Calculus (RCC) introduced in [Randell et al., 1992b]. Its original version RCC-8 consists of 8 topological relations: “DC”, “EC”, “PO”, etc. As shown in Figure 1.2, “DC” means “disconnected”, “EC” is short for “externally connected” and “PO” is “partially overlapping” (Figure 1.2).

Figure 1.2: Three RCC-8 relations: DC (disconnected), EC (externally connected), and PO (partially overlapping) [Randell et al., 1992b].



Spatial intelligence achieved by QSRR consists of at least two parts: representation ability (the ability to represent spatial aspects that are indispensable for modeling a certain problem) and reasoning power (deriving new knowledge for problem solving based on information at hand [Cohn and Renz, 2008]). Correspondingly, two of the most fundamental research problems in this field are: first, how spatial scenarios are described in the forms of qualitative spatial relations between objects; and second, how to derive spatial knowledge that is essential in problem solving.

QSRR provides a qualitative and relational understanding of our spatial world, where each scenario is described by the set of qualitative spatial relations between existing objects. A qualitative spatial representation of the scenario shown in Figure 1.3³ can be “red-left-blue, blue-left-orange, and red-left-orange”, where “red”, “blue” and “orange” represent the three robots in corresponding color respectively and “left” is a

³Source: <https://www.pinterest.ie/pin/230668812446945035/>

spatial relation that indicates a robot is on the left-side of another robot from the reader's perspective.

Figure 1.3: Three robots in different colours: red, blue, and orange. They stand in a row and are in certain direction relations.



Given the same scenario, an example of qualitative spatial reasoning is: given the facts of *red-left-blue* and *blue-left-orange*, it is not difficult to infer *red-left-orange*. Such a reasoning paradigm is called “compositional reasoning”, which can be concluded as deriving the relation between object A and object C given the relation between object A and object B and that between object B and object C. There are also other forms of reasoning. For example, converse reasoning: inferring the relation between A and B given the relation between B and A; and consistency checking: to check whether there is at least one configuration of objects that can satisfy all the constraints implied by their relations. The problem of “in the same frame of reference, is it possible to find three robots A, B, and C that are in relations of *A-left-B*, *B-left-C*, and *A-right-C*?” can be solved by consistency checking. The answer is “no” because from *A-left-B* and *B-left-C* we can infer *A-left-C*, which is conflict with *A-right-C*.

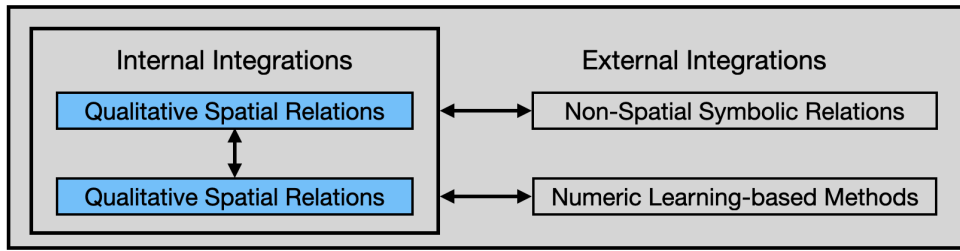
QSRR models are close to human cognition because spatial relations between real-world objects are described in a qualitative manner [Renz and Nebel, 2007]. For example, in many cases, we human beings tend to describe the distance relation between two objects with “near” or “far” and neglect the exact numeric value of distance: “my desk is close to my bed” rather than “my desk is 10.25 centimeters away from my bed”. Therefore, these models are helpful for achieving better human-computer interaction by, for example, learning to interpret spatial semantics in natural language in the form of qualitative spatial relations [Kordjamshidi et al., 2013]. There are also other forms of QSRR applications. For example, detecting rule conflicts in marine navigation [Dylla, 2009; Kreutzmann et al., 2013].

1.3 Research Problems

In the past three decades, numerous qualitative spatial representation models have been introduced to describe spatial knowledge in specific scenarios with qualitative spatial relations. There are also a large number of developments on reasoning techniques so that, in certain cases, it is possible to implement converse reasoning,

compositional reasoning, and consistency checking in an automatic manner. However, there are few real-world QSRR applications [Renz and Nebel, 2007]; and there are also few discussions on improving the practical usefulness of QSRR in real-world scenarios. In other words, previous research on QSRR pays more attention on theoretical analysis of representation models and reasoning techniques rather than practical issues when applying QSRR in solving real-world problems.

Figure 1.4: Internal integrations between qualitative spatial relations; external integrations between qualitative spatial relations and other artificial intelligence representations and techniques.



The first three QSRR open problems proposed in [Dylla et al., 2017] are: *combining qualitative abstractions*, *integration with other symbolic methods*, and *integration with quantitative approaches*. Inspired by their discussions, the argument to be defended in this thesis is “to achieve better practical usefulness, qualitative spatial relations should be internally integrated with each other according to their implicit connections; and externally integrated with techniques from other relevant fields”. A visualized and more detailed interpretation of integrations is shown in Figure 1.4, which is based on the following observations.

First, in most QSRR models, qualitative spatial relations are determined in a fixed frame of reference while in real-world scenarios the frames of reference are changeable. For example, in Figure 1.3, the direction relations between robots are different when described from different frames of reference: from readers’ perspective, the red robot is to the left of the blue robot, while from the blue robot’s perspective, the red robot is at its righthand side. These implicit frames of reference have to be considered when we deal with descriptions made in various frames of reference. The same issue also exists when describing distance relations or size relations.

Second, QSRR models only represent and reason about spatial knowledge. Spatial knowledge and non-spatial knowledge should be manipulated together in order to provide complete solutions to real-world problem. For example, from a bird’s eye view, place A is to the left of place B. If only spatial knowledge is considered, as shown in Figure 1.5(a), we may ask an agent to “turn left” to reach place A from place B. However, as shown in Figure 1.5(b), there might be an obstacle between place A and B so that place A is neither reachable nor observable from place B. Without considering non-spatial knowledge, misleading navigation instructions might be given to users.

Third, QSRR deals with qualitative input and is only able to provide qualitative

Figure 1.5: (a) Given the direction relation between place A and B, we may reach A from B by turning left. (b) However, place A is not reachable from B if there is an obstacle in-between.



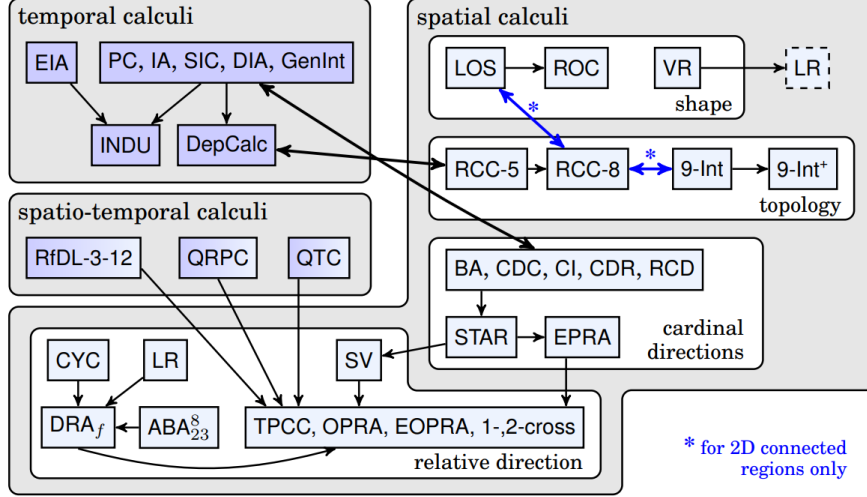
solutions to problems related to spatial representation and reasoning. In contrast, learning-based approaches [LeCun et al., 2015] are based on large-scale numeric data and their usefulness in solving practical problems have already been proved in recent years. As introduced earlier, QSRR is close to human cognition while learning-based approaches are weak in human understandability. Since researchers in these two fields follow distinct philosophies and methodologies, it is challenging to apply qualitative spatial relations to improve the human-likeness of learning-based approaches.

Fourth, as shown in Figure 1.6, QSRR models are inter-connected; belong to various categories (spatial, temporal or spatio-temporal) and cover different spatial aspects (topology, direction or shape). In most previous applications, people manually decide which qualitative spatial representation model(s) to use while there are very few discussions on how to implement an integrative model selection strategy, where the inter-connections, similarities, and differences between representation models are taken into consideration. Other model selection relevant problems are: How to measure the appropriateness of the selected models? Is it possible to develop an effective automatic model selection strategy that is capable of selecting appropriate representation models for problem solving in various scenarios? If so, how to prove its effectiveness regarding unbiased model selection and its scalability in manipulating large scale data?

It is, of course, impossible to develop an integration guideline that is applicable in all scenarios. Instead, we aim to solve the following specific research problems to enlighten future research on achieving better practical usefulness of QSRR by integration.

- How to integrate qualitative direction relations that are described in the same or different frames of reference? (Chapter 3)
- How to integrate spatial and non-spatial relations between landmarks to achieve human-like localization and navigation? (Chapter 4)
- How to integrate structured symbolic spatial representation with unstructured

Figure 1.6: Categories of typical QSRR models and their relationships [Dylla et al., 2017]. Some of them will be introduced in detail in Chapter 2.



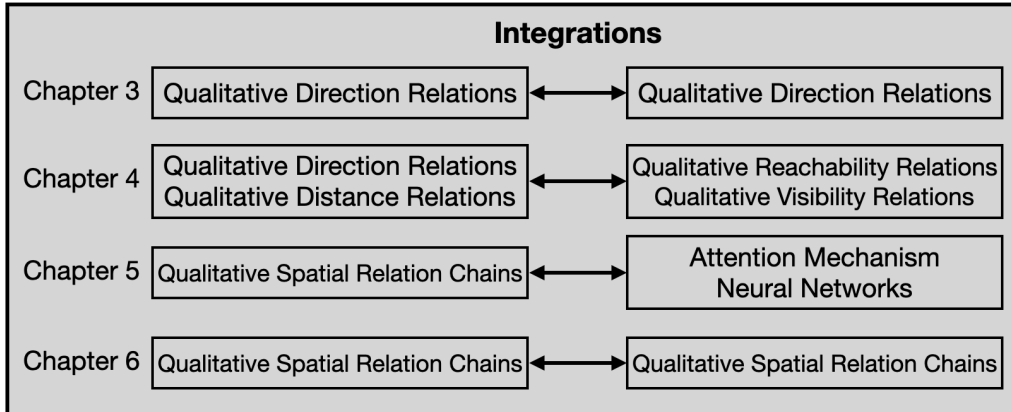
numeric learning methods? (Chapter 5)

- How to integrate various spatial representation models to achieve an efficient model selection strategy? (Chapter 6)

1.4 Thesis Outline

We will examine the above research problems in detail in this thesis. First, in Chapter 2, we provide an overview of qualitative spatial representation and qualitative spatial reasoning techniques; and introduce typical QSRR applications in two domains: navigation and action description. Then, as shown in Figure 1.7, the rest of this thesis is organized as follows.

Figure 1.7: An overview of integrations in Chapter 3-6.



In Chapter 3, we focus on solving the problem of qualitative direction relation integration. First we discuss and formally define three different categories of frames of reference that are, in most cases, implicit in describing direction relations. We name the implicit deciding direction of a frame of reference as “anchor”. Based on these formal definitions, an innovative spatial constraint language $\mathcal{DAF}_m$ is introduced to express and integrate direction information across different frames of reference. We also propose a set of rules that enables across-frames-of-reference converse reasoning and compositional reasoning. Moreover, we are the first to define the constraint satisfaction problem in the context of direction relations across different frames of reference. To enlighten future complete solutions on this constraint satisfaction problem, we conduct formal discussions on local solutions as well as theoretical properties of our proposed spatial constraint language. In a case study at the end of Chapter 3, we prove our integrative constraint language and reasoning rules are able to model and reason about direction relations across frames of reference.

Our goal in Chapter 4 is to enable integrative qualitative localization and navigation by integrating qualitative spatial relations (direction relations and distance relations) with non-spatial symbolic relations (reachability relations and visibility relations). We implement a series of mechanisms that is able to localize and navigate agents given only qualitative knowledge of places. Qualitative place knowledge can be extracted from natural language descriptions and our mechanisms are motivated by human communication and aim to assist human localization and navigation in an intuitive and interactive way. Specifically, extracted qualitative place knowledge is stored in a map. This qualitative map is a graph where the vertices represent places and edges are labelled with different categories of qualitative relations between places. Based on such maps, an interactive localization algorithm is developed that is capable of locating agents based on eliminating landmarks that is unable to satisfy the constraints implied by user observations; and a reliable navigation system is introduced which (different from previous methods) takes number of landmarks and qualitative distance information into consideration. We conduct a case study to further explain our mechanisms and a series of experiments to prove our solution is better than previous solutions on qualitative localization and navigation.

Our first trial on integrating qualitative spatial relations with learning-based methods will be described in detail in Chapter 5. We propose an innovative neuro-symbolic approach that recognizes actions from videos with explanations. These explanations are in the form of qualitative spatial object relation chains and thus human-understandable. We first introduce the process of representing videos symbolically by qualitative spatial object relation chains. We further utilize an attention mechanism that captures the correlations between such object relation chains and actions. Object relation chains are salient if they are deciding in recognizing specific actions. We evaluate our methods on two video datasets and the experimental results show our approach achieves better performance in both recognition accuracy and quality of action explanations than previous symbolic approaches.

In Chapter 6, we aim to explore solutions for the problem of automatic integrative qualitative spatial representation determination. In most previous work, the set

of qualitative spatial calculi for representation is manually chosen. In this chapter, we discuss the disadvantages of manual determination and propose three different representation determination strategies: naive, random climbing, and fast selection. The naive strategy is to choose the best combination from checking the performance of all possible combinations; the random climbing strategy is to randomly choose a new qualitative spatial calculus and adjust the set of current choice according to the change in performance; and the fast selection strategy is to fast determine the set of qualitative spatial calculi for representation by prior knowledge and several specific assumptions. We demonstrate and evaluate these three strategies with a small-scale simulated dataset.

We conclude this thesis and outline future work in Chapter 7. We summarize our four key contributions in four research problems: qualitative spatial relation integration, integrative qualitative localization and navigation, integrating QSRR with deep learning, and integrative qualitative spatial representation selection. Correspondingly, we discuss our impacts on four higher-level research domains: qualitative spatial relation integration, human-like localization and navigation, trustworthy AI agents, and automatic spatial representation determination. We discuss future work in several promising directions on achieving more practically-useful spatial intelligence.

Background and Related Work

“I say with Didacus Stella, a dwarf standing on the shoulders of a giant may see farther than a giant himself.” – Robert Burton¹

This thesis describes our efforts on improving the applicability of qualitative spatial representation and reasoning (QSRR) by an integrative methodology. In this chapter, we will introduce research that is relevant to QSRR and its applications. As a quick overview of qualitative spatial representation and reasoning (QSRR), in the first two sections, we will introduce a couple of representative qualitative spatial calculi and explain mainstream qualitative spatial reasoning techniques. Readers can refer to [Renz and Nebel, 2007; Cohn and Renz, 2008; Chen et al., 2015; Dylla et al., 2017] for a complete and detailed investigation of QSRR. In Section 2.3, we will introduce typical QSRR applications.

2.1 Qualitative Spatial Representation

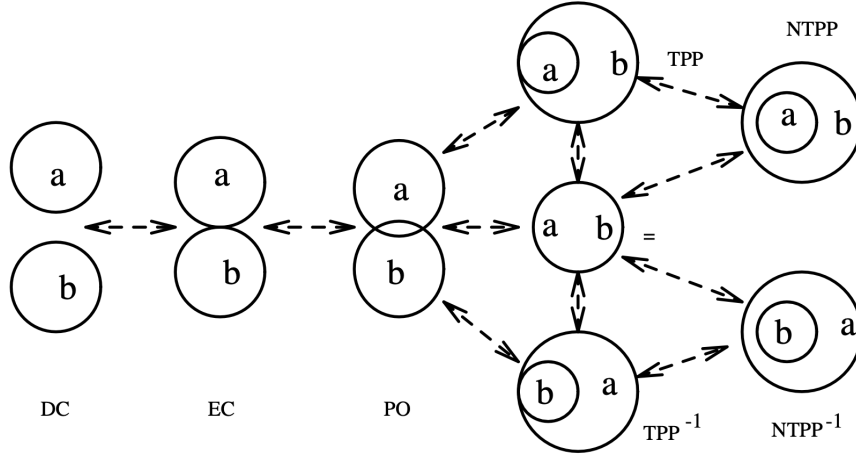
Qualitative spatial relations are formally defined in qualitative spatial calculi. In this section, we will focus on the representation ability of several representative qualitative spatial calculi. Specifically, for each introduced qualitative spatial calculus, we will describe its components, the spatial aspect it covers, its granularity, object forms and relation arities. We will also introduce

2.1.1 Topology

Topological relations are the relations between object frontiers. The Region Connection Calculus (RCC) [Randell et al., 1992b] is one of the most well-known and widely applied qualitative spatial calculi. RCC describes the topological relations between regions and its original form is RCC-8, which consists of eight binary relations as in Figure 2.1: DC (disconnected), EC (externally connected), = (equal), PO (partially overlapping), TPP (tangential proper part), TPPi (tangential proper part inverse), NTPP (non-tangential proper part) and NTPPi (non-tangential proper part inverse).

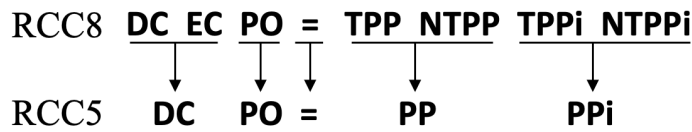
¹In the second edition of *The Anatomy of Melancholy*, 1624.

Figure 2.1: Visualization of RCC-8 relations [Randell et al., 1992b].



RCC-5 is a coarser version of RCC-8 because it is derived from RCC-8 by combining tangential relations with its neighbor relations [Randell et al., 1992b]. For example, as shown in Figure 2.2, PP is a RCC-5 relation and it is a combination of two RCC-8 relations: TPP and NTTP. Other coarser RCC versions are RCC-2, RCC-3, and RCC-4 [Gatsoulis et al., 2016]. From RCC-5 to RCC-2, topological relations are defined with gradually coarser granularities: from RCC-5 to RCC-4, = is combined with PP; in RCC-3, PPi is neglected; and RCC-2 only discriminates whether two regions are connected. There are also finer versions of RCC. For example, RCC-15 and RCC-23 consider whether the convex hull of a region is inside, partly inside, or outside that of another region [Cohn et al., 1997]; RCC-62 [OuYang et al., 2007] is a refinement of RCC-23 by extending the 9-intersection model [Egenhofer and Herring, 1990] into the 16-intersection model.

Figure 2.2: Deriving RCC-5 from RCC-8 [Randell et al., 1992b; Chen et al., 2015].

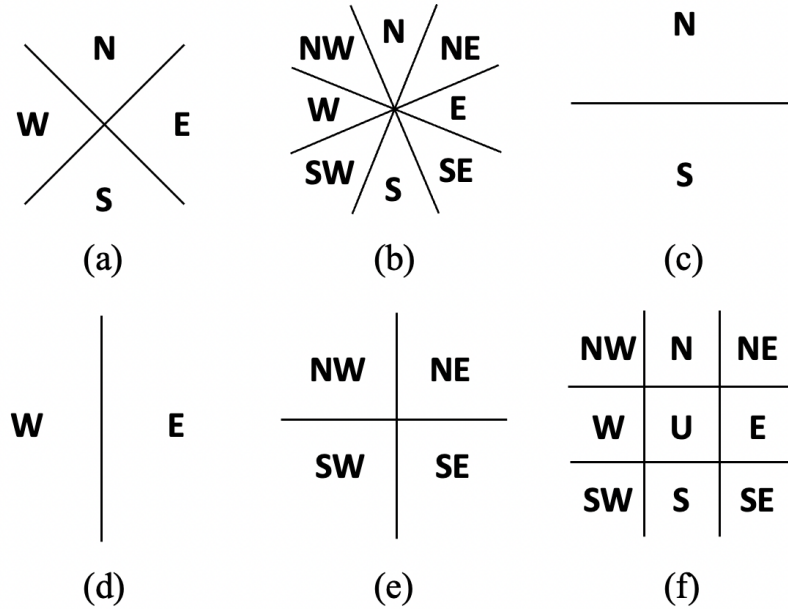


2.1.2 Direction

Direction relations are common in daily communications. In the past three decades, numerous qualitative direction calculi have been created to model various forms of direction relations. Cardinal directions are defined in a fixed implicit frame of reference. For example, Object A is to the north of Object B if A is closer to the north pole. In this case, north is a cardinal direction and the baseline for direction deciding is the direction pointing to the north pole. As shown in Figure 2.3, [Frank, 1991] has

proposed two forms of cardinal direction relation representations: cone-based and projection-based. For example, as in Figure 2.3 (f), the space is divided into 9 sections by two projections and considering the neural zone. Given two objects A and B, the direction relation of B to A is decided by the region B locates with A in the center. If B is in the NW section, their qualitative cardinal direction relation is “NW”; if B is in U, their direction relation is “U” (i.e. uncertain).

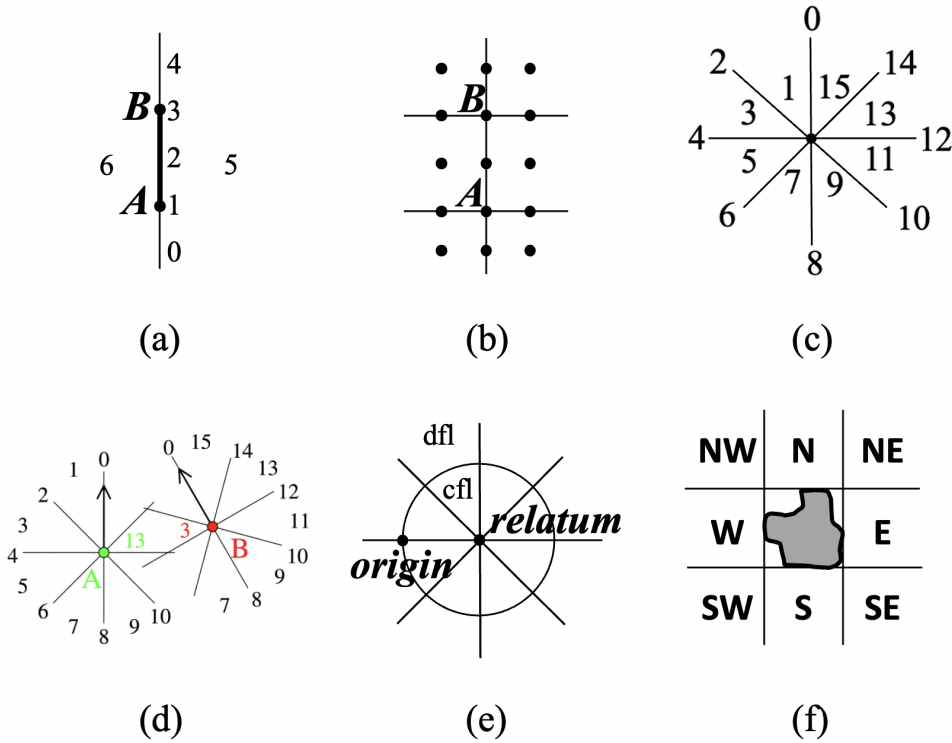
Figure 2.3: Cone-based and projection-based cardinal direction relations [Frank, 1991]. (a) 4 cones. (b) 8 cones. (c) One projection: north/south division. (d) One projection: west/east division. (e) Two projections: combining the north/south division with the west/east division. (f) Two projections with neural zone (U).



By contrast, relative direction relations are described in relative frames of reference. As shown in Figure 2.4 (a), in the Flipflop calculus [Ligozat, 1993], to describe the direction relation between a triple of objects A, B, and C, the space is divided into 6 parts: behind A (0), at A (1), between A and B (2), at B (3), in front of B (4), right of AB (5), and left of AB (6). The direction relation among A, B, and C is decided by where C locates given this space division. In this case, the frame of reference for direction relation deciding is relative because it is determined by Line AB, which is changeable given different A and B. The Double Cross Calculus (DCC) [Freksa, 1992b] extends the Flipflop calculus by adding two lines that go through A and B respectively and are vertical to AB so that these three lines form two crosses and the 2D plane is divided into 15 parts (Figure 2.4 (b)).

Other more recent qualitative direction calculi are designed with various space division strategies and object forms: $\mathcal{STAR}_m$ [Renz and Mitra, 2004] extends [Frank, 1991] by allowing the space to be divided into arbitrary number of sectors (Figure 2.4

Figure 2.4: (a) In the Flipflop calculus, the space is divided into 6 parts with AB as the baseline. (b) DCC divides the space into 15 sectors by 3 lines. (c) In $\mathcal{STAR}_4$, the 2D space is divided into 16 sectors by 4 lines. (d) $\mathcal{OPRA}_m$: Oriented point A is in relation of (3, 13) with oriented point B because A is in Sector 3 of the space division centered with B and B is in Sector 13 of that with A. (e) In TPCC, the frame of reference for direction relation is decided by the line from the relatum to the origin. Distance relation is also considered. For example, “cfl” is short for “close-front-left” while “dfl” is for “distant-front-left”. (f) In CDR, the space is divided into 9 parts, where the central tile is decided by the shape of the relatum object.



(c)); The Oriented Point Relation Algebra ($\mathcal{OPRA}_m$) [Moratz et al., 2005] models the relative direction relations between oriented points (Figure 2.4 (d)); The Ternary Point Configuration Calculus (TPCC) [Moratz and Ragni, 2008] describes ternary relative direction relations between points (Figure 2.4 (e)); and Cardinal Direction Relations (CDR) [Skiadopoulos and Koubarakis, 2004] focuses on cardinal direction relations between regions (Figure 2.4 (f)).

2.1.3 Distance, Motion, and Size

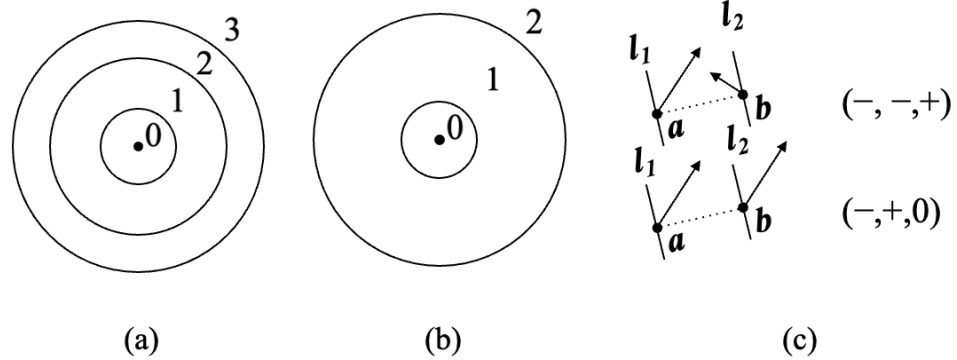
The Qualitative Distance Calculus (QDC) [Clementini et al., 1997] describes the qualitative distance relations between points. Similar to $\mathcal{STAR}_m$, QDC is also designed with arbitrary granularity: the number of space divisions is flexible. The span of each distance division is also configurable in various scenarios. Two possible con-

figurations of QDC are shown in Figure 2.5 (a) and (b). In real-world applications, sector indexes might correspond to various terms in natural language. For example, “0” might indicate “close” or “contact”.

The Qualitative Trajectory Calculus (QTC) [Van de Weghe et al., 2006] is able to describe the qualitative moving direction relations and speed relations between moving points. Figure 2.5 (c) shows an example of QTC_B , which is the basic form of QTC. There are also other more sophisticated versions of QTC. For example, QTC_C applies the Double Cross Calculus to describe moving directions and QTC_N is designed for movements in networks.

[Sridhar et al., 2011a] has presented three relative size relations “huge”, “large”, and “small” but in their paper, there is no further discussion on qualitative size relations. In some cases, the Point Algebra (PA) [Van Beek, 1992] is used to compare the relative size of objects. PA consists of three relations “>”, “<”, and “=”, which, in the scenario of size comparison, denotes the size of an object is greater than, smaller than or equal to that of another object [Gerevini and Renz, 2002a].

Figure 2.5: (a) One possible configuration of QDC_4 , where the space is divided into 4 sections based on the distance from the center point. (b) One possible configuration of QDC_3 , where the space is divided into 3 sections. (c) The arrows starting from a or b show the moving directions of a or b. Longer arrow indicates greater speed. (top) The first – means a is moving towards b; the second – indicates b is also moving towards a; the last + means a’s speed is greater than that of b. (bottom) The first – means a is moving towards b; the second + shows b is moving away from a; the last 0 means they move in the same speed.



2.1.4 Integrative Qualitative Spatial Representation

The majority of above introduced qualitative spatial calculi covers only one spatial aspect: topology, direction, distance, motion or size; and most above discussions also only concern one specific qualitative spatial calculi. However, there are also qualitative spatial calculi that cover multiple spatial aspects or discussions that are relevant to the conversions between different qualitative spatial calculi. For example,

[Isli et al., 2001] explore the conversions between cardinal direction relations and relative direction relations; in [Gerevini and Renz, 2002a], RCC-8 and qualitative size relations are integrated and formally discussed; [Li, 2007] first discusses the combination of topological and directional information for spatial reasoning, which is followed by [Liu et al., 2009; Li and Cohn, 2012; Cohn et al., 2014]; TPCC [Moratz and Ragni, 2008] is an integration of qualitative direction relations and qualitative distance relations; [Cohn et al., 2012] propose an integrative spatial representation formalism CORE-9 and argue their formalism is able to describe the topological, directional, size, and distance relations between objects; and [Clementini, 2013] discuss how to convert direction relations described in various frames of reference into those defined in a unified frame of reference.

[Renz and Nebel, 2007] point out that there are two ways of integrating two qualitative spatial calculi. Each qualitative spatial calculus can be regarded as the combination of its set of qualitative spatial relations and corresponding reasoning rules. The first calculi integration strategy is to form a new set of relations from the cross product of the two original sets of relations; the second strategy is to keep the original representation and develop new reasoning algorithms that consider the inter-dependences of relations.

Multiple qualitative spatial calculi are usually used together to describe real-world scenarios. More details please refer to Section 2.3.

2.2 Qualitative Spatial Reasoning

Reasoning is the process of deriving new knowledge from the given information [Cohn and Renz, 2008]. In this section, we will introduce common reasoning problems and techniques in the context of QSRR.

Figure 2.6: (a) Converse table of Point Algebra (PA). $r^\smile$ means the converse of r . (b) Composition table of PA. Source: [Dylla et al., 2017]

(a)	r	$r^\smile$	(b)	$r \setminus s$	$<$	$=$	$>$
	$<$	$>$		$<$	$\{<\}$	$\{<\}$	$\{<,,>\}$
	$=$	$=$		$=$	$\{<\}$	$\{=\}$	$\{>\}$
	$>$	$<$		$>$	$\{<,,>\}$	$\{>\}$	$\{>\}$

2.2.1 Converse Reasoning

If the relation between object A and B is R, then the relation between B and A is the converse of R. Given R, converse reasoning is to infer the converse of R. For most qualitative spatial calculi, it is not difficult to infer the converse of its relations given their definitions. Converse table is the table that stores all converse relations for a

specific calculus. As shown in Figure 2.6 (a), for Point Algebra [Van Beek, 1992], the converse of “<” is “>”; the converse of “=” is itself; and the converse of “>” is “<”.

A binary relation is a relation between a pair of objects. If R is a binary relation, then there is only one converse relation of R . For each relation among a triple of objects (i.e. ternary relation), according to [Zimmermann and Freksa, 1996; Scivos and Nebel, 2001], there are five converse relations, which are named as *inversion*, *short cut*, *short cut inverse*, *homing*, and *homing inverse* respectively. Let R_{abc} denotes the relation among object A , B and C , then its inversion is R_{bac} ; its short cut is R_{acb} ; its short cut inverse is R_{cab} , its homing is R_{bca} , and its homing inverse is R_{cba} .

2.2.2 Compositional Reasoning

Compositional reasoning is to infer a new relation by composing two relations [Renz and Nebel, 2007]. Let R_{ab} denote the relation between object A and B , and R_{bc} the relation between B and C , the composition of R_{ab} and R_{bc} is R_{ac} . Formally, $R_{ac} = R_{ab} \circ R_{bc}$, where $\circ$ denotes the composition operation. From the perspective of constraint-based reasoning, composition operation can be interpreted as inferring the constraints on object A and C from composing the constraints implied by R_{ab} with those implied by R_{bc} . Composition table is the table that stores all precomputed compositions. The composition table of the Point Algebra is shown in Figure 2.6 (b). Based on this table, the composition of “<” and “<” is “<”; and the composition of “<” and “>” is $\{<, >, =\}$, which means “< $\circ$ >” could be “<”, “>”, or “=”.

The composition table of RCC-8 is presented in [Randell et al., 1992a]. There are many different ways of composing two ternary relations. For example, R_{abd} could be inferred from the composition of R_{abc} and R_{acd} [Isli and Cohn, 2000]. Other meaningful ways of ternary relation compositions can be derived from this form by permutation of relations [Scivos and Nebel, 2001]. There are also discussions on the converse and composition of n -ary relations, which is out of the focus of this thesis. Readers interested at this topic can refer to [Condotta et al., 2006; Dylla et al., 2017] for more details.

2.2.3 Consistency Checking

In the field of QSRR, as discussed in [Renz and Nebel, 1999], the most important reasoning problem is to check whether a set of qualitative spatial relations is consistent (i.e. the consistency problem) or equivalently whether there is at least one set of objects that can satisfy the constraints implied by these relations. A basic assumption in QSRR is that domains of spatial objects are continuous and infinite. Hence it is impossible to solve the consistency problem by backtracking objects (which is the main technique for dealing with common constraint satisfaction problems (CSPs)). Most of the existing solutions to the consistency problem are local consistency-based: there are subsets that include all basic relations and for these subsets checking local consistency (e.g. path consistency or strong 4-consistency) is tractable and sufficient for consistency. This local consistency-based strategy is at least applicable to qual-

itative relations of the Interval Algebra ($\mathcal{IA}$) [Nebel and Bürckert, 1995], the Cardinal Direction Algebra ($\mathcal{CDA}$) [Ligozat, 1998], and the Region Connection Calculus ($\mathcal{RCC}$) [Renz, 1999].

However, local consistency-based methods are not suitable for many other categories of qualitative spatial relations (which means there are cases where relations are locally consistent but not consistent) and for them there is almost no solution for the consistency problem. For example, binary orientation relations [Isli and Cohn, 2000], fine-grained absolute direction relations [Renz and Mitra, 2004], and relative direction relations [Wolter and Lee, 2010]. A seemingly feasible strategy is to find at least one consistent initialization of the variables involved in the set of qualitative relations because once a solution is found it is sufficient to prove consistency. This strategy could be achieved by attempts in which entities are initialized with exact values. These attempts are straightforward and intuitive. As pointed out in [Wolter and Wallgrün, 2013], it is easier for people to understand a scenario with a quantitative example than with a qualitative description.

As mentioned earlier, two of the most famous qualitative temporal calculi are Point Algebra ($\mathcal{PA}$) [Vilain et al., 1990] and Interval Algebra ($\mathcal{IA}$) [Allen, 1990]. $\mathcal{PA}$ consists of three basic relations (i.e. “<”, “=”, and “>”) over the domain of 1D points [Vilain et al., 1990] while $\mathcal{IA}$ consists of 13 basic relations over the domain of 1D intervals. The consistency problem of $\mathcal{PA}$ is tractable while that of $\mathcal{IA}$ is intractable. However, the only tractable subset of $\mathcal{IA}$ that containing all basic relations has been found in [Nebel and Bürckert, 1995].

The consistency problem of 2D orientations with a fixed granularity (left, right, equal, and opposite) has been discussed in [Isli and Cohn, 2000]: for ternary orientation calculus $\mathcal{CVC}_t$ there are tractable subsets that including all basic relations so that the consistency problem can be solved by checking strong 4-consistency; while for binary orientation calculus $\mathcal{CVC}_b$ there might be no such tractable subset because path-consistency cannot detect inconsistencies even for basic relations of $\mathcal{CVC}_b$ [Isli and Cohn, 2000].

There are three categories of direction relations (absolute, intrinsic and relative) that are different from how the frame of reference is decided [Levinson, 1996]. Cardinal direction relations are defined in absolute frames of reference. As introduced earlier, the Star Calculus ($\mathcal{STAR}_m$) [Renz and Mitra, 2004] is designed to model cardinal direction relations [Frank, 1991] with adjustable granularity. According to [Renz and Mitra, 2004], $\mathcal{STAR}_1$ is tractable because a $\mathcal{STAR}_1$ relation can be represented by a pair of $\mathcal{PA}$ relations; $\mathcal{STAR}_2$ is actually equivalent to $\mathcal{CDA}$ so that its consistency problem can also be solved by enforcing path-consistency; and for finer-grained $\mathcal{STAR}_m$ with $m > 2$ the consistency problem cannot be solved by local consistency-based methods.

There are also many qualitative direction relation calculi that are defined in intrinsic or relative frames of reference (e.g. the Double Cross Calculus [Freksa, 1992c], $\mathcal{OPRA}_m$ [Moratz, 2006], and $\mathcal{SV}_m$ [Lee et al., 2013]). The consistency problem of intrinsic or relative direction relations remains unsolved as pointed out by a conjecture in [Wolter and Lee, 2010]: there is no polynomial time algorithm for the consistency

checking of direction relations involving left and right distinction.

The problem of across-calculi consistency checking (or the across-calculi consistency problem) is to check whether there is at least one consistent solution for two or more sets of qualitative spatial relations that come from different calculi. Two strategies of combining relations from two different calculi have been discussed in [Renz and Nebel, 2007]: one is to deal with new relations that originate from taking the cross product of the original two sets of relations; the other is to search for the correlations between the original two sets of relations. However, the disadvantage of the former strategy is that the number of basic relations is huge if the first strategy is used; and we have to find the correlations manually if the second strategy is applied.

There are also very few attempts on solving particular cases of this problem and both of the following two examples belong to the second strategy mentioned above: [Isli et al., 2001] provides an algorithm that can achieve an extent of consistency that is between local consistency for each sub-calculus and across-calculi consistency; the BI-PATH algorithm in [Gerevini and Renz, 2002b] is designed for a set of relations in one of the three maximal tractable sets of $\mathcal{RCC}$ -8 and a set of qualitative size relations. The BI-PATH algorithm seems promising but it might not be extendable to the across-calculi consistency problem of other calculi because it requires tractable consistency checking algorithm for each sub-calculus while as discussed in the previous subsection for many calculi there is no such tractable algorithm.

2.2.4 Neighborhood-based Reasoning

Two qualitative spatial relations are conceptual neighbors if (1) they are defined in the same qualitative spatial calculi; and (2) they can be transformed into one another with only one deformation [Freksa, 1992a]. For example, in the Point Algebra, “>” and “=” are conceptual neighbors while “>” and “<” are not because it is impossible that, as the size of one object changes, their size relation jumps from “>” to “<” (their relation should be “=” at certain intermediate stage). Conceptual neighborhood graph is a graph that consists of relation nodes and edges that connect conceptual neighbors. Figure 2.1 also presents the conceptual neighborhood graph of $\mathcal{RCC}$ -8, from which we can tell, for example, “=” is the conceptual neighbor of PO, TPP, TPPi, NTPP, and NTPPi. Conceptual neighborhood graph stores the conceptual closeness and similarity of relations, which can be used to analyze qualitative state transitions [Dylla, 2009] or to remove accidental relation changes when abstracting relations from noisy video [Sridhar et al., 2011b].

2.3 QSRR Applications

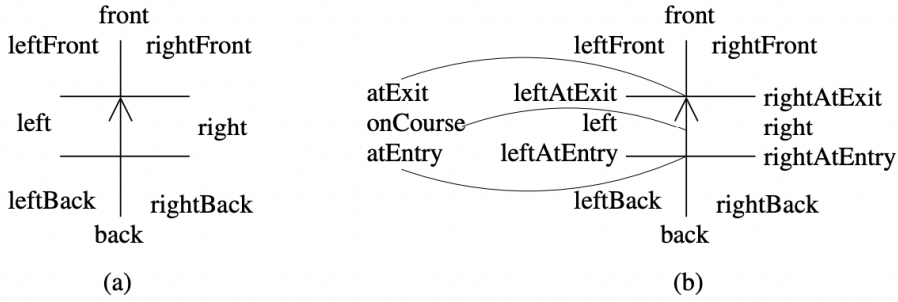
The main advantage of QSRR is that it is close to the way humans describe spatial scenarios and reason about spatial problems [Renz and Nebel, 2007]. We are not aiming to provide a complete list of previous QSRR applications. Instead, we choose to describe typical QSRR applications in navigation and action description so that readers understand the usefulness of QSRR in these two domains.

2.3.1 Navigation

In this subsection, we will introduce QSRR applications in navigation [Perico et al., 2021], which is one of the natural QSRR application domains claimed by [Wolter and Wallgrün, 2013].

Direction description is essential in navigation. Qualitative direction relations are useful in human machine interaction because they abstract away unnecessary numeric details and are interchangeable with natural language terms. For example, as shown in Figure 2.7, the Double Cross Calculus [Freksa, 1992c] is applied to describe direction relations so that it is possible to fulfill the communication between a driver and an intelligent wheelchair [Krieg-Brückner and Shi, 2006].

Figure 2.7: (a) A basic version of the correspondence between the Double Cross Calculus relations and their natural language interpretations. (b) An extended version of such correspondence. Source: [Krieg-Brückner and Shi, 2006].

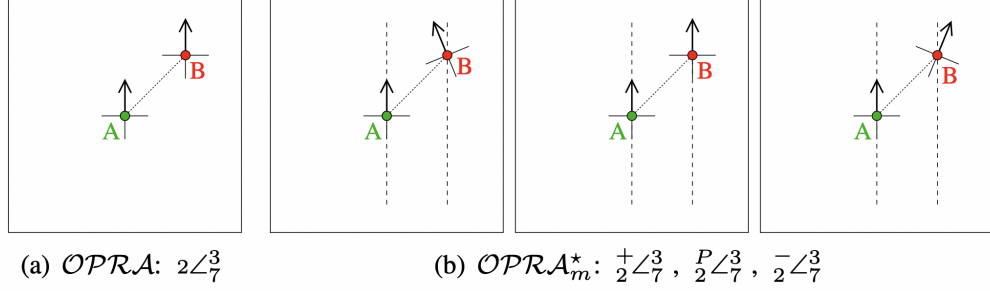


In [Fogliaroni et al., 2009], agents are assumed with limited sensor abilities and the authors aim to solve the problem of qualitative localization and navigation. Specifically, they propose a model that considers the qualitative visibility and occlusion relations between extended objects; divide space into different regions according to their visibility and occlusion relations with existing objects; and construct a topological map that records the adjacency relations between these divided regions. Localization is fulfilled by identifying which region the agent is in via local observations; and navigation is realized by a graph search of the next satisfying region.

[Dylla, 2009] utilizes an extended version of $OPRA_m$ to express common sea navigation scenarios (Figure 2.8); proposes the action-augmented conceptual neighborhood (ACN) to analyze critical sea navigation state transitions; and applies consistency checking in ensuring rule-compliant agent behaviors. $OPRA_m$ is applied in a different scenario in [Lücke et al., 2011]. In their paper, a street network is described in the form of a set of $OPRA_s$ relations via local observations. They conduct experiments to show their method is comparable to another qualitative navigation paradigm [Klippel et al., 2005].

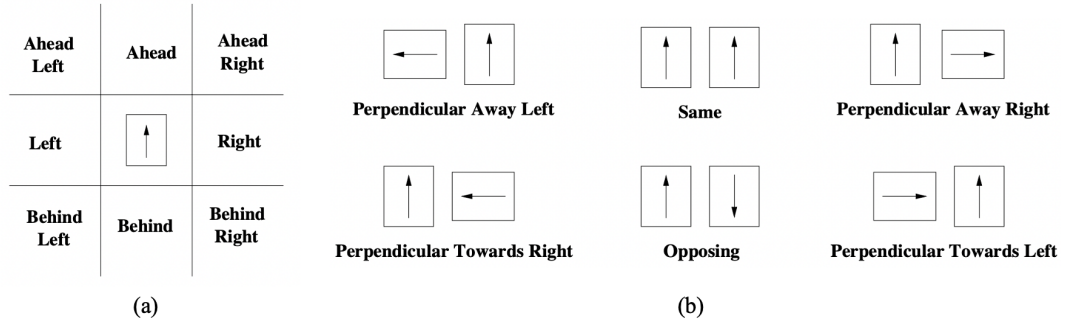
For navigation problems, there are at least two main sources of qualitative direction relations: first, quantitative data in geographical information systems (e.g. OpenStreetMap data) as pointed out by [Lücke et al., 2011]; and second, real-world

Figure 2.8: (a) $\mathcal{OPRA}_2$ is applied to described two moving boats. (b) $\mathcal{OPRA}_2$ is extended so that it is possible to express two boats are moving closer (+), moving in parallel (P), or moving away from each other ($-$). Source: [Dylla, 2009].



natural language place relation descriptions [Vasardani et al., 2013; Kordjamshidi et al., 2013; Liu et al., 2014; Chen et al., 2018]. Given large-scale real-world data, it is impossible to manually abstract qualitative spatial relations from quantitative data. Over the past twenty years, a couple of QSRR tools has been implemented. For example, SparQ [Dylla et al., 2006], the Qualitative Algebra Toolkit (QAT) [Condotta et al., 2006], Generic Qualitative Reasoner (GQR) [Westphal et al., 2009], and QSRLib [Gatsoulis et al., 2016]. Based on these tools, automatic qualitative spatial relation extraction is now possible.

Figure 2.9: Two predefined sets of qualitative spatial relations to describe moving vehicle relations [Ferryhough et al., 2000]. (a) Qualitative relative position relations. (b) Qualitative relative motion direction relations.



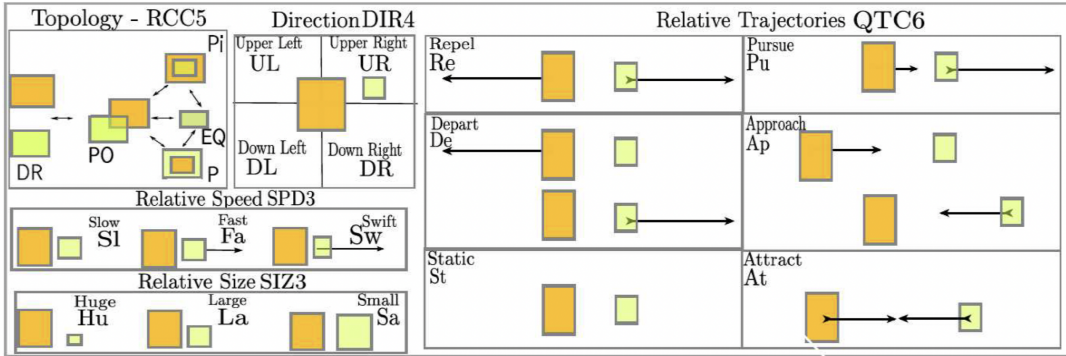
2.3.2 Action Description

There has been a long history of applying qualitative spatial relations to describe action state transitions. [Ferryhough et al., 2000] predefine a set of qualitative relative position relations (Figure 2.9 (a)) and a set of qualitative relative motion direction relations (Figure 2.9 (b)) to describe the relations between vehicles in realistic traffic scenarios. Traffic actions like following or overtaking are learnt as qualitative

spatio-temporal models, which are series of qualitative scenario descriptions and each scenario description is actually the set of qualitative spatial relations between detected vehicle contours. A developed solution is proposed in [Galata et al., 2002], where the set of qualitative spatial relations is not predefined but learnt by clustering numeric direction, velocity, and distance relations between vehicles. Similar relation learning technique is applied in [Bleser et al., 2015] to model industrial assembly and production workflows.

[Sridhar et al., 2010; Duckworth et al., 2019] work on the problem of unsupervised action clustering. In [Sridhar et al., 2010], RCC-5 [Randell et al., 1992b] is applied to describe object interactions and the Interval Algebra (IA) [Allen, 1990] is used to record the temporal relations between these interactions. Unsupervised learning is realized by a generative probabilistic procedure. They claim that their approach is able to find meaningful actions, which is supported by the results of experiments on real data. In contrast, [Duckworth et al., 2019] use QDC [Clementini et al., 1997], TPCC [Moratz and Ragni, 2008], and QTC [Van de Weghe et al., 2006] to describe interactions between human body and objects. Latent Dirichlet Allocation [Blei et al., 2003] is applied to recover semantically-meaningful actions. They evaluate their approach on two datasets and prove its performance is comparable to supervised SVM.

Figure 2.10: The qualitative action descriptions in [Sridhar et al., 2011a] consider five sets of qualitative spatial relations: topological relations; direction relations; speed relations; size relations; and moving trajectory relations.



Such representation technique has also been applied in [Sridhar et al., 2011a] to conduct a series of benchmarking experiments on video action classification, detection, and clustering. As shown in Figure 2.10, their qualitative action description is based on predefined qualitative spatial relations in five aspects: topology, direction, speed, size, and motion. The result from an ablation study shows that the most inclusive representation (i.e. the one that considers all five spatial aspects) always perform the best. [Tayyub et al., 2014] propose three categories of qualitative action representations: qualitative temporal, qualitative spatial, and quantitative spatial for daily action recognition. Their qualitative spatial representation is based on counting the occurrence of various chains of RCC-3 relations between objects. Object relation

chains are like, for example, o_1-o_2-dc , $o_2-o_3-dc-po$, and $o_1-o_3-dc-po-pp$, where o_1, o_2, o_3 are objects and dc, po, pp are RCC-3 relations. Similar to the results in [Sridhar et al., 2011a], the best recognition accuracy is also achieved by combining these three forms of representations.

[Dubba et al., 2015] propose an inductive logic programming framework that learns the best descriptions for each action in a supervised manner. For each action, there is a set of positive and negative examples (namely videos that are relevant or irrelevant to a specific action). Similar to [Sridhar et al., 2010], action descriptions are also based on qualitative temporal and spatial relations. RCC-5 relations are manually chosen to describe object interactions in airport scenarios. Actions like airplane arrival or departure are represented as changes in the RCC relations between airplanes, vehicles, and ground zones. QTC relations are used to describe actions in a dataset that consists of short videos of action verbs like approach, exchange, or jump. These learnt qualitative action descriptions are then used to recognize actions in unseen videos.

[Young and Hawes, 2015] present an approach that allows agents to learn from observing experts. Their knowledge presentation is based on four qualitative spatial calculi RCC, *STAR*, QDC, and QTC. Specifically, the qualitative movement states of agents are described by the qualitative topological relations, direction relations, distance relations, and trajectory relations. Qualitative action representations are combined with machine learning methods to learn RoboCup soccer action like kick or dash. According to their experiment results, their approach achieves a high degree of accuracy.

Qualitative Direction Relation Integration

"Your direction is more important than your speed." – Richard L. Evans

Direction relations are some of the most commonly used spatial relations in human communication. A number of formalisms have been proposed to capture direction information, but all of them suffer from the same problem: people use direction relations as if they are binary relations (e.g. A is to the left of B) when they are in fact ternary relations (e.g. A is to the left of B when viewed from C). This implicit third piece of information that is required, but typically missing, is often referred to as the frame of reference (FoR). Given only binary direction relations without knowing the FoR means that we cannot do proper spatial reasoning (e.g. A is to the left of B and A is to the right of B are both consistent), we cannot integrate direction information from different sources, not even from the same source, and it can be difficult to understand what exactly someone means.

In this chapter, we will provide 1) the spatial constraint language $\mathcal{DAF}_m$ that can represent and compose direction relations across different FoRs; and 2) the foundations for deciding the overall consistency. We hope our model of representing and reasoning can bring research on direction relations to another level and will finally make it possible to properly use them in a comprehensive way.

3.1 Introduction

Representing spatial information and being able to reason about it is crucial in the field of Artificial Intelligence [Cohn and Renz, 2008]. Direction relations are very common in practical spatial descriptions and therefore have been studied extensively in the last two decades. Numerous formalisms have been created to model and reason about direction relations [Frank, 1991; Freksa, 1992b; Ligozat, 1993; Isli et al., 2001; Renz and Mitra, 2004; Clementini and Billen, 2006; Moratz, 2006; Moratz and Ragni, 2008; Lee et al., 2013]. They are defined in different categories of **frames of reference (FoR)**, which provides a coordination system [Frank, 1998] based on which direction relations (and distance relations [Moratz and Ragni, 2008]) can be

described. Several different groups of terms have been used to represent the roles of spatial entities in direction relation descriptions. For example, **figure**, **ground**, and **viewpoint** [Levinson, 1996]; **locatum**, **relatum** [Tenbrink and Kuhn, 2011], and **observer**. The latter group of terms are used throughout the chapter, where locatum is the entity that is to be located by relating to the position of a relatum; observers provide perspectives and they might or might not appear in a description [Tenbrink and Kuhn, 2011]. According to Levinson's taxonomy [Levinson, 1996], there are three categories of FoRs (**absolute**, **intrinsic**, and **relative**). Their difference can be explained with a practical route description example "go south at the next intersection until you see a bus station to your left; facing to the bus station you can see the library is to the right of the station" (Figure 3.1(a)):

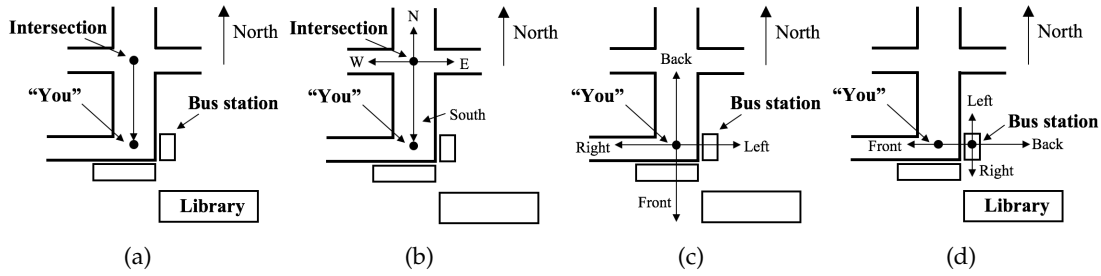


Figure 3.1: (a) Start at the intersection and then go south until the library is located by relating to the bus station. (b) Absolute FoR: based on north, the direction "you" go is south. (c) Intrinsic FoR: based on the direction "you" are facing to the bus station is to the left of "you". (d) Relative FoR: based on the direction from the bus station to "you" the library is to the right of the station.

- In absolute FoRs, an invariant direction [Levinson, 1996] (e.g. north) is used as the reference. For example, the direction "you" (relatum) go is "south" because it is opposite to north (Figure 3.1(b)). Examples of formalisms defined in absolute FoRs are Cardinal Direction Algebra ($CD\mathcal{A}$) [Frank, 1991] and $STAR_m$ [Renz and Mitra, 2004].
- In intrinsic FoRs, direction relations among spatial entities are described according to the "inherent features" [Levinson, 1996] of relatums. For example, "a bus station" (locatum) is "to your left" based on the direction "you" (relatum) are facing to (Figure 3.1(c)). Examples of formalisms defined in intrinsic FoRs are Oriented Point Relation Algebra ($OPRA_m$) [Moratz, 2006]; and SV_m [Lee et al., 2013].
- In relative FoRs, the line that connects the observer and the relatum is applied as the baseline. For example, "the library" (locatum) is described as "to the right of the station" based on the fact that "you" (observer) look at "the station" (relatum) (Figure 3.1(d)). Examples of formalisms defined in relative FoRs are Single Cross & Double Cross [Freksa, 1992b]; flipflop calculus [Ligozat, 1993]; and Ternary Point Configuration Calculus (TPCC) [Moratz and Ragni, 2008].

Most existing direction relation formalisms are defined in a specific category of FoR (absolute, intrinsic or relative). However, in practice direction relations could be expressed in different categories of FoRs (e.g. the above route description example), which means that by using any single existing formalism it is almost impossible to express, integrate and reason about direction information from different sources. This problem is critical because for example a set of direction relations that is consistent in absolute FoRs (Figure 3.2(a)) might turn out to be invalid when combining with another set of direction relations described in relative FoRs (Figure 3.2(b)).



Figure 3.2: (a) "b is to the north of a and c is to the north of b", which is inconsistent with (b) "from a's perspective, c is in front of b", because based on the former, b is between a and c while based on latter b is not.

To solve this problem, in this chapter we provide 1) an inclusive representation that can express and integrate direction information across different FoRs; 2) methods of across-FoR compositions; and 3) the foundations for deciding the overall consistency.

3.2 Related work

Direction relations are very common and numerous formalisms have been created to model and reason about direction relations in the last two decades. For example, cone-based and projection-based approaches for cardinal directions [Frank, 1991]; the flipflop calculus [Ligozat, 1993]; cardinal direction relations with arbitrary granularity ($STAR_m$) [Renz and Mitra, 2004]; relative direction relations between oriented points ($OPRA_m$) [Moratz, 2006]; ternary relative direction relations between points (TPCC) [Moratz and Ragni, 2008]; and observer orientation-based relative directions (SV_m) [Lee et al., 2013].

However, previous research on the problem of representing and reasoning about direction relations across different FoRs is limited and we are the first to solve this problem in a general sense. For example, in [Hernandez, 1994] Hernandez described how to transform direction relations in intrinsic, extrinsic, and deictic FoRs into implicit FoRs; in [Clementini, 2013] Clementini provides a framework that can convert relations in different FoRs into his 5-intersection model. However, in both of these two works across-category composition and reasoning have not been discussed. [Kreutzmann and Wolter, 2014] focuses on unifying computations in the form of linear programming, while symbolic representation and reasoning are not unified in their work.

Our model could be the most inclusive direction relation formalism so far because

it is compatible to direction information expressed in most of the previous direction relation formalisms. For example, $STAR_m$ [Renz and Mitra, 2004], $OPRA_m$ [Moratz, 2006] and TPCC [Moratz and Ragni, 2008] can be regarded as special cases of our model because similar space division strategy is applied while each of them just cover one category of FoRs.

3.3 Representation

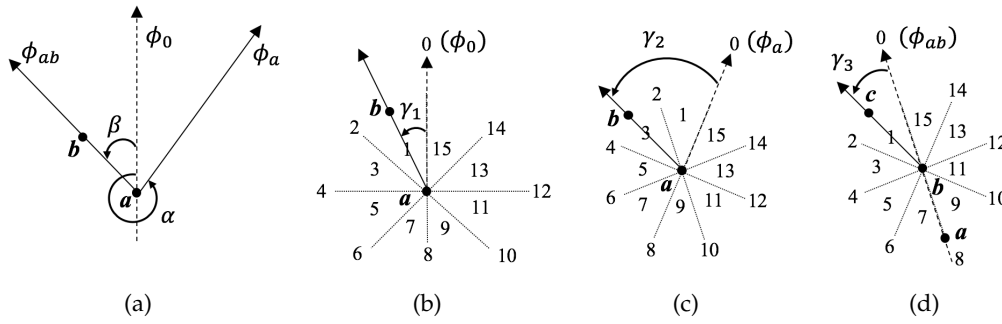


Figure 3.3: (a) The value of an intrinsic anchor ϕ_a or a relative anchor ϕ_{ab} is decided by the value of the anticlockwise angle deviated from the absolute anchor ϕ_0 (α for ϕ_a , β for ϕ_{ab}). (b) In an absolute FoR, the 2-D plane is divided into 16 sectors with the relatum a as the centroid and the absolute anchor ϕ_0 decides the direction of Sector 0. The absolute direction relation between a and b is decided by the value of the angle that is formed by vector ab anticlockwise deviating from ϕ_0 . (c) In an intrinsic FoR, the 2-D plane is divided into 16 sectors with the relatum a as the centroid and the intrinsic anchor ϕ_a (i.e. a 's orientation) decides the direction of Sector 0. The intrinsic direction relation between a and b is decided by the value of the angle that is formed by vector ab anticlockwise deviating from ϕ_a . (d) In a relative FoR, the 2-D plane is divided into 16 sectors with the relatum b as the centroid and the relative anchor ϕ_{ab} (i.e. the orientation from the observer a to the relatum b) decides the direction of Sector 0. The relative direction relation between a , b and c is decided by the value of the angle that is formed by vector bc anticlockwise deviating from ϕ_{ab} .

3.3.1 Spatial entities

Spatial entities can be modeled as points e.g. in [Frank, 1991; Ligozat, 1993; Isli et al., 2001; Renz and Mitra, 2004; Moratz, 2006; Moratz and Ragni, 2008; Lee et al., 2013]; or as regions e.g. in [Randell et al., 1992b; Clementini and Billen, 2006]. Point representation of spatial entities is more popular for theoretical analysis because irrelevant features could be ignored while regions are more close to real-life spatial objects. In our model, spatial entities are oriented points [Mossakowski and Moratz, 2012] in 2-D Euclidean plane: given an oriented point $a = (x_a, y_a, \phi_a) \in D_{op}$ where

$D_{op} = \mathbb{R}^2 \times \Theta$ and $\Theta = [0, 2\pi)$, $(x_a, y_a) \in \mathbb{R}^2$ decides the position of a and the value of $\phi_a \in \Theta$ decides the orientation of a .

3.3.2 Frames of reference

As described in the introduction FoRs can be absolute, intrinsic, or relative (which is also called as the **category** of a FoR). Each FoR consists of three elements: the centroid (which is usually the relatum), the **anchor** and sectors. The anchor is the orientation of a FoR [Frank, 1998; Levinson, 1996] and the baseline from which the direction relation between a locatum and a relatum is determined. Since there are three categories of FoR, correspondingly absolute/intrinsic/relative anchors can be defined in Definition 1.

Definition 1 (Absolute/Intrinsic/Relative anchor). *The absolute/intrinsic/relative anchor is the anchor used in absolute/intrinsic/relative FoRs to decide the direction relations between oriented points.*

The absolute anchor is fixed and can be expressed as the orientation of a fake oriented point $\mathbf{0}$ whose position is unknown and its orientation is $\phi_0 = 0$.

The intrinsic anchor is the orientation (or the intrinsic front) of a relatum. For example, given a relatum a , its orientation ϕ_a might be used as an intrinsic anchor.

The relative anchor is the perspective decided by the observer looking into the relatum. For example, given an observer a and a relatum b , the orientation from a to b can be used as a relative anchor (denoted as ϕ_{ab}).

The value of an intrinsic or relative anchor is decided by the value of the anticlockwise angle deviated from the absolute anchor (e.g. $\phi_a = \alpha$ and $\phi_{ab} = \beta$ in Figure 3.3(a)). All angle values throughout this chapter are by default taken modulo 2π . Sectors are decided by the strategy of space division and will be explained in the following section.

3.3.3 Space division and sectors

Our space division model is similar to $\mathcal{OPRA}_m$ [Moratz et al., 2005; Mossakowski and Moratz, 2012]: given a certain anchor, the 2-D plane is divided into $4m$ sectors by m straight lines ($m \in \mathbb{N}$). The value of m decides the granularity of space division [Lee et al., 2013]. Three $m = 4$ examples are in Figure 3.3(b)-3.3(d). These sectors are indexed anticlockwise from 0 to $4m - 1$. The anchor of a FoR is in the same direction as Sector 0. Sectors with even indexes are linear sectors (also called as even sectors) while sectors with odd indexes are planar sectors (also called as odd sectors) [Moratz et al., 2005]. All results of integer operations in the remainder of this chapter are by default taken modulo $4m$. The domain of Sector i (D_i) is defined as follows [Mossakowski and Moratz, 2012]:

$$D_i = \begin{cases} \frac{2\pi i}{4m}, & \text{if } i \text{ is even;} \\ (\frac{2\pi(i-1)}{4m}, \frac{2\pi(i+1)}{4m}), & \text{if } i \text{ is odd.} \end{cases}$$

For example, in Figure 3.3(b) $D_0 = 0$ and $D_1 = (0, \frac{\pi}{4})$.

3.3.4 The spatial constraint language $\mathcal{DAF}_m$

Direction relations Across different Frames of references (or $\mathcal{DAF}_m$) is an inclusive constraint language that can express and integrate direction information across different FoRs in a general sense. It is defined over the domain of oriented points (D_{op}) and m is the granularity parameter [Lee et al., 2013]. Its $16m + 5$ basic relations consist of 5 same position relations, $4m$ absolute direction relations, $4m$ intrinsic direction relations, $4m$ relative direction relations and $4m$ anchor relations, which are defined as follows.

Same position relations are defined before direction relations because it is almost meaningless to discuss direction relations if spatial entities are in the same position. **Binary same position relation** (Definition 2) can be regarded as a special case of direction descriptions in absolute or intrinsic FoRs that are related to two spatial entities. **Ternary same position relations** (Definition 3) should also be defined to correspond to direction descriptions in relative FoRs that cover three spatial entities.

Definition 2 (Binary same position relation). *Given two oriented points $a, b \in D_{op}$, they are in the binary same position relation (denoted as S_{ab}^2) iff $x_a = x_b \wedge y_a = y_b$. The negation of S_{ab}^2 is denoted as $\neg S_{ab}^2$.*

Definition 3 (Ternary same position relations). *Given three oriented points a, b and $c \in D_{op}$, they are in the ternary same position relation (denoted as S_{abc}^3) iff S_{ab}^2 and S_{bc}^2 .*

A degenerate case of ternary same position relation is ternary 12 same position relation (i.e. only the first two oriented points are in the same position). It is denoted as S_{abc}^{12} iff S_{ab}^2 and $\neg S_{bc}^2$.

Similarly, they are in the ternary 23 same position relation (denoted as S_{abc}^{23}) iff $\neg S_{ab}^2$ and S_{bc}^2 ; they are in the ternary 13 same position relation (denoted as S_{abc}^{13}) iff S_{ac}^2 and $\neg S_{bc}^2$.

In a certain FoR, direction relations between oriented points are determined by the value of the anticlockwise included angle between the anchor and the vector from the relatum to the locatum (e.g. γ_1 in Figure 3.3(b), γ_2 in Figure 3.3(c), and γ_3 in Figure 3.3(d)). Specifically, direction relations in absolute FoRs (Definition 4) are always determined according to the absolute anchor ϕ_0 (Figure 3.3(b)); direction relations in intrinsic FoRs (Definition 5) are decided by the orientation of the relatum (e.g. ϕ_a in Figure 3.3(c)); while in relative FoRs direction relations (Definition 6) are determined by the perspective of the observer looking into the relatum (e.g. ϕ_{ab} in Figure 3.3(d)).

Definition 4 (Absolute direction relations). *Given $a, b \in D_{op}$ and $i \in \{0, 1, \dots, 4m - 1\}$, a and b are in an absolute direction relation i^A (denoted as A_{ab}^i) iff $\phi_{ab} - \phi_0 \in D_i$ and $\neg S_{ab}^2$. Note that ϕ_{ab} represents the value of the included angle between the vector from a to b and ϕ_0 .*

Definition 5 (Intrinsic direction relations). *Given $a, b \in D_{op}$ and $i \in \{0, 1, \dots, 4m - 1\}$, a and b are in an intrinsic direction relation i^I (denoted as I_{ab}^i) iff $\phi_{ab} - \phi_a \in D_i$ and $\neg S_{ab}^2$.*

Definition 6 (Relative direction relations). *Given $a, b, c \in D_{op}$ and $i \in \{0, 1, \dots, 4m - 1\}$, a, b, c are in a relative direction relation i^R (denoted as R_{abc}^i) iff $\phi_{bc} - \phi_{ab} \in D_i$ and $\neg S_{ab}^2$ and $\neg S_{bc}^2$.*

For example, A_{ab}^1 means in the absolute FoR a and b are in the absolute direction relation 1^A (Figure 3.3(b)); in Figure 3.3(c) from a 's point of view b is in Sector 3, which can be denoted as I_{ab}^3 ; while in Figure 3.3(d), R_{abc}^1 can be used to indicate from a 's point of view c is in Sector 1 of b .

With these definitions, we can easily show that

Lemma 1. *The basic relations of $\mathcal{DAF}_m$ are not JEPD.*

It can be proved by the fact that S_{abc}^{13} is a subset of R_{abc}^{2m} . For example, in Figure 3.3(d) if c is removed to the same position as a and b remains unchanged, it is not difficult to conclude that S_{abc}^{13} is a special case of R_{abc}^8 .

Except for the same position relations and direction relations, the relations between anchors should also be modeled because as will be seen in the following sections it is the prerequisite of reasoning about direction relations across different FoRs. To be uniform with the relations defined above, the relations between anchors are also expressed by the relations between oriented points (Definition 7).

Definition 7 (Anchor relations). *Given $a, b \in D_{op}$ and $i \in \{0, 1, \dots, 4m - 1\}$, a and b are in an anchor relation i^Φ (denoted as Φ_{ab}^i) iff $\phi_b - \phi_a \in D_i$.*

Φ_{a0}^i denotes that an intrinsic anchor ϕ_a and the absolute anchor ϕ_0 are in the anchor relation i^Φ (recall that absolute anchor is the orientation of a fake oriented point 0). A relative anchor relates to two oriented points. To express its relation with other anchors, a dedicated new oriented point with the same orientation as the relative anchor can be created. For example, given a relative anchor ϕ_{ab} that is decided by an observer a and a relatum b , we can create a new oriented point b' whose orientation is $\phi_{b'} = \phi_{ab}$.

A set of relations can indicate the indeterminacy about the relation between oriented points. For example, the absolute direction relation between a and b can be $\{0^A, 2^A\}$ (denoted as $A_{ab}^{\{0,2\}}$), which means the absolute direction relation between a and b can be 0^A or 2^A . Note that $\{1^c, \dots, 5^c\}$ is different from $\{5^c, \dots, 1^c\}$ where $C \in \{\mathcal{A}, \mathcal{I}, \mathcal{R}, \phi\}$ indicates relation category. The former means $\{1^c, 2^c, 3^c, 4^c, 5^c\}$ while the latter means $\{5^c, \dots, (4m - 1)^c, 0^c, 1^c\}$. **Non-informative** (or **universal**) absolute direction relation is a special case of sets of absolute direction relations and is defined as follows. Non-informative intrinsic/relative direction relations and anchor relations are defined analogously.

Definition 8 (Non-informative/universal absolute direction relations). *Two oriented points a and b are in the non-informative/universal absolute direction relation U^A (denoted as A_{ab}^U) if there is no information about their relation in absolute FoRs. Namely $U^A = \{0^A, 1^A, \dots, (4m - 1)^A\}$ or $A_{ab}^U = A_{ab}^{\{0,1,\dots,4m-1\}}$.*

Each **instance** of $\mathcal{DAF}_m(\Theta)$ consists of a finite set of same position relations, absolute/intrinsic/relative direction relations, and anchor relations over the domain of oriented points.

The converse and composition of binary and ternary same position relations are trivial by definition. For example,

1. S_{ab}^2 iff S_{ba}^2 .
2. $(S_{ab}^2 \text{ and } S_{bc}^2) \Rightarrow S_{ac}^2$.
3. $S_{abc}^{12} \Rightarrow S_{cab}^{23}$.
4. S_{abc}^3 iff S_{*}^3 , where $*$ is the permutations of abc .

However, converse and composition of direction relations and anchor relations should be defined according to corresponding relation semantics [Renz and Mitra, 2004] as will be introduced in the following sections.

3.4 Converse of Direction Relations

In this section the converse of absolute/intrinsic/relative direction relations will be discussed separately.

Proposition 1 (Converse of absolute direction relations). *Given two oriented points $a, b \in D_{op}$ and their absolute direction relation A_{ab}^i where $i \in \{0, 1, \dots, 4m - 1\}$, the converse of A_{ab}^i is $A_{ba}^{(i+2m)}$ (denoted as $A_{ab}^{i\smile}$).*

Proof. By definition, A_{ab}^i iff $\phi_{ab} - \phi_0 \in D_i$. Assume A_{ba}^j then $\phi_{ba} - \phi_0 \in D_j$. Because $\phi_{ba} = \phi_{ab} + \pi$ (e.g. $\phi_{ab} = \alpha$ and $\phi_{ba} = \beta = \alpha + \pi$ in Figure 3.4(a)), $\phi_{ba} - \phi_0 = \phi_{ab} - \phi_0 + \pi \in D_i + \pi$. Then it is not difficult to infer $j = i + 2m$ based on the definition of sector domains. Note that angle values throughout this chapter are by default taken modulo 2π and integer operations are by default taken modulo $4m$. $\square$

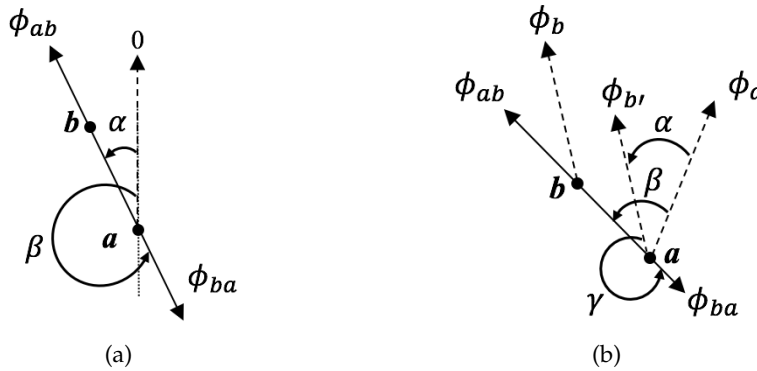


Figure 3.4: (a) $\beta = \alpha + \pi$ where $\alpha = \phi_{ab}$ and $\beta = \phi_{ba}$ (b) $\gamma = \beta + \pi - \alpha$ where $\alpha = \phi_b - \phi_a$, $\beta = \phi_{ab} - \phi_a$, and $\gamma = \phi_{ba} - \phi_b$.

Proposition 2 (Converse of intrinsic direction relations). *Given two oriented points $a, b \in D_{op}$, their intrinsic direction relation I_{ab}^i and their anchor relation Φ_{ab}^j ($i, j \in \{0, 1, \dots, 4m - 1\}$), the converse of I_{ab}^i is I_{ba}^k (denoted as $I_{ab}^{i\smile}$) where $k = \Omega_{i,j,i+2m-j}$ and $\Omega_{x,y,z} =$*

$$\begin{cases} z, & \text{if } x \text{ or } y \text{ are even;} \\ \{z-1, z, z+1\}, & \text{if } x \text{ and } y \text{ are odd.} \end{cases}$$

Proof. By definition, I_{ab}^i iff $\phi_{ab} - \phi_a \in D_i$ and Φ_{ab}^j iff $\phi_b - \phi_a \in D_j$. Assume I_{ba}^k then $\phi_{ba} - \phi_b \in D_k$. Since $\phi_{ba} - \phi_b = (\phi_{ab} + \pi) - (\phi_b - \phi_a + \phi_a) = (\phi_{ab} - \phi_a) + \pi - (\phi_b - \phi_a)$, $\phi_{ba} - \phi_b \in D_i + \pi - D_j$. For example, in Figure 3.4(b), $\alpha = \phi_b - \phi_a$, $\beta = \phi_{ab} - \phi_a$, $\gamma = \phi_{ba} - \phi_b$ and $\gamma = \beta + \pi - \alpha$.

Let $z = i + 2m - j$. If i and j are both even, $D_i + \pi - D_j = \frac{2\pi z}{4m}$ then $k = z$; If one of i and j is even and the other is odd, $D_i + \pi - D_j = (\frac{2\pi(z-1)}{4m}, \frac{2\pi(z+1)}{4m})$ then k is also equal to z ; If i and j are both odd, $D_i + \pi - D_j = (\frac{2\pi(z-2)}{4m}, \frac{2\pi(z+2)}{4m})$ then $k = \{z-1, z, z+1\}$. $\square$

Different from absolute/intrinsic direction relations, relative direction relations are ternary relations. According to [Zimmermann and Freksa, 1996; Scivos and Nebel, 2001], there are five converse relations for each ternary relation.

Proposition 3 (Converse of relative direction relations). *Given three oriented points $a, b, c \in D_{op}$ that are in different positions and their relative direction relation R_{abc}^i where $i \in \{0, 1, \dots, 4m-1\}$, following the terminology in [Zimmermann and Freksa, 1996],*

1. **Inversion** of R_{abc}^i is $R_{abc}^{i \sim inv} = R_{bac}^j$ where $j =$

$$\begin{cases} 2m, & \text{if } i = 0; \\ \{0, 2m\}, & \text{if } i = 2m; \\ \{2m+1, \dots, i+2m-\Delta_i\}, & \text{if } 0 < i < 2m; \\ \{i-2m+\Delta_i, \dots, 2m-1\}, & \text{if } 2m < i < 4m. \end{cases}$$

where Δ_i is defined as

$$\Delta_i = \begin{cases} 1, & \text{if } i \text{ is even;} \\ 0, & \text{if } i \text{ is odd.} \end{cases}$$

2. **Short cut** $R_{abc}^{i \sim sc} = R_{acb}^j$. Interestingly, the transition result of $R_{abc}^{i \sim sc}$ are the same as that of $R_{abc}^{i \sim inv}$.

3. **Short cut inverse** $R_{abc}^{i \sim sci} = R_{cab}^j$ if R_{bac}^{-j} .

4. **Homing** $R_{abc}^{i \sim hm} = R_{bca}^j$ if R_{acb}^{-j} .

5. **Homing inverse** $R_{abc}^{i \sim hmi} = R_{cba}^{-i}$.

Proof. It is trivial to prove Rule 5 because by definition R_{abc}^i iff $\phi_{bc} - \phi_{ab} \in D_i$, R_{cba}^j iff $\phi_{ba} - \phi_{cb} \in D_j$ and $\phi_{bc} - \phi_{ab} = (\phi_{cb} + \pi) - (\phi_{ba} + \pi) = -(\phi_{ba} - \phi_{cb})$. Rule 3 and 4 can be proved analogously.

For Rule 1, by definition R_{abc}^i iff $\phi_{bc} - \phi_{ab} \in D_i$, R_{bac}^j iff $\phi_{ac} - \phi_{ba} \in D_j$. If $i = 0$ or $2m$ the proof is trivial. If $0 < i < 2m$, based on the geometric properties of triangle abc (Figure 3.5(a)), $0 < \beta < \alpha$ where $\alpha = \phi_{bc} - \phi_{ab}$ and $\beta = \phi_{ac} - \phi_{ba} - \pi$. Then

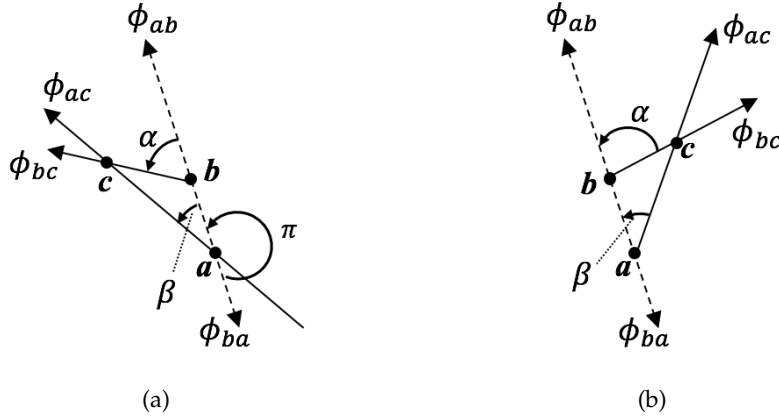


Figure 3.5: (a) when R_{abc}^i and $0 < i < 2m$, $0 < \beta < \alpha$ where $\alpha = \phi_{bc} - \phi_{ab}$ and $\beta = \phi_{ac} - \phi_{ba} - \pi$. (b) when R_{abc}^i and $2m < i < 4m$, $0 < \beta < \alpha$ where $\alpha = 2\pi - (\phi_{bc} - \phi_{ab})$ and $\beta = \pi - (\phi_{ac} - \phi_{ba})$.

$\phi_{ac} - \phi_{ba} \in (\pi, \phi_{bc} - \phi_{ab} + \pi)$. If i is even, $D_j = (\frac{2\pi(2m)}{4m}, \frac{2\pi((i+2m-1)-1)}{4m})$ so $j = \{2m + 1, \dots, i + 2m - 1\}$; If i is odd, $D_j = (\frac{2\pi(2m)}{4m}, \frac{2\pi((i+2m)+1)}{4m})$ so $j = \{2m + 1, \dots, i + 2m\}$.

If $2m < i < 4m$, similarly, as shown in Figure 3.5(b), $0 < \beta < \alpha$ where $\alpha = 2\pi - (\phi_{bc} - \phi_{ab})$ and $\beta = \pi - (\phi_{ac} - \phi_{ba})$. Then $\phi_{ac} - \phi_{ba} \in (\phi_{bc} - \phi_{ab} - \pi, \pi)$. If i is even, $D_j = (\frac{2\pi((i-2m+1)-1)}{4m}, \frac{2\pi(2m)}{4m})$ so $j = \{i - 2m + 1, \dots, 2m - 1\}$; if i is odd, $D_j = (\frac{2\pi((i-2m)-1)}{4m}, \frac{2\pi(2m)}{4m})$ so $j = \{i - 2m, \dots, 2m - 1\}$. Rule 2 can be proved analogously. $\square$

3.5 Composing Direction Relations

Definition 9 (Direction relation composition). *Direction relation composition is to infer new direction relation by composing direction relations from the same or different categories of FoRs. Let $\circ^{c_1 c_2 c_3}$ denote the direction relation composition operator where $C_x \in \{\mathcal{A}, \mathcal{I}, \mathcal{R}\}$ indicates the category of the corresponding direction relation of x^{th} element of the operation and $x \in \{1, 2, 3\}$. If $C_1 = C_2 = C_3$, $\circ^{c_1 c_2 c_3}$ is an **in-category composition**; if not, it is an **across-category composition**.*

For example, given three direction relations $i^{c_1}, j^{c_2}, k^{c_3}$ where $i, j, k \in \{0, 1, \dots, 4m - 1\}$, $i^{c_1} \circ^{c_1 c_2 c_3} j^{c_2} = k^{c_3}$ indicates that k^{c_3} can be inferred by composing i^{c_1} and j^{c_2} . In theory, there are at least 3 forms of in-category composition (i.e. $\circ^{AAA}$, $\circ^{III}$, and $\circ^{RRR}$) and 24 forms of across-category composition.

3.5.1 In-category composition

Proposition 4 ($\circ^{AAA}$). *Given two absolute direction relations A_{ab}^i and A_{bc}^j , their in-category composition is $A_{ab}^i \circ^{AAA} A_{bc}^j = A_{ac}^k$ where $k =$*

$$\begin{cases} i, \text{ if } j - i = 0; \\ U, \text{ if } j - i = 2m, \text{ and } i \text{ is odd}; \\ \{i, j\}, \text{ if } j - i = 2m \text{ and } i \text{ is even}; \\ \{i + \Delta_i, \dots, j - \Delta_j\}, \text{ if } 0 < j - i < 2m; \\ \{j + \Delta_j, \dots, i - \Delta_i\}, \text{ if } 2m < j - i < 4m. \end{cases}$$

Proof. The proof is similar to that of Rule 1 in Proposition 3. By definition A_{ab}^i iff $\phi_{ab} - \phi_0 \in D_i$, A_{bc}^j iff $\phi_{bc} - \phi_0 \in D_j$ and A_{ac}^k iff $\phi_{ac} - \phi_0 \in D_k$. If $j - i = 0$ or $2m$ the proof is trivial. If $0 < j - i < 2m$, based on the geometric properties of triangle abc (Figure 3.6(a)), $0 < \beta < \alpha$ where $\alpha = \phi_{bc} - \phi_{ab}$ and $\beta = \phi_{ac} - \phi_{ab}$. Then $\phi_{ac} - \phi_0 \in (\phi_{ab} - \phi_0, \phi_{bc} - \phi_0)$. The rest of the proof can refer to that of Rule 1 in Proposition 3. If $2m < j - i < 4m$ (Figure 3.6(b)), $0 < \beta < \alpha$ where $\alpha = 2\pi - (\phi_{bc} - \phi_{ab})$ and $\beta = 2\pi - (\phi_{ac} - \phi_{ab})$ and the proof is similar. $\square$

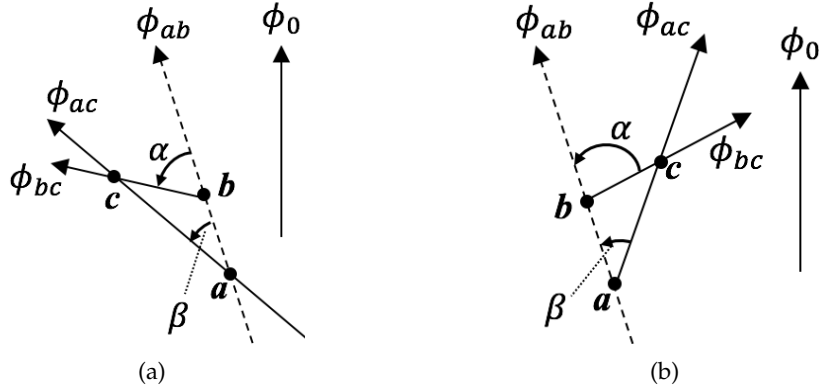


Figure 3.6: (a) when A_{ab}^i, A_{bc}^j and $0 < j - i < 2m$, $0 < \beta < \alpha$ where $\alpha = \phi_{bc} - \phi_{ab}$ and $\beta = \phi_{ac} - \phi_{ab}$. (b) when A_{ab}^i, A_{bc}^j and $2m < j - i < 4m$, $0 < \beta < \alpha$ where $\alpha = 2\pi - (\phi_{bc} - \phi_{ab})$ and $\beta = 2\pi - (\phi_{ac} - \phi_{ab})$.

By contrast, an intrinsic direction relation I_{ab}^i cannot be directly composed with I_{bc}^j because they might be defined in different FoRs along with different anchors (i.e. ϕ_a and ϕ_b). To achieve $\circ^{III}$, intrinsic direction relations should be translated into the same FoR and then refer to the composition rules of $\circ^{AAA}$. Direction relation translation is defined as follows and corresponding translation rules are in Proposition 5.

Definition 10 (Direction relation translation). *Direction relation translation is to translate a direction relation decided in one FoR into that in another based on the relation between the anchors of the FoRs. It can be regarded as a composition of a direction relation and an anchor relation and is denoted as $\circ^T$.*

Proposition 5 ($\circ^T$). *Given three oriented points $a, b, c \in D_{op}$ and $i, j, k \in \{0, 1, \dots, 4m - 1\}$,*

1. $A_{ab}^i \circ^T \Phi_{0a}^j = I_{ab}^k$.
2. $I_{ab}^i \circ^T \Phi_{a0}^j = A_{ab}^k$.

3. $I_{ab}^i \circ^\tau \Phi_{ab}^j = I_{ba}^{k+2m}$. 4. $A_{bc}^i \circ^\tau \Phi_{0b'}^j = R_{abc}^k$.
 5. $R_{abc}^i \circ^\tau \Phi_{b'0}^j = A_{bc}^k$. 6. $I_{bc}^i \circ^\tau \Phi_{bb'}^j = R_{abc}^k$.
 7. $R_{bc}^i \circ^\tau \Phi_{b'b}^j = I_{bc}^k$.

where $\phi_{b'} = \phi_{ab}$ and $k = \Omega_{i,j,i-j}$. Note that $\circ^\tau$ is commutative (e.g. $A_{ab}^i \circ^\tau \Phi_{0a}^j = \Phi_{0a}^j \circ^\tau A_{ab}^i$).

Proof. For Rule 1, by definition A_{ab}^i iff $\phi_{ab} - \phi_0 \in D_i$, Φ_{0a}^j iff $\phi_a - \phi_0 \in D_j$, and I_{ab}^k iff $\phi_{ab} - \phi_a \in D_k$. Since $\phi_{ab} - \phi_a = (\phi_{ab} - \phi_0) - (\phi_a - \phi_0)$ (Figure 3.7(a)), $D_k = D_i - D_j$ and then $k = \Omega_{i,j,i-j}$. Rule 2-7 can be proved analogously and the proof of commutativity is trivial. $\square$

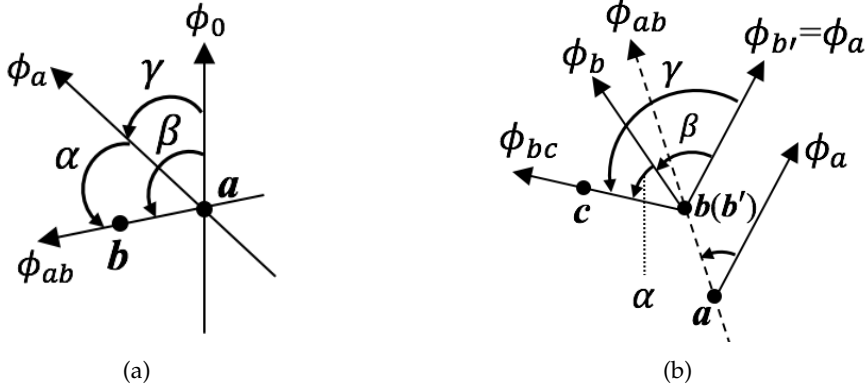


Figure 3.7: (a) $\alpha = \beta - \gamma$ where $\alpha = \phi_{ab} - \phi_a$, $\beta = \phi_{ab} - \phi_0$ and $\gamma = \phi_a - \phi_0$. (b) $\gamma = \alpha + \beta$ where $\alpha = \phi_{bc} - \phi_b$, $\beta = \phi_b - \phi_a$ and $\gamma = \phi_{bc} - \phi_a$.

Proposition 6 ($\circ^{III}$). Given two intrinsic direction relations I_{ab}^i , I_{bc}^j and Φ_{ab}^k the anchor relation between a and b , to achieve $I_{ab}^i \circ^{III} I_{bc}^j$, I_{bc}^j should be first translated into $I_{b'c}^{j'}$ where $S_{b'b}^2$, $\Phi_{b'a}^0$ and $j' = j + k$ (Figure 3.7(b)). Then $I_{ab}^i \circ^{III} I_{b'c}^{j'}$ can refer to the result of $A_{ab}^i \circ^{AAA} A_{b'c}^{j'}$ because I_{ab}^i and $I_{b'c}^{j'}$ are defined in the same FoR.

As discussed in [Isli et al., 2001; Scivos and Nebel, 2001], there are many possible ways to compose two ternary relations. However, it does not mean all the composition results are **informative** (Definition 11). In our representation, for example, the composition of two ternary direction relations with only one common spatial entity is always **non-informative** [Scivos and Nebel, 2004] (e.g. R_{abc}^i and R_{cde}^j). With two common spatial entities, there might be two types of informative composition which differ at which entity is omitted (Definition 12 and 13). For example, both R_{abd}^k and R_{acd}^l might be inferred from composing R_{abc}^i with R_{bcd}^j .

Definition 11 (Informative/non-informative operation). An operation is informative if the operation result could include informative relations; an operation is non-informative if the operation result always includes only non-informative relations.

Definition 12 (Type-1 of $\circ^{\mathcal{RRR}}$). Given four oriented points in a fixed order of a, b, c and d , type-1 ternary relative direction relation composition (denoted as $\circ^{31}$) is to infer the relation among a, b and d by composing the relation among a, b and c with the relation among b, c and d . Composition rules are given in Proposition 7.

Definition 13 (Type-2 of $\circ^{\mathcal{RRR}}$). Given four oriented points in a fixed order of a, b, c and d , type-2 ternary relative direction relation composition (denoted as $\circ^{32}$) is to infer the relation among a, c , and d by composing the relation among a, b , and c with the relation among b, c , and d . Composition rules are given in Proposition 8.

Proposition 7 ($\circ^{31}$). Given two ternary direction relations R_{abc}^i and R_{bcd}^j , their type-1 ternary composition $R_{abc}^i \circ^{31} R_{bcd}^j = R_{abd}^k$ where $k =$

$$\begin{cases} i, & \text{if } j = 0; \\ \{i, i + 2m\}, & \text{if } j = 2m; \\ \{i + \Delta_i, \dots, i + j - \Psi_{i,j}\}, & \text{if } 0 < j < 2m; \\ \{i + j + \Psi_{i,j}, \dots, i - \Delta_i\}, & \text{if } 2m < j < 4m. \end{cases}$$

where $\Psi_{i,j}$ is defined as

$$\Psi_{i,j} = \begin{cases} 1, & \text{if both } i \text{ and } j \text{ are even;} \\ -1, & \text{if both } i \text{ and } j \text{ are odd;} \\ 0, & \text{otherwise.} \end{cases}$$

Proof. By definition R_{abc}^i iff $\phi_{bc} - \phi_{ab} \in D_i$, R_{bcd}^j iff $\phi_{cd} - \phi_{bc} \in D_j$ and R_{abd}^k iff $\phi_{bd} - \phi_{ab} \in D_k$. If $j = 0$ or $2m$ the proof is trivial. If $0 < j < 2m$, based on the geometric properties of triangle bcd (Figure 3.8(a)), $0 < \gamma - \alpha < \beta$ where $\alpha = \phi_{bc} - \phi_{ab}$, $\beta = \phi_{cd} - \phi_{bc}$ and $\gamma = \phi_{bd} - \phi_{ab}$. Then $\phi_{bd} - \phi_{ab} \in (\phi_{bc} - \phi_{ab}, \phi_{bc} - \phi_{ab} + \phi_{cd} - \phi_{bc})$. The rest of the proof is similar to that of Rule 1 in Proposition 3. If $2m < i < 4m$, $0 < \alpha - \gamma < 2\pi - \beta$ where $\alpha = \phi_{bc} - \phi_{ab}$, $\beta = \phi_{cd} - \phi_{bc}$ and $\gamma = \phi_{bd} - \phi_{ab}$ (Figure 3.8(b)) and the rest of the proof can refer to the case when $0 < j < 2m$. $\square$

Proposition 8 ($\circ^{32}$). Given two ternary direction relations R_{abc}^i and R_{bcd}^j , their type-2 ternary composition $R_{abc}^i \circ^{32} R_{bcd}^j = R_{acd}^k$ where $k =$

$$\begin{cases} j, & \text{if } i = 0; \\ \{j, j + 2m\}, & \text{if } i = 2m; \\ \{j + \Delta_j, \dots, i + j - \Psi_{i,j}\}, & \text{if } 0 < i < 2m; \\ \{i + j + \Psi_{i,j}, \dots, j - \Delta_j\}, & \text{if } 2m < i < 4m. \end{cases}$$

The proof is analogous to that of $\circ^{31}$.

3.5.2 Across-category composition

Across-category composition also allows inference of new informative direction relations. In theory there are at least 24 forms of across-category composition, and

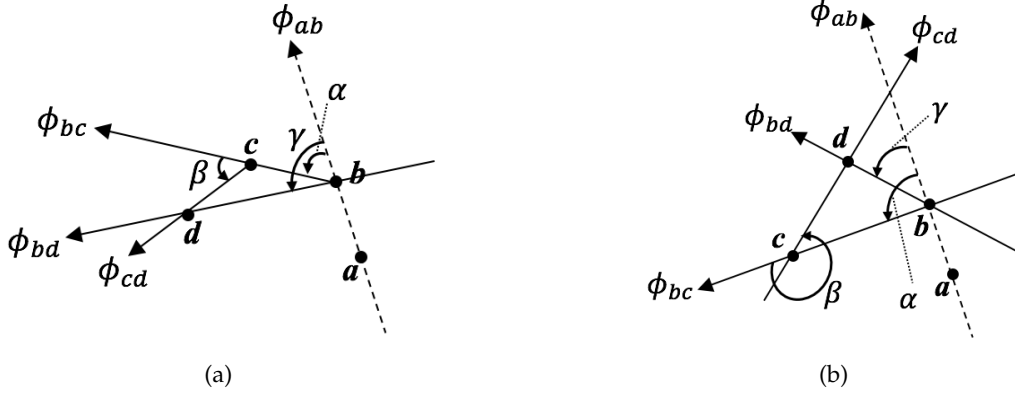


Figure 3.8: (a) when R_{bcd}^j and $0 < j < 2m$, $0 < \gamma - \alpha < \beta$ where $\alpha = \phi_{bc} - \phi_{ab}$, $\beta = \phi_{cd} - \phi_{bc}$ and $\gamma = \phi_{bd} - \phi_{ab}$. (b) when R_{bcd}^j and $2m < j < 4m$, $0 < \alpha - \gamma < 2\pi - \beta$ where $\alpha = \phi_{bc} - \phi_{ab}$, $\beta = \phi_{cd} - \phi_{bc}$ and $\gamma = \phi_{bd} - \phi_{ab}$.

most of them rely on translating direction relations in one FoR into that in another FoR (**translation-dependent composition**) because across-category composition covers direction relations from different FoRs. However, there are 5 forms of across-category composition that do not depend on direction relation translation (**translation-independent composition**). For example, $\circ^{AAR}$, $\circ^{ARA}$ and their composition rules are given in Proposition 9 and 10 respectively.

Proposition 9 ($\circ^{AAR}$). Given two absolute direction relations A_{ab}^i and A_{bc}^j , R_{abc}^k can be inferred from $A_{ab}^i \circ^{AAR} A_{bc}^j$ where $k = \Omega_{i,j,j-i}$.

The proof is trivial given the fact that $\phi_{bc} - \phi_{ab} = (\phi_{bc} - \phi_0) - (\phi_{ab} - \phi_0)$.

Proposition 10 ($\circ^{ARA}$). Given an absolute direction relations A_{ab}^i and a relative direction relation R_{abc}^j , A_{bc}^k can be inferred from $A_{ab}^i \circ^{ARA} R_{abc}^j$ where $k = \Omega_{i,j,i+j}$.

The proof is trivial given the fact that $\phi_{bc} - \phi_0 = (\phi_{bc} - \phi_{ab}) + (\phi_{ab} - \phi_0)$. The rules of $\circ^{RAA}$, $\circ^{TRI}$ and $\circ^{RII}$ are analogous.

Different from translation-independent compositions, translation-dependent compositions are based on one or more direction relation translations (Definition 10), for example $\circ^{AZA}$ should be first transformed into $\circ^{AAA}$ (Proposition 11).

Proposition 11 ($\circ^{AZA}$). Given an absolute direction relation A_{ab}^i and an intrinsic direction relation I_{bc}^j , their across-category composition $A_{ab}^i \circ^{AZA} I_{bc}^j$ can be transformed into $A_{ab}^i \circ^{AAA} A_{bc}^j$ by translating I_{bc}^j into A_{bc}^j according to Rule 2 of $\circ^T$ (Proposition 5).

Similar strategy can be used in $\circ^{AII}$, $\circ^{AAI}$, $\circ^{IIA}$, $\circ^{IAA}$, $\circ^{IAI}$, $\circ^{IIR}$ and in across-category compositions that cover all three categories (there are 6 of them and for example $\circ^{AIR}$). The remaining 6 forms of across-category compositions (i.e. $\circ^{RAA}$, $\circ^{ARR}$, $\circ^{RAR}$, $\circ^{RRI}$, $\circ^{IRR}$ and $\circ^{RII}$) that have not been discussed so far include exactly two relative direction relations and all of them are non-informative.

3.6 Anchor Relation Operations

As can be seen in the previous section, without the information about the relation between anchors, it is almost impossible to connect direction information described in one FoR to that in another (except for absolute direction relations as they are defined in the same absolute anchor). Anchor relations might be directly described in natural languages. For example, "the chair is facing north" [Levinson, 1996], where the orientation of the chair (ϕ_c) is north (ϕ_0) and can be expressed as Φ_{c0}^0 . The relation between two anchors can also be inferred if the direction relation between two oriented points is described according to both of them (Proposition 13) or by composing another two anchor relations (Proposition 14).

Proposition 12 (Converse of anchor relations). *Given two oriented points $a, b \in D_{op}$ and their anchor relation Φ_{ab}^i where $i \in \{0, 1, \dots, 4m - 1\}$, the converse of Φ_{ab}^i is Φ_{ba}^{-i} (denoted as $\Phi_{ab}^{i\sim}$).*

Proof. By definition, Φ_{ab}^i iff $\phi_b - \phi_a \in D_i$. Assume Φ_{ba}^j then $\phi_a - \phi_b \in D_j$. It is not difficult to infer $i + j = 0$ from $\phi_a - \phi_b = -(\phi_b - \phi_a)$ (Figure 3.9(a)). $\square$

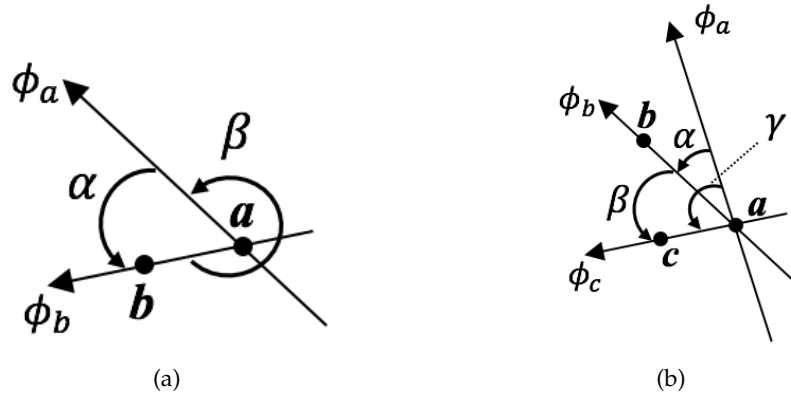


Figure 3.9: (a) $\alpha + \beta = 0$ where $\alpha = \phi_b - \phi_a$ and $\beta = \phi_a - \phi_b$. (b) $\gamma = \alpha + \beta$ where $\alpha = \phi_b - \phi_a$, $\beta = \phi_c - \phi_b$ and $\gamma = \phi_c - \phi_a$.

Proposition 13 (Anchor relation inference). *Given three oriented points $a, b, c \in D_{op}$ and $i, j, k \in \{0, 1, \dots, 4m - 1\}$,*

1. A_{ab}^i and $I_{ab}^j \Rightarrow \Phi_{0a}^k$
2. I_{ab}^i and $I_{ba}^j \Rightarrow \Phi_{0a}^k$
3. A_{bc}^i and $R_{abc}^j \Rightarrow \Phi_{0b'}^k$
4. I_{bc}^i and $R_{abc}^j \Rightarrow \Phi_{bb'}^k$, where $\phi_{b'} = \phi_{ab}$ and $k = \Omega_{i,j,i-j}$.

Proof. For Rule 1, the proof is trivial given the fact that $\phi_a - \phi_0 = (\phi_{ab} - \phi_0) - (\phi_{ab} - \phi_a)$ (Figure 3.7(a)). Rule 2-4 can be proved analogously. $\square$

Proposition 14 (Anchor relation composition). *Let $\circ^\phi$ denote anchor relation composition. Given Φ_{ab}^i and Φ_{bc}^j two anchor relations, $\Phi_{ab}^i \circ^\phi \Phi_{bc}^j = \Phi_{ac}^k$ where $k = \Omega_{i,j,i+j}$.*

Proof. The proof is trivial given the fact that $\phi_c - \phi_a = (\phi_b - \phi_a) + (\phi_c - \phi_b)$ (Figure 3.9(b)). $\square$

3.6.1 Operations over sets of basic relations

The rules of relation operation (i.e. converse and composition) in the previous sections are for basic relations. For sets of basic relations, the converse of a union of basic relations is equal to the union of the converse of all basic relations. Namely, $(\{i, j, \dots, k\}^c)^\sim = (i^c)^\sim \cup (j^c)^\sim \cup \dots \cup (k^c)^\sim$, where $\mathcal{C} \in \{\mathcal{A}, \mathcal{I}, \mathcal{R}, \phi\}$. $\sim$ can also be converse of ternary relations e.g. $\sim_{inv}$.

The composition of two unions of basic relations is equal to the union of compositions of pairs of relations from these two sets. Namely, $\{i, j, \dots, k\}^{c_1} \circ^* \{u, v, \dots, w\}^{c_2} = (i^{c_1} \circ u^{c_2}) \cup (i^{c_1} \circ^* v^{c_2}) \cup \dots \cup (i^{c_1} \circ^* w^{c_2}) \cup (j^{c_1} \circ u^{c_2}) \cup (j^{c_1} \circ^* v^{c_2}) \cup \dots \cup (j^{c_1} \circ^* w^{c_2}) \cup \dots \cup (k^{c_1} \circ u^{c_2}) \cup (k^{c_1} \circ^* v^{c_2}) \cup \dots \cup (k^{c_1} \circ^* w^{c_2})$, where $\circ^*$ can be $\circ^\phi$, $\circ^\tau$ or $\circ^{c_1 c_2 c_3}$ and $c_1, c_2, c_3 \in \{\mathcal{A}, \mathcal{I}, \mathcal{R}\}$.

3.7 Towards Constraint-based Reasoning

The core problem in qualitative spatial reasoning is deciding whether a set of spatial relations is **consistent** because most of the other problems can be reduced to it [Renz and Nebel, 1999]. The consistency problem can be transformed to a **constraint satisfaction problem** (CSP) because spatial relations can also be regarded as constraints between spatial entities [Renz and Nebel, 2007]. In our case, the most important reasoning problem over Θ (an instance of $\mathcal{DAF}_m$) is whether it is **across-category consistent**.

Definition 14 (Across-category consistency). Θ (an instance of $\mathcal{DAF}_m$) is across-category consistent or equivalently Θ is satisfiable iff there exists at least one set of oriented points that can satisfy all the constraints originated from its relations.

Classical composition and path-consistency based methods cannot be used to check across-category consistency because the basic relations of $\mathcal{DAF}_m$ are not JEPD (Lemma 1) and some of the relations are not closed over converse or composition. However, $CSP(\mathcal{DAF}_m)$ is decidable (Theorem 1) because the constraints can be transformed into a set of inequalities [Moratz and Ragni, 2008] and it is possible to transform $\mathcal{DAF}_m$ relations so that they are JEPD (Theorem 2).

Theorem 1. $CSP(\mathcal{DAF}_m)$ is NP-hard and in PSPACE.

Proof. Given an instance of $\mathcal{DAF}_m$ and its set of relations (consisting of same position relations, absolute/intrinsic/relative direction relations and anchor relations), by definition same position relations and anchor relations are linear inequalities. For example, S_{ab}^2 iff $x_a = x_b$ and $y_a = y_b$; Φ_{ab}^i iff $\phi_b - \phi_a \in D_i$. If i is even, $\phi_b - \phi_a = \frac{4m}{2\pi i}$; if i is odd, $\frac{4m}{2\pi(i-1)} < \phi_b - \phi_a < \frac{4m}{2\pi(i+1)}$.

Absolute direction relations can be transformed into intrinsic direction relations. For example, A_{ab}^i can be transformed into $I_{a'b}^i$ where $S_{a'a}^2$ and $\Phi_{a'0}^0$ (Figure 3.10(a)). Intrinsic direction relations can be transformed into relative direction relations. For example, I_{ab}^i can be transformed into $R_{a'ab}^i$ where a' is an arbitrary point that is in the opposite direction of ϕ_a from a so that $\phi_{a'a} = \phi_a$ (Figure 3.10(b)). According to [Moratz and Ragni, 2008], relative direction relations can be transformed into polynomial inequalities of power 2 with integer coefficients, which can be solved in polynomial space [Renegar, 1992]. The NP-hardness of $\text{CSP}(\mathcal{DAF}_m)$ is based on the results in [Scivos and Nebel, 2001]. $\square$

Theorem 2. *A set of $\mathcal{DAF}_m$ relations can be transferred to a set of JEPD ternary relative direction relations and ternary same position relations without information loss.*

Proof. First we prove that a basic absolute direction relation can be transferred to a basic intrinsic direction relation without information loss. As in the proof of Theorem 1, $A_{ab}^i \Rightarrow I_{a'b}^i$ where $S_{a'a}^2$ and $\Phi_{a'0}^0$ (as shown in Figure 10). There is no information loss in this process because conversely $I_{a'b}^i \wedge S_{a'a}^2 \wedge \Phi_{a'0}^0 \Rightarrow A_{ab}^i$. Similarly, $R_{a'ab}^i \wedge (\phi_a = \phi_{a'a}) \Rightarrow I_{ab}^i$ then it is proved that a basic intrinsic direction relation can be transferred to a basic relative direction relation without information loss. The only difference is to store the extra information $\phi_{a'a} = \phi_a$. So a set of basic absolute, intrinsic and relative direction relations can be transferred to a set of basic relative direction relations without information loss. For non-basic direction relations the proof is similar.

For same position relations, it is trivial to prove that a set of basic binary same position relations can be equivalently transformed into a set of basic ternary same position relations. For non-basic same positions relations, the proof is also trivial. For example, $S_{ab}^2 \wedge (S_{bc}^2 \vee \neg S_{bc}^2) \Leftrightarrow S_{abc}^3 \vee S_{abc}^{12}$.

Next we will prove that a set of relative direction relations can be transferred to a set of JEPD ternary relative direction relations and same position relations without information loss. Note that a set of JEPD basic ternary relations can be $\{S^{12}, S^{23}, S^{13}, S^3, 0'^R, 1'^R, \dots, (4m-1)'^R\}$ where i'^R ($i \in \{0, 1, \dots, 4m-1\}$) is different from i^R because of an extra constraint $\neg S^{13}$. A set of basic ternary relative direction relations and same position relations can be transformed into this set of JEPD relations by the following rules:

If $i \neq 2m$, replace R_{abc}^i with $R_{abc}^{i'}$; if $i = 2m$ and $\neg S_{abc}^{13}$, replace R_{abc}^i with $R_{abc}^{i'}$; and if $i = 2m$ and S_{abc}^{13} , replace R_{abc}^i with S_{abc}^{13} . There is no information loss since conversely if $i \neq 2m$, $R_{abc}^{i'} \Rightarrow R_{abc}^i$; if $i = 2m$, $R_{abc}^{i'} \Rightarrow \neg S_{abc}^{13}$ and R_{abc}^i ; and if S_{abc}^{13} we can get R_{abc}^i . $\square$

As discussed earlier, path-consistency cannot be used to check across-category consistency because our basic relations are not JEPD. Instead, another form of local consistency (i.e. converse consistency in Definition 15) is used as a prerequisite of across-category consistency.

Definition 15 (Converse consistency). *A relation is converse consistent iff there is at least one consistent scenario that can satisfy its converse relation(s) and itself.*

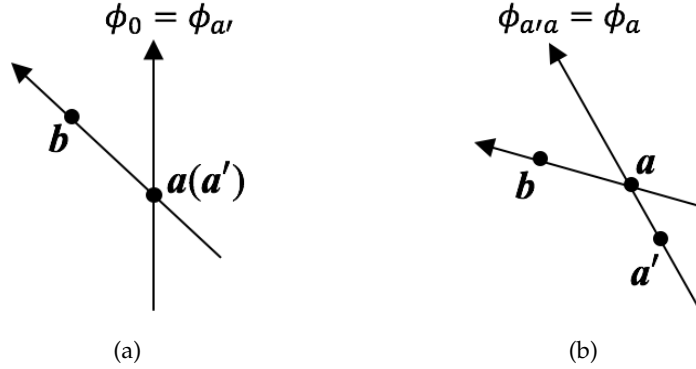


Figure 3.10: (a) An absolute direction relation A_{ab}^i can be transformed into an intrinsic direction relation $I_{a'b}^i$ where $S_{a'a}^2$ and $\Phi_{a'0}^0$. (b) An intrinsic direction relation I_{ab}^i can be transformed into a relative direction relation $R_{a'ab}^i$ where $\phi_a = \phi_{a'a}$.

Converse consistency can be used as a necessity of across-category consistency because if any relation of $\mathcal{DAF}_m$ is not converse consistent, it is not across-category consistent. For same position relations, absolute/intrinsic direction relations and anchor relations, converse consistency check is trivial because they are binary relations and there is only one form of converse relations. However, for relative direction relations, it is non-trivial because there are five converse relations and corresponding converse consistency check can refer to Proposition 15.

Proposition 15. *Given three relative direction relations R_{abc}^i , R_{cab}^j and R_{bca}^k among oriented points a , b and c , they are converse consistent iff there exist at least one set of three oriented points a , b and c that can satisfy all of the following constraints:*

1. R_{cba}^{-i} ; 2. R_{bac}^{-j} ; 3. R_{acb}^{-k} ;
4. $\text{Triangle}(2m - i, 2m - j, 2m - k)$ where $\text{Triangle}(x, y, z)$ is true iff there exist three angles $\alpha \in D_x$, $\beta \in D_y$ and $\gamma \in D_z$ that can form a triangle [Mossakowski and Moratz, 2012].

Proof. If a relative direction relation is converse consistent, there exists at least one set of three oriented points a , b and c that can satisfy this relative direction relation and all its converse relations. The proof of Constraint 1-3 can refer to Rule 4-6 in Proposition 3 (or as shown in Figure 3.11(a), $\alpha + \beta = 0$). For Constraint 4, based on the geometric properties of triangle abc in Figure 3.11(a), γ_1 , γ_2 and γ_3 can form a triangle where $\gamma_1 = \pi - (\phi_{bc} - \phi_{ab})$, $\gamma_2 = \pi - (\phi_{ab} - \phi_{ca})$ and $\gamma_3 = \pi - (\phi_{ca} - \phi_{bc})$. $\square$

3.8 Case Study and Applications

Our case study is inspired by the hare-shooting story in [Ligozat, 1993]: there are two hunters Allen and Ben who dislike each other because Allen (a) insists on using

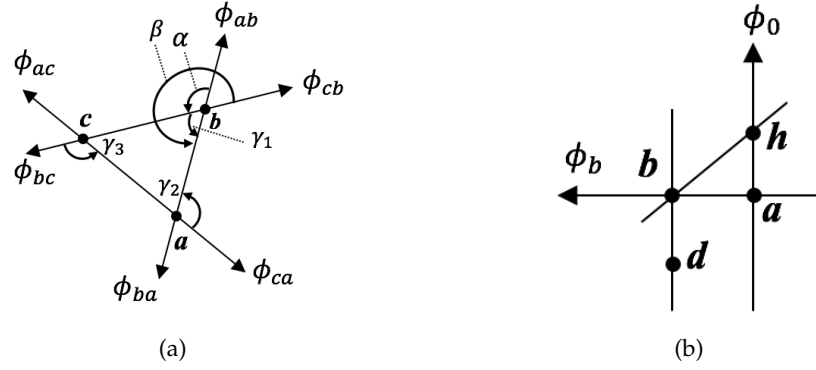


Figure 3.11: (a) $\alpha + \beta = 0$ and γ_1, γ_2 and γ_3 can form the triangle abc . (b) Relative positions of the hunters (**a** and **b**), the dog (**d**) and the hare (**h**). ϕ_0 indicates the north and ϕ_b is the orientation of **b**.

absolute FoRs to describe direction information while Ben (**b**) prefers to describe direction relations in his own orientation. One day they go for hare-shooting with a "smart" hunting dog (**d**). Given the following information (Figure 3.11(b)), if Allen asks the dog to chase the hare (**h**) how can the dog know where the hare is? Assume these two hunters keep moving so that the strategy of first moving to one of the hunters cannot be used.

(1) Allen yells to Ben "Hey a hare in the north (of me)! Why do you face to west? It is in your northeast!".

(2) Ben keeps moving forwards and replies "Stop yelling at me! I am not interested at the hare if it is found by you. Why don't you ask the dog that is standing to my left to get it!?".

(3) The dog looks at Ben and finds out Allen is to the right of him.

It is almost impossible to solve this problem with the existing formalisms because most of them can just deal with direction information defined in one category of FoRs. However, useful information can be inferred by applying $\mathcal{DAF}_4$. The direction information in (1) can be expressed by A_{ah}^0 , Φ_{0b}^4 , and A_{bh}^{14} ; that in (2) can be described by I_{bd}^4 ; while R_{dba}^{12} can express the direction information in (3). Our aim is to infer the direction relation between the hare and dog which can be denoted as A_{dh}^i where i is unknown. Based on the inference rules introduced in the previous sections, $i = \{13, \dots, 3\}$ (induction details are as follows and refer to Figure 3b for a pictorial view of these sectors), which means the dog should move to a rough sense of north.

- #1. $\Phi_{0b}^4 = \Phi_{b0}^{12}$ from (1).
- #2. $I_{bd}^4 \circ^T \Phi_{b0}^{12} = A_{bd}^8$ from (2) and #1.
- #3. $A_{bh}^{14} = A_{hb}^6$ from (1).
- #4. $A_{ah}^0 \circ^{AAR} A_{hb}^6 = R_{ahb}^6$ from (1) and #3.
- #5. $R_{ahb}^6 \circ^{sci} R_{bah}^{\{3, \dots, 7\}} = R_{bah}^{\{3, \dots, 7\}}$ from #4.
- #6. $R_{dba}^{12} \circ^{31} R_{bah}^{\{3, \dots, 7\}} = R_{dbh}^{\{13, \dots, 3\}}$ from (3) and #5.
- #7. $A_{bd}^8 = A_{db}^0$ from #2.

#8. $A_{db}^0 \circ^{ARA} R_{dbh}^{\{13, \dots, 3\}} = A_{bh}^{\{13, \dots, 3\}}$ from #7 and #6.

#9. $A_{db}^0 \circ^{AAA} A_{bh}^{\{13, \dots, 3\}} = A_{dh}^{\{13, \dots, 3\}}$ from #7 and #8.

In practice, sets of relations can be used to model inaccuracy of inaccurate direction information. For example, $A_{ah}^{\{15, 0, 1\}}$ can indicate h is roughly north of a .

Direction relations occur very frequently in human communication and therefore also on Social Media and (future) human-robot interaction. Unlike humans, current AI agents are unable to understand and deal with these expressions properly. Our spatial constraint language $\mathcal{DAF}_m$ is a breakthrough in solving this problem, which could stimulate applications of qualitative spatial reasoning in areas like robot navigation, navigation regulation, direction relation recognition, query-answering and so on [Wolter and Wallgrün, 2013].

3.9 Conclusion

To the best of our knowledge, we have presented the first formalism that can represent and reason direction relations across different frames of reference. Direction relations in three different categories of FoRs have been unified and the implicit factor (i.e. anchor) that is critical in deciding direction relations has been modeled. We provided approaches to infer the anchors used and based on which direction relations in one FoR can be translated into another. We also discussed how across-category composition and inconsistency check can be done.

Integrative Qualitative Localization and Navigation

“The journey with a 1000 miles begins with one step.” – Confucius.

GPS-based services (e.g. Google Maps) are very popular in our daily life, while there are still many GPS-denied environments (e.g. indoor and underground scenarios) in which they cannot be used. In these situations, localization and navigation are still important, for example in emergency evacuation or indoor navigation. In this chapter, we aim to solve the problem of localizing and navigating humans or robots in GPS-denied environments based on landmarks. Our work is inspired by human daily communications about localization and navigation, for example someone who has been to a shopping mall many times can localize and guide another person to get to a certain shop via conversations over the cellphone. Our goal is to build a system with the same capability.

In this chapter, we propose a system that relies on qualitative information of places (e.g. the direction relations between landmarks involved in route descriptions), where localization can be achieved in an interactive manner and by analyzing observations provided by users. Our system decides the “best” route from one place to another by three factors: the number of landmarks; the number of ambiguous turns; and the qualitative distance (e.g. near and far). According to the experimental results, the number of queries in the interactive localization process is acceptable and our route planning algorithm outperforms previous methods in several cases.

4.1 Introduction

How to correctly move from the current place to the destination is fundamental in our daily life. This can be called the navigation problem, which is also a research problem that has been widely and extensively discussed in several different fields (e.g. Geographical Information System, Computer Science and Robotics). The following sub-problems have to be considered in most cases though researchers from different fields have different focuses:

Map. Are online map services (e.g. Google Maps) available in the current environment? It is ideal to have a complete and precise map at-hand but navigation is still possible without such maps. For example, in a scenario where an out-of-towner asks a local resident how to get to a certain place, the local resident can describe a route by referring to the imaginary map in his/her mind.

Localization. If GPS is available, it is trivial to locate the current position. However, there are still many GPS-denied environments where other forms of localization mechanisms are still required: indoor and underground environments where there is only unstable or no GPS-signal (e.g. shopping malls and mines, or inner cities with many high-rises); battlefield situations where GPS is jammed, or events where it is not allowed or not possible to use GPS (e.g. orienteering).

Route Planning. How to find the best route from the origin to the destination? Another problem that might be equally important is how to define the “best” route? Dijkstra [Dijkstra, 1959] is the most popular algorithm to find the shortest path. However, in some cases the “simplest” route is more preferable [Duckham and Kulik, 2003]. “Simplest” can be defined in different ways. For example, the route with fewest turns or the route that is the easiest to be described [Richter and Duckham, 2008].

Route Instruction. Route instructions can be either textual (e.g. route descriptions in natural language) or pictorial (e.g. routes that are illustrated in a sketched map or on the screen of a mobile device). How to describe a route in a less ambiguous way? For example, at an intersection there is more than one right turn, it is better to indicate which right turn to take (e.g. “next intersection take the first right turn” rather than just “next intersection turn right”).

Decision points and Landmarks. Agents have to make decisions (e.g. go straight or make turns) at decision points (e.g. intersections or corners). Landmarks are salient places so that their positions can be regarded as “known” and be used to locate other places. Decision points are indispensable in navigation and landmarks are helpful in simplifying route instructions and eliminating ambiguities [Michon and Denis, 2001].

Agent. Humans, robots or vehicles that follow and execute certain route instructions to get from one place to another will be called “agents” in [Westphal and Renz, 2011]. Agents can be different in several aspects. Except for moving ability and the form of route instructions that can be utilized, agents can be also different in cognitive ability (e.g. whether they are able to recognize landmarks).

Requery. It is possible that an agent is lost during a navigation process. In such cases, requeries are necessary so that agents can be re-located and the navigation process can continue by following a re-planned route.

In this chapter, we will focus on the problem of localizing and navigating agents in GPS-denied environments based on landmarks. We also assume there is neither

a precise and complete map nor an online map service like Google Maps available to the user for the area we are navigating in [Winter et al., 2019; Wijewardena et al., 2016]. Solving this problem is important for indoor navigation such as in subway stations or shopping centres, but also in extreme scenarios like natural disasters where quickly and reliably navigating to safe places can save lives [Winter et al., 2011]. For example, when there is a fire in a shopping mall, how to quickly find the route to the Exit from the current place (e.g. the Converse shop) only given the following two previously recorded route descriptions?

(1) Can you see the Adidas shop? Go there and turn left, you can see the Burberry shop. At the corner near Burberry turn left and walk for about 1-2 minutes and you will arrive at the Converse shop.

(2) Continue along this way until you can find the Burberry shop. At the elevator next to it turn right and go straight for about 50-60 meters where you can see Domino's Pizza. The Exit is very close to Domino's. Just make a slight right turn.

Such descriptions can be found in scenarios where people are asked to describe routes from memory. In the following we will describe a navigation system that can locate and navigate in GPS-denied environments given just incomplete and imprecise knowledge about the positions of places like the examples given above. Based on the above two route descriptions, our system should be able to infer a route from the Converse shop to the Exit.

The rest of this chapter is organized as follows: in Section 2 we introduce related work; in Section 3 we describe how to construct a map of places that is helpful in the process of localization (Section 4) and navigation (Section 5); in Section 6 we demonstrate how our navigation system can be used by solving the problem of how to get to the Exit from the Converse shop; our solution is evaluated by experiments in Section 7; In Section 8 we conclude the chapter and also discuss future work.

4.2 Related Work

Dijkstra's algorithm [Dijkstra, 1959] (or the shortest path algorithm) is to find the shortest path between two nodes in a directed graph where edges are labelled with weights (by default the distance between the two nodes but also can be other metrics like travel time). Following this pattern, directed graphs are also used to represent real-world traffic networks (where there will be an edge from one node to another if there is a realistic path in the traffic network) [Haque et al., 2006; Duckham and Kulik, 2003] because it is convenient for further analysis.

In [Haque et al., 2006], the shortest most reliable path algorithm (which is an adapted version of the shortest path algorithm) is provided to find the most reliable path. Shorter path is more preferable to break ties when two paths are equally reliable. The unreliability of a path is calculated by summing the unreliability of all turns involved in this path and the unreliability of a turn is decided by the number of other turns that are instruction equivalent at the same intersection. For example, when arriving at a certain intersection from a certain direction, there are two turns

on the left. The unreliability of each left turn is 1 because for each of them there is one other instruction equivalent turn.

In contrast, the simplest path algorithm [Duckham and Kulik, 2003] aims to find the simplest path under the assumption that it is simpler to describe paths with less number of turns or turns in less complex intersections. The complexity of different intersections has been discussed in [Winter, 2002]. The simplest path algorithm is also adapted from the shortest path algorithm. The computational complexity of both algorithms is $O(|E|^2)$ or $O(|V|^2)$ where $|E|$ is the number of edges and $|V|$ is the number of nodes in the directed graph under the assumption that the directed graphs that representing traffic networks are sparse (then it is safe to have $O(|E|) = O(|V|)$).

Following the simplest path algorithm, Richter and Duckham [Richter and Duckham, 2008] propose the simplest instruction algorithm and claim that by using their algorithm the best paths can be described with much smaller number of instructions (on average a 50% decrease) when compared to both of the shortest path algorithm and the simplest path algorithm with an acceptable increase of the path length (on average a 10% increase). This improvement relies on considering the complexity of different intersections, referring to landmarks and applying the technique of spatial chunking. Landmarks are frequently referred to in route instructions because they can be used to indicate crucial turns (e.g. “turn right at the bus station”); locate other places (e.g. “the library is to the left of the bus station”); and ensure the route instructions are being correctly followed (e.g. “if you can find a bus station then you are in the right way”) [Richter and Duckham, 2008]. Spatial chunking is used to describe several continuous decision points in one instruction rather than multiple instructions. For example, “turn left at the second traffic light” rather than “go straight at the first traffic light and then turn left at the second”.

There are quite a few discussions on landmarks since they are frequently utilized in navigation as mentioned above. As pointed out by [Michon and Denis, 2001], landmarks are helpful in unfamiliar environments and people tend to use landmarks more frequently when they have to make turns. A review of research on navigation with landmarks can be found in [Duckham et al., 2010] where three approaches for landmark identification have been introduced and also a new navigation model based on landmarks has been proposed. Their landmark navigation model actually focuses on generating simple route instructions that include landmarks. In the navigation system presented in [Serrão et al., 2014], the complete and precise map of a building is given by an indoor geographic information system. Visual landmarks (i.e. landmarks that can be recognized using vision) are used for localization.

There is also research on generating more sophisticated “human-like” landmark-based route instructions. The developed prototype “IndoorNav” [Russo et al., 2014] can locate a user by requiring them to scan the nearest QRCode and provide route instructions to the destination by considering all the direction relations of visible landmarks along a certain route. Grouping (similar to spatial chunking) is used to simplify textual instructions. The Indoor Landmark Navigation Model (ILNM) proposed in [Fellner et al., 2017] relies on spatial objects that are in common categories (e.g. doors or elevators) rather than landmarks with unique or salient visual, se-

mantic or structural characteristics. Deep learning methods can also be helpful in generating route instructions in natural languages as described in [Oßwald et al., 2014; Daniele et al., 2017].

According to [Daniel and Denis, 1998], the majority of landmark descriptions are in the form of direction relations. Duckham et al. [Duckham et al., 2010] also argue that numerical descriptions of distance or turning angles are very rare in human route descriptions. To be closer to human understanding of spatial knowledge used in navigation (e.g. the direction relations and topological relations between landmarks) and to achieve better human-robot interaction, it is natural to describe landmarks with their qualitative spatial relations. Qualitative Spatial Representation and Reasoning focuses on the representation and reasoning of qualitative spatial relations between spatial entities [Renz and Nebel, 2007; Cohn and Renz, 2008]. Qualitative direction calculi (which are systematic models of sets of qualitative direction relations) are helpful in modeling directions that are heavily used in route descriptions. For example, the double-cross calculus [Freksa, 1992c] that can model the direction relation between three points is applied in [Krieg-Brückner and Shi, 2006] to generate route descriptions to support the communication between a driver and an intelligent wheelchair; in [Dylla, 2009; Lücke et al., 2011] $\mathcal{OPRA}_m$ (a qualitative calculus that can deal with the relative direction relations between pairs of oriented points) has been used for navigating agents. Please refer to Chapter 2 for more details.

Such qualitative direction relations can be converted from quantitative data in geographical information systems (e.g. OpenStreetMap data) as in [Lücke et al., 2011]; or be extracted from real-world natural language place descriptions (e.g. the two descriptions in the introduction) in the form of “Place A Relation Place B” [Liu et al., 2014] and be formalized as place graphs [Vasardani et al., 2013; Chen et al., 2018].

More and more robotics researchers have made efforts to handle qualitative input (e.g. natural language direction description [Kyriacou et al., 2005; Krieg-Brückner and Shi, 2006] or sketched maps [Chronis and Skubic, 2004; Shah and Campbell, 2013; Boniardi et al., 2015]) because qualitative representations are more suitable for human-robot interaction though traditionally most navigation strategies in this field are based on quantitative input (e.g. metric data collected from sensors or cameras, OpenStreetMap data, or floor plans). Their research focuses more on interpreting qualitative input and applying it for path planning rather than route planning.

Note that in some cases the two concepts of “path” and “route” are used as synonyms (e.g. the shortest path and the shortest route). In this chapter, to avoid ambiguity, “path” and “route” are regarded as different terms in the sense that route planning is to find the “best” route from the origin to the destination though the definition of “best” varies and in contrast path planning is to deal with the problems about how to precisely move (e.g. motion trajectory planning; and obstacle avoidance).

In this chapter, we focus on qualitative navigation (i.e. solving the navigation problem by qualitative input [Fogliaroni et al., 2009; Levitt, 1990]), which is different from quantitative navigation that heavily relies on quantitative input. To the best of

our knowledge, our solution is the first that combines the following features: qualitative knowledge (e.g. the direction and distance relations between places) as input; landmark-based; interactive localization; and deciding the best route by considering number of landmarks, number of ambiguous turns and qualitative distance. We believe our solution is different from previous solutions in reliability and is closer to how humans are doing it and therefore is more intuitive for humans to use.

4.3 Qualitative Place Map

In a GPS-denied environment, localization and navigation are still possible by referring to the descriptions of places from people who are familiar with this area. Note that in this chapter places can be either landmarks (e.g. the Adidas shop) or non-salient places (e.g. turns or elevators, or not so well-known branded shops). A qualitative place map is a map of places in a qualitative form, which is inspired by the qualitative constraint networks of landmarks discussed in [Li et al., 2013]. They focus on solving some of the relevant theoretical problems while we aim to apply this form of representation on solving real-world problems. In this section, we will introduce the formal definition of qualitative place map and also provide construction algorithms in different scenarios.

Definition 16 (Qualitative Place Map). *A qualitative place map (QPM) is a directed graph whose vertices are places and the edges between vertices are labelled with the qualitative knowledge about places. Namely a QPM is a tuple (V, E) where $V = \{v_1, v_2, \dots, v_n\}$ is the set of vertices (n is the number of vertices) and $E = \{e_1, e_2, \dots, e_m\}$ is the set of edges (m is the number of edges).*

In a QPM, each vertex $v \in V$ has two attributes: the name of its corresponding place and whether the place is a landmark. The naming function is defined to be $pn : V \rightarrow S$ where S is a set of place names; and the landmark function is $lm : V \rightarrow \{t, f\}$ (t : true and f : false). For example let v_1 represent the “Adidas shop” and v_2 represent the “corner” in the first description then $pn(v_1) = Adidas$, $pn(v_2) = corner$, $lm(v_1) = t$ and $lm(v_2) = f$. An intuitive landmark judging strategy that only relies on place names is used in this chapter while more sophisticated techniques can be found in the relevant research introduced in Section 2: in a certain environment places with unique and semantic names are landmarks (e.g. the Adidas shop or the Big Clock) while places with names that are not unique, not well-known, or only indicate their functions are not (e.g. turns, elevators or doors).

Each edge $e \in E$ is associated with a label consisting of six elements. Namely, $e = (v_s, v_e, r_{rea}, r_{vis}, R_{dir}, r_{dis})$ where $v_s, v_e \in V$ indicate the starting vertex and the ending vertex respectively, $r_{rea} \in \{t, f\}$ (t : true and f : false) denotes whether v_e is directly reachable from v_s (namely there is a direct path between them), $r_{vis} \in \{t, f\}$ denotes whether v_e is visible from v_s , R_{dir} is the set of qualitative direction relations between v_s and v_e and r_{dis} is the qualitative distance relation between them.

4.3.1 Direction Relations

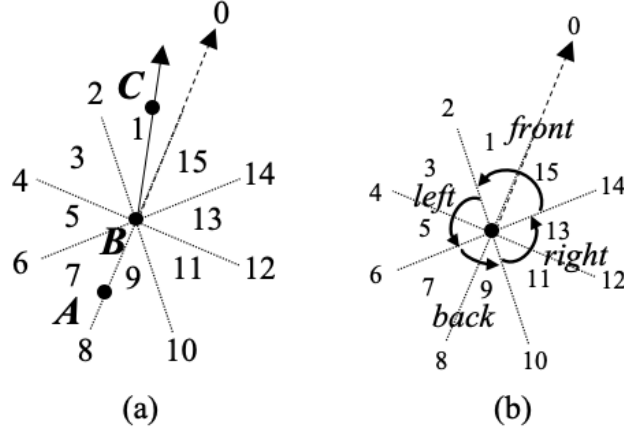


Figure 4.1: (a) By using the qualitative direction relation calculus $\mathcal{RDR}_4$, the direction relation between A , B and C is 1. (b) The direction relation labelling function that assigns the 16 relations with the labels “front, left, back and right”.

The qualitative direction relation calculus used in this chapter is Relative Direction Relation ($\mathcal{RDR}_m$), which is part of $\mathcal{DAF}_m$ [Hua et al., 2018] and also an adaptation of the $\mathcal{STAR}_m$ calculus [Renz and Mitra, 2004] and the double-cross calculus [Freksa, 1992c] to model the direction relations of place triples with arbitrary granularities. As the example in Figure 1a shows there are three places A , B and C and $m = 4$ and the 2D plane is divided into $4 * 4$ sectors by 4 lines and these sectors are indexed from 0 to $4 * 4 - 1$ (i.e. 15). Sectors indexed with even numbers are even sectors (half lines in the 2D plane) and those indexed with odd numbers are odd sectors (convex regions in the 2D plane). The direction relation between A , B and C is 1 because C is in Sector 1 given B as the centroid and the orientation of Sector 0 the same as $A \rightarrow B$. The Direction Relation Labelling function is defined to be: $l_{dir} : Z_{4m} \rightarrow L_{dir}$ where $Z_{4m} = \{0, 1, \dots, 4m - 1\}$ and L_{dir} is a set of direction relation labels in natural language (e.g. $\{front, left, back, right\}$). Its converse function is defined as $l_{dir}^{-1} : L_{dir} \rightarrow Z_{4m}$. In this chapter, the default qualitative direction relation calculus is $\mathcal{RDR}_4$ and the default l_{dir} function (as in Figure 1b) is

- $l_{dir}(0, 1, 15) = front$
- $l_{dir}(2, 3, 4, 5, 6) = left$
- $l_{dir}(7, 8, 9) = back$
- $l_{dir}(10, 11, 12, 13, 14) = right$

The value of m and the corresponding l_{dir} function are configurable because in different scenarios the size of sectors and the natural language terms to describe each sector might vary. For example, the sector model in [Klippel et al., 2005] is shown

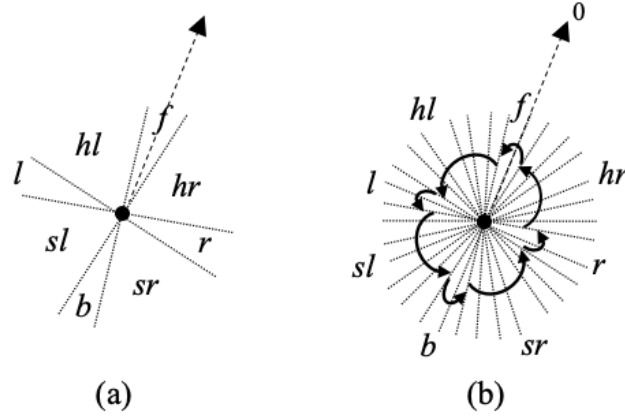


Figure 4.2: (a) The sector model used by Klippel et al. f: front, hl: half left, l: left, sl: sharp left, b: back, sr: sharp right; r: right; hr: half right. (b) An equivalent sector model with $m = 16$ and a dedicated direction relation labelling function.

in Figure 2a: the 2D plane is not equally divided into even and odd sectors and “half left”, “left” and “sharp left” are used instead of just “left”. Their model can be represented by $\mathcal{RDR}_m$ with a greater m and adjusting the l_{dir} function as in Figure 2b (which originates from the mapping strategy in [Lücke et al., 2011]).

4.3.2 Distance Relations

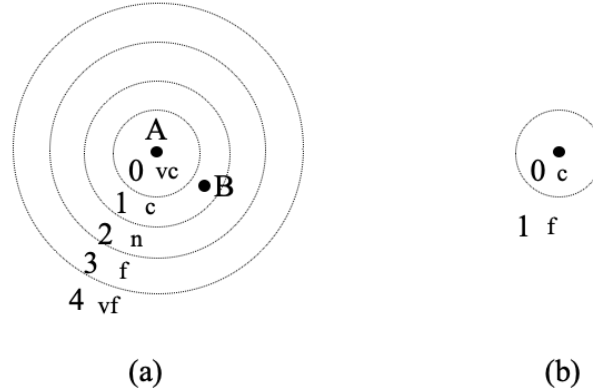


Figure 4.3: (a) The default qualitative distance calculus $\mathcal{QDR}_5$. vc: very_close, c: close, n: neural, f: far, vf: very_far. (b) Another qualitative distance calculus $\mathcal{QDR}_2$. c: close, f: far.

The qualitative distance calculus used in this chapter Qualitative Distance Relation ($\mathcal{QDR}_m$) is first discussed in [Hernandez et al., 1995] to model the qualitative distance relations between places with arbitrary granularity. As in Figure 3a the 2D plane is divided into a circle and several other circular rings based on how far

the points are from the center. These regions are indexed from 0 to $5 - 1$ (i.e. 4). The distance relation between A and B is 1 because B is in Region 1 of A (Figure 3a). The Distance Relation Labelling function is defined to be: $l_{dis} : Z_m \rightarrow L_{dis}$ where $Z_m = \{0, 1, \dots, m - 1\}$ and L_{dis} is a set of distance relation labels in natural language (e.g. $\{very_close, close, neural, far, very_far\}$). Its converse function is defined as $l_{dis}^{-1} : L_{dis} \rightarrow Z_m$. In this chapter, the default qualitative distance calculus used is QDR_5 and the default l_{dis} function (as in Figure 3a) is

- $l_{dis}(0) = very_close$
- $l_{dis}(1) = close$
- $l_{dis}(2) = neural$
- $l_{dis}(3) = far$
- $l_{dis}(4) = very_far$

The value of m and the corresponding l_{dis} function can also be configured by users because in different cases the number of regions, region widths and the natural language terms to describe each region might be also different. For example, people might just use “close” and “far” (i.e. $m = 2$ as in Figure 3b) to describe distance relations rather than the more refined 5 terms used in the default l_{dis} function. Also, the range of “near” varies in different contexts [Richter et al., 2013].

4.3.3 Extracting Qualitative Knowledge from Natural Language Descriptions

It is possible to extract direction relations from route descriptions by applying natural language processing techniques discussed in [Liu et al., 2014]. Such qualitative knowledge extraction is more related to Natural Language Processing and is not the main focus of this chapter. We just assume qualitative knowledge can be extracted from natural language descriptions. For example, the direction relations in “going from A to B you can see C is on the left” and “going from D to B you can see C is on the right” can be denoted as $r_{dir}(A, B, C) = l_{dir}^{-1}(left)$ and $r_{dir}(D, B, C) = l_{dir}^{-1}(right)$ respectively. Since there might be different incoming places (e.g. A and D), the deciding orientation of the direction relation between two places might also be multiple. If the incoming place is unknown as in the description “ B is on A ’s leftside”, the direction relation between A and B has to be denoted as $r_{dir}(null, A, B) = l_{dir}^{-1}(left)$, which actually provides no information regarding the relative position of B to A if no further information is given.

It is also possible to extract distance relations, reachability relations and visibility relations from natural language descriptions. For example, the distance relation in “ A is very close to B ” can be denoted as $r_{dis}(A, B) = l_{dis}^{-1}(very_close)$; the reachability relation in “from A you can go to B ” can be denoted as $r_{rea}(A, B) = t$; and the visibility relation in “at A you can see B ” can be denoted as $r_{vis}(A, B) = t$.

4.3.4 Map Construction

Given the qualitative knowledge of an environment in the form of a set of direction relations, distance relations, reachability relations and visibility relations, we can construct a qualitative place map by applying Algorithm 1. The input of Algorithm 1 is the set of relations $S_{dir}, S_{dis}, S_{rea}, S_{vis}$, and its output is a qualitative place map (QPM); its computational complexity is linear to the number of relations.

Algorithm 1 Qualitative Place Map Construction

```

1: procedure CONSTRUCT( $QPM, S_{dir}, S_{dis}, S_{rea}, S_{vis}$ )
2:   InferAndAddNewRelation( $S_{dir}, S_{dis}, S_{rea}, S_{vis}$ )
3:    $QPM = (V, E)$ 
4:   for  $r \in S_{dir}$  do
5:      $r = (p_1, p_2, p_3, i)$ 
6:     CheckPlace( $r, V$ )
7:      $e = \text{CheckEdge}(p_2, p_3, E)$ 
8:      $e.R_{dir}.add((v_1, i))$ 
9:   for  $r \in S_{dis}$  do
10:     $r = (p_1, p_2, i)$ 
11:    CheckPlace( $r, V$ )
12:     $e = \text{CheckEdge}(p_1, p_2, E)$ 
13:     $e.r_{dis} = i$ 
14:   for  $r \in S_{rea}$  do
15:     $r = (p_1, p_2, t)$ 
16:    CheckPlace( $r, V$ )
17:     $e = \text{CheckEdge}(p_1, p_2, E)$ 
18:     $e.r_{rea} = t$ 
19:   for  $r \in S_{vis}$  do
20:     $r = (p_1, p_2, t)$ 
21:    CheckPlace( $r, V$ )
22:     $e = \text{CheckEdge}(p_1, p_2, E)$ 
23:     $e.r_{vis} = t$ 
24:
25: procedure CHECKPLACE( $r, V$ )
26:   for  $p \in r.places$  do
27:     if  $p.name.type = semantic$  then
28:       if  $p.name \notin V.placeNames$  then
29:          $pn(v) = p.name$ 
30:          $lm(v) = t$ 
31:          $V.add(v)$ 
32:       if  $p.name.type = functional$  then
33:          $pn(v) = p.name$ 
34:          $lm(v) = f$ 
35:          $V.add(v)$ 

```

Before iterating those qualitative relations, we first infer new relations by the following rules and add them to the corresponding sets of relations (Line 2):

- If $r_{dir}(A, B, C) = i$, $r_{dir}(C, B, A) = 4m - i$, which is based on the converse rules discussed in [Hua et al., 2018].
- If $r_{dis}(A, B) = i$, $r_{dis}(B, A) = i$. It is based on the assumption that the distance from A to B is the same as that from B to A .
- If $r_{rea}(A, B) = t$, $r_{rea}(B, A) = t$, which is based on the assumption that if B is reachable from A then A is also reachable from B .
- If $r_{rea}(A, B) = t$, $r_{vis}(A, B) = t$. It is assumed if B is reachable from A then it is also visible from A .
- If $r_{vis}(A, B) = t$, $r_{vis}(B, A) = t$, which is based on the assumption that if B is visible from A then A is also visible from B .

Note that most of the above rules are based on simplified assumptions, which is subject to change in different scenarios. For example, it is possible that Place A is directly reachable from Place B but not visible from Place B because they are connected by a curved path and there is an obstacle inbetween so that Place A is not visible from Place B . Other qualitative spatial reasoning techniques (e.g. composition) can also be used to infer new qualitative relations. However, it is still an open problem whether they are that helpful in our case. For example, the direction relation between Place A , B and D (i.e. $r_{dir}(A, B, D)$) can be inferred by composing $r_{dir}(A, B, C)$ and $r_{dir}(B, C, D)$. But it is possible that D is neither visible nor not directly reachable from B , which means the newly inferred direction relation between A , B , and D may not be helpful in localization or navigation.

Then we iterate the set of direction relations S_{dir} (Line 4). Given a direction relation between three places for example $r_{dir}(p_1, p_2, p_3) = i$ and $i \in Z_{4m}$ (Line 5), check whether p_1 , p_2 , and p_3 have already been represented by vertices in V (Line 6) and the edge from the vertex of p_2 to the vertex of p_3 is in E (Line 7). If the name of a place is semantic (Line 27), we can check whether there is a vertex in V that has the same place name (if yes it is already in V); if the name of a place is functional (Line 32), we just add it to V because so far it is impossible to judge for example whether two intersections are actually referring to the same place. In the QPM, the direction relation between v_2 and v_3 is denoted as (v_1, i) and it will be added as one of the direction relations between v_2 and v_3 in the label of the edge from v_2 to v_3 (Line 8) where v_1, v_2, v_3 are the vertices representing p_1, p_2 , and p_3 respectively.

Similarly, next we go through the set of distance relations S_{dis} (Line 9). Given a distance relation between two places for example $r_{dis}(p_1, p_2) = i$ and $i \in Z_m$ (Line 10), check whether these two places and the edge between them are already in the QPM (Line 11-12). Then add (i) as the distance relation in the label of the edge from v_1 to v_2 (Line 13). The same procedure is applied in iterating through the set of reachability relations S_{rea} (Line 14-18) and visibility relations S_{vis} (Line 19-23).

Note that in a qualitative place map, if no other information is given, the default direction relation between two vertices is $U_{dir} = \{0, 1, \dots, 15\}$ (namely can be any relation in this set); the default distance relation is 2; both of the default reachability relation and visibility relation are f .

Algorithm 2 Qualitative Place Map Construction by Exploration

```

1: procedure CONSTRUCT( $V, E, p$ )
2:    $p.explored = true$ 
3:   AddVertice( $p, V$ )
4:   for  $p_{nb} \in p.neighbors$  do
5:     if  $p_{nb}.explored \neq true$  then
6:       AddEdge( $p, p_{nb}, E$ )
7:       RecordRelations()
8:       Construct( $V, E, p_{nb}$ )
9:       AddEdge( $p_{nb}, p, E$ )
10:      RecordRelations()
  
```

A qualitative place map can also be constructed by exploration if there is a traveller or a robot that can move around and record observations. The exploration strategy can be based on a depth-first search (DFS) [Tarjan, 1972] so that we can traverse all places in a certain order. Start from a random place, we can mark it as “explored” and then move to an arbitrary next place that is directly reachable. If the next place has already been marked as “explored”, go back to the incoming place and choose the next directly reachable place. Record all of the observations about the direction relations, distance relations, reachability relations and visibility relations between places during the exploration. The exploration process can stop if at a certain place all of the directed reachable places have been explored. Algorithm 2 describes this process of starting from a random place p in a place network and the QPM constructed is represented by (V, E) .

4.4 Interactive Localization

Localization is important because the starting position of an agent might be unknown, or it might be lost during the navigation process. It is trivial to locate a user if GPS is available while there are still many GPS-denied environments. In these cases our localization mechanism is still available given the prelearned qualitative place map (QPM) of the environment.

Our localization algorithm (Algorithm 3) is interactive because it is based on the interaction with the agent. Details are as follows:

(1) Initialize a candidate set V_c that includes all landmarks in the given QPM (Line 3-5).

(2) The agent is asked to move to an arbitrary next place and to provide descriptions about its current position (Line 7): it can go to the nearest salient place and tell us the name of the place (*place*); or it can describe the places that are visible from its

Algorithm 3 Interactive Localization

```

1: procedure INTERACTIVELOCALIZATION( $QPM$ )
2:    $(V, E) = QPM$ 
3:   for  $v \in V$  do
4:     if  $lm(v) = t$  then
5:        $V_c.add(v)$ 
6:   while  $|V_c| > 1$  do
7:      $(place, S_{vis}, S_{dir}) = \text{AgentDescription}()$ 
8:      $\text{Locate}(place, S_{vis}, S_{dir}, V, E, V_c)$ 
9:   return  $V_c$ 
10: procedure LOCATE( $place, S_{vis}, S_{dir}, V, E, V_c$ )
11:   if  $place \neq \emptyset$  then
12:     for  $v \in V$  do
13:       if  $pn(v) = place.name$  then
14:          $V_c.removeAll()$ 
15:          $V_c.add(v)$ 
16:       return
17:   if  $S_{vis} \neq \emptyset$  then
18:     for  $v \in V_c$  do
19:        $isValid = true$ 
20:       for  $r \in S_{vis}$  do
21:          $r = (p_1, p_2, t)$ 
22:         if  $r_{vis}(v, v(p_2)) = f$  then
23:            $isValid = false$ 
24:       if  $isValid = false$  then
25:          $V_c.remove(v)$ 
26:   if  $S_{dir} \neq \emptyset$  then
27:     for  $v \in V_c$  do
28:        $isValid = true$ 
29:       for  $r \in S_{dir}$  do
30:          $r = (p_1, p_2, p_3, i)$ 
31:         if  $r_{dir}(v(p_1), v, v(p_3)) \neq i$  then
32:            $isValid = false$ 
33:       if  $isValid = false$  then
34:          $V_c.remove(v)$ 

```

current position (S_{vis}); or it can tell us from which incoming landmark the direction relations between its current place and other visible landmarks (S_{dir}). We assume in each communication only one form of descriptions is provided by the agent.

(3) If the information given by the agent is the name of its current place (Line 11), we check whether it is included as a landmark in the QPM. If yes then we know the current position of the agent and let V_c only contain this landmark (Line 13-16).

(4) If the information given by the agent is the names of the places that are visible from its current position (Line 17), we can remove landmarks from V_c if from it any of the places mentioned are not visible (Line 22). $v(p_2)$ is the vertex that represents p_2 . $r_{vis}(v_1, v_2)$ returns the visibility relation between v_1 and v_2 .

(5) If the information given by the agent is the direction relations between its current position and other landmarks from a certain incoming landmark (Line 26), we can remove landmarks from V_c if from the incoming landmark the direction relations between it and the landmarks do not match the direction relations provided by the agent (Line 31). $v(p_1)$ is the vertex that represents p_1 and $v(p_3)$ is the vertex that represents p_3 . $r_{dir}(v_1, v_2, v_3)$ returns the direction relation between v_2 and v_3 given v_1 as the incoming landmark.

(6) If the number of elements in V_c is greater than 1, go back to step (2) (Line 6).

It is assumed that the orientation of an agent is the same as the orientation from the incoming place to its current place. The incoming landmark must be included in a direction relation description because, as discussed above, if it is not provided the direction relation given is informationless. Our localization algorithm can always succeed to locate an agent if the information provided by the agent is consistent and the agent is not allowed to keep moving in a loop of places (which is supported by the experimental results in Section 7). If the current location of the agent cannot be determined in one query, the agent has to requery. The number of requeries for localization (i.e. the number of interactions with the system minus 1) should not be greater than a certain value (e.g. 2), otherwise the agent will be frustrated. In Section 7 we will explore whether our localization algorithm can always succeed in agent localization and also the average number of requeries.

4.5 Reliable Navigation

4.5.1 Route Planning

Route planning aims to find the “best” route from the origin to the destination though “best” is defined by different criteria in different route planning algorithms. For example, the classical Dijkstra algorithm [Dijkstra, 1959] is to find the route with the shortest length while the shortest most reliable path algorithm [Haque et al., 2006] (which is adaptation of the Dijkstra algorithm) aims to find the most reliable route and route length is used to break ties when two routes share the same unreliability. These two algorithms are also different in number of nodes involved in each assignment of the weighting function (which is designed to assign weight to each edge or a pair of edges): 2 in the Dijkstra algorithm because the weighting function

is only related to one edge; 3 in the shortest most reliable path algorithm because the main weighting function is related to two edges which satisfies that ending node of the first edge is the same as the starting node of the second edge. To avoid the repetitive description on how the Dijkstra algorithm is adapted, we just focus on how to calculate the total cost of a route. The cost function is defined to assign the cost to a route and it indicates the cost of moving from the first node to the last node by making turns or going straight via the intermediate nodes. For example, the cost function used in [Haque et al., 2006] is:

$$cost[v_1, v_2, \dots, v_n] = \sum_{i=1}^{n-2} (ie[v_i, v_{i+1}, v_{i+2}] - 1) \quad (4.1)$$

where $cost[v_1, v_2, \dots, v_n]$ is the cost of a route from v_1 to v_n and $ie[v_i, v_{i+1}, v_{i+2}]$ is the number of turns that are instruction equivalent to the turn from the edge (v_i, v_{i+1}) to the edge (v_{i+1}, v_{i+2}) (or more formally the number of vertices v that satisfy

$$r_{dir}(v_i, v_{i+1}, v) = r_{dir}(v_i, v_{i+1}, v_{i+2})). \quad (4.2)$$

By contrast, the cost function used in our route planning algorithm $cost^*[v_1, v_2, \dots, v_n]$ is defined as

$$\sum_{i=1}^{n-2} \left(\frac{\log_{a_1}(ie[v_i, v_{i+1}, v_{i+2}] - 1 + a_1)}{\log_{a_2}(lm_{\#}[v_{i+2}] + a_2)} \right) + qd[v_1, v_2, \dots, v_n] \quad (4.3)$$

where $lm_{\#}[v_{i+2}]$ is the number of landmarks that are visible at v_{i+2} and $qd[v_1, v_2, \dots, v_n]$ is the qualitative distance from v_1 to v_n (or in a more formal form

$$qd[v_1, v_2, \dots, v_n] = \sum_{i=1}^{n-1} (r_{dis}(v_i, v_{i+1}))). \quad (4.4)$$

The above cost function is designed based on the assumption that one route is less costly if its distance is shorter, with less ambiguous turns, or with more landmarks. Our route planning algorithm considers not only the distance between two nodes and the unreliability of each turn but also the number of landmarks involved in a certain route because it is less possible for an agent to be lost and the route instructions can be simpler with greater number of landmarks. Its cost function consists of two parts: the first part is the sum of the ratios of the unreliability of the turns to the number of landmarks. For each turn, it is better to calculate the ratio of these two factors because they might be relevant (e.g. greater number of landmarks might incur the increase of the unreliability as in Figure 4). Note that logarithmic functions are applied to smooth the influence of much greater values of the unreliability of a turn and the number of landmarks. a_1 and a_2 are adjustable constants and their values can be decided by exploratory experiments as discussed in Section 7. The second part is the qualitative distance from the origin to the destination.

Other details of our route planning algorithm are omitted because the only difference between our algorithm and the shortest most reliable path algorithm [Haque

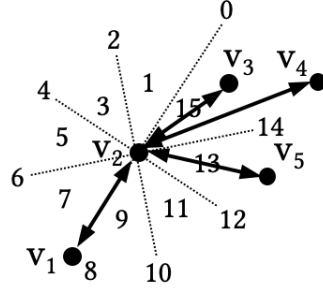


Figure 4.4: Assuming v_4 is a landmark and is visible from v_3 . When moving from v_1 to v_3 via v_2 , v_4 is helpful as a landmark that might be used to locate v_3 . However, v_4 can also increase the unreliability because it is the turn of (v_1, v_2, v_3) is instruction equivalent to the turn of (v_1, v_2, v_4) .

et al., 2006] is how the cost function is defined. The discussion about the computational complexity is also similar (namely $O(|E|^2)$).

Algorithm 4 Reliable Navigation

- 1: **procedure** NAVIGATION(QPM, v_d)
 - 2: $v_c = \text{InteractiveLocalization}(QPM)$
 - 3: $route = \text{FindTheMostReliableRoute}(QPM, v_c, v_d)$
 - 4: **for** $(v_1, v_2) \in route$ **do**
 - 5: $\text{GenerateInstructions}(v_1, v_2)$
-

4.5.2 Navigation System

As in Algorithm 4, by applying the interactive localization mechanism discussed in Section 4 our navigation algorithm can navigate an agent without knowing its current place (Line 2).

Given the qualitative place map, the current vertex v_c and the destination vertex v_d , the route planning algorithm discussed in the previous subsection can be used to find the most reliable route (Line 3).

Extract the edges in the most reliable route in sequence and generate route instructions (Line 5). A template-based route generation strategy is used in our system while more sophisticated strategies can be found in [Oßwald et al., 2014; Daniele et al., 2017]. For example, given v_1 as the current place and v_2 the next place to go, first describe landmarks that are visible from v_1 in the form of “at $pn(v_1)$ you can see...” and then indicate whether to go straight or take turns to get to v_2 based on the direction relation between v_1 and v_2 ; next describe landmarks that are visible from v_2 by using the same template “at $pn(v_2)$ you can see...”. This mechanism is to ensure that each time the agent arrives at a new place, descriptions of visible landmarks from the new place will be provided to check whether the agent is at the correct place, which is better than [Haque et al., 2006; Westphal and Renz, 2011] where an

agent can only realize they are in the wrong way when there is no valid navigation instruction to follow.

Whenever the agent reports it is lost, just restart the navigation procedure.

4.6 Case Study

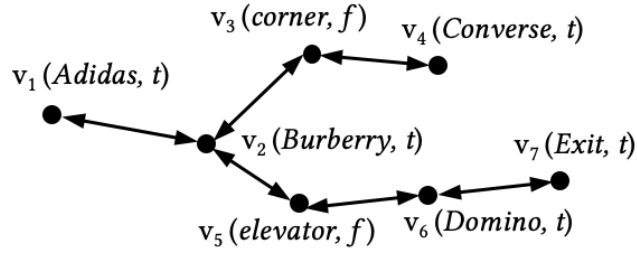


Figure 4.5: A qualitative place map constructed from two route descriptions.

In this section we will describe how to solve the problem mentioned in the introduction by using our navigation system. As shown in Figure 5 a qualitative place map can be constructed based on the qualitative knowledge extracted from the two descriptions and applying the inference rules discussed in Section 3.4. In the QPM, each vertex is labelled with a tuple where the first element is the name of the place and the second element indicates whether it is a landmark. The labels of edges are as follows:

- $e_{12} = (v_1, v_2, t, t, \{(null, l_{dir}^{-1}(left))\}, 2)$
- $e_{21} = (v_2, v_1, t, t, \{(null, U_{dir})\}, 2)$
- $e_{23} = (v_2, v_3, t, t, \{(v_1, U_{dir})\}, 1)$
- $e_{32} = (v_3, v_2, t, t, \{(v_4, l_{dir}^{-1}(right))\}, 1)$
- $e_{34} = (v_3, v_4, t, t, \{(v_2, l_{dir}^{-1}(left))\}, 3)$
- $e_{43} = (v_4, v_3, t, t, \{(null, U_{dir})\}, 3)$
- $e_{25} = (v_2, v_5, t, t, \{(null, U_{dir})\}, 0)$
- $e_{52} = (v_5, v_2, t, t, \{(v_6, l_{dir}^{-1}(left))\}, 0)$
- $e_{56} = (v_5, v_6, t, t, \{(v_2, l_{dir}^{-1}(right))\}, 1)$
- $e_{65} = (v_6, v_5, t, t, \{(v_7, l_{dir}^{-1}(left))\}, 1)$
- $e_{67} = (v_6, v_7, t, t, \{(v_5, l_{dir}^{-1}(right))\}, 0)$
- $e_{76} = (v_6, v_5, t, t, \{(null, U_{dir})\}, 0)$

Applying the routing planning algorithm discussed in Section 5, it is not difficult to know that the best route from “Converse” (v_4) to “Exit” (v_7) is $v_4v_3v_2v_5v_6v_7$. An agent can follow the route instructions generated by the template-based strategy discussed in Section 5.2:

- From Converse go to the corner and turn right to Burberry. At the corner you can see Burberry.
- At Burberry you can see Adidas. From Burberry go to the elevator and turn right to Domino. At the elevator you can see Domino.
- From Domino you can see Exit. From Domino go to Exit.

It is not difficult to recognize the templates used in the above route instructions. Generating more human-like instructions are left as future work.

4.7 Experiments

We evaluate the whole localization and navigation process in simulated environments. Qualitative place maps (QPMs) are generated by the Exploration Algorithm (Algorithm 2) from graphs of places with different configurations in terms of number of vertices and edges. Specifically, the four configurations being used are 200 vertices & 200 edges, 200 vertices & 300 edges, 500 vertices & 500 edges and 500 vertices & 750 edges. The vertices may be recognizable landmarks or non-salient places. The landmark vertices may be reachable by the agent (e.g. a bookshop) or only visible but unreachable (e.g. an advertisement board hanging on the ceiling). The qualitative place maps are built with the default qualitative spatial models used in this chapter: $\mathcal{RDR}_4$ and $\mathcal{QDR}_5$.

Then, we simulate the guiding process between a navigator with the qualitative place maps and an agent who aims to travel from randomly selected starting points to destination points, which corresponds to the situation where based on prior knowledge one person can localize and navigate the other via remote communications. Note that it is assumed the starting point is unknown to both the navigator and the agent at first. The navigator needs to locate the agent based on its qualitative place map and the observations provided by the agent. Another assumption we made here is that the agent will correctly interpret and execute route instructions, which means, the agent can only make mistakes when facing with ambiguous turns.

4.7.1 Evaluation of Interactive Localization

In this subsection, our interactive localization algorithm is evaluated by analyzing the number of requeries needed for obtaining the correct location of the agent. In our localization mechanism, each requery results in a movement of the agent, therefore it is important to keep the number of requeries small. Table 4.1 demonstrates the percentage of requery amounts performed over 10000 testing cases in each place

graph. The accuracy of our qualitative localization algorithm is 100% (i.e. the agent can always be located). More than 80% of the localization processes need zero or one requery, and more than 90% of the localization processes need no more than 2 requeries. Based on the experimental results, it is safe to claim our method works reasonably well though there is no similar purely qualitative localization algorithm to compare with.

It can also be observed from the results that with the increase of number of vertices and edges, the percentage of tests with more than 1 requeries slightly increases. One possible reason is that more requeries are required to resolve the ambiguity brought by the increased vertices and their connections.

Table 4.1: Evaluation results of interactive localization. ‘v’ represents the number of vertices in the place graph; ‘e’ represents the number of edges in the place graph; ‘r’ is the number of requeries; ‘ r_{ave} ’ is the average number of requeries made per guidance.

	v=200,e=200	v=200,e=300	v=500,e=500	v=500,e=750
$r = 0$	38.01%	37.32%	34.19%	34.31%
$r = 1$	49.13%	49.05%	52.12%	47.60%
$r = 2$	6.49%	8.03%	7.14%	10.53%
$r > 2$	6.37%	5.60%	6.55%	7.56%
r_{ave}	0.90	0.87	0.92	0.99

4.7.2 Evaluation of Reliable Navigation

Table 4.2: Evaluation results for different cost functions on a place graph with 200 vertices and 200 edges. “#instructions” is the number of instructions provided by the navigator to the agent. “#wrong turns” is the number of incorrect turns made by the agent because of ambiguous turns. “distance” is the travelling distance (in units of length) of the agent including wrong movements and movements by the interactive localization mechanism. All the data in the following 4 tables are averaged per trip. “SMR” is short for the shortest most reliable path algorithm and “ILRN” is short for our interactive localization and reliable navigation algorithm.

	Dijkstra	SMR	ILRN
# instructions	9.27	9.14	8.90
# wrong turns	0.75	0.72	0.61
distance	4.03	4.03	4.10

In this subsection, we evaluate the navigation process which includes interactive localization and route planning. We monitor the number of instructions given by the navigator (namely the number of places from the origin to the destination), the dis-

Table 4.3: Evaluation results for different weight configurations on a place graph with 200 vertices and 300 edges.

	Dijkstra	SMR	ILRN
# instructions	7.99	7.94	7.68
# wrong turns	1.00	0.93	0.71
distance	3.53	3.43	3.54

Table 4.4: Evaluation results for different weight configurations on a place graph with 500 vertices and 500 edges.

	Dijkstra	SMR	ILRN
# instructions	11.74	11.60	11.28
# wrong turns	0.76	0.71	0.61
distance	5.23	5.18	5.12

Table 4.5: Evaluation results for different weight configurations on a place graph with 500 vertices and 750 edges.

	Dijkstra	SMR	ILRN
# instructions	10.55	10.34	9.54
# wrong turns	1.42	1.30	0.95
distance	4.86	4.72	4.50

tance that the agent actually travels in the simulation and the number of wrong turns the agent makes during the travel. We compare our results with the Dijkstra algorithm and also the shortest most reliable path algorithm (SMR) described in [Haque et al., 2006]. Since it is assumed that the exact distances between places are unknown to the navigator, the distances used in the three algorithms are all qualitative distances with granularity of 5. However, the total travel distance for measuring the performance of the three algorithms are exact distances from the ground truth of the place graphs. In the evaluation below, the two constants in our cost function (i.e. a_1 and a_2 in Equation (3)) are set to 4 and 2 respectively based on the exploratory experiments on our algorithm.

Dijkstra, SMR, and our interactive localization and reliable navigation (ILRN) algorithm are all applied in 10000 randomly generated test cases in each of the same four graphs (as those in Section 7.1). The average number of instruction, average number of wrong turns and average distance are calculated for comparison. As shown in Table 4.2, 4.3, 4.4 and 4.5, we can observe that our algorithm provided the smallest number of instructions among the three algorithms in all four different configurations. By using our algorithm the average number of wrong turns is much smaller than those by using the other two algorithms, which demonstrates our algo-

rithm is capable of finding the most reliable route and can reduce the frequency of getting lost. Note that in two situations wrong turns will be detected: 1) descriptions from the agent are different from the expectation of the navigator; 2) the agent cannot observe the landmark according to the route instruction given by the navigator.

For the relatively large graphs with 500 vertices, by using our algorithm the smallest travel distance can be achieved. However, for graphs with 200 vertices, the physically shortest paths cannot be found by using our algorithm. But as discussed above, our method is still preferable when compared to the other two algorithms because it can largely reduce the number of instructions from the navigator and the number of wrong turns.

4.8 Conclusion and Future Work

In this chapter, we have provided a series of mechanisms that enable localization and navigation in GPS-denied environments given only qualitative knowledge of places. This is motivated by human communication and aims to assist human localization and navigation in an intuitive and interactive way. The qualitative knowledge between places (landmarks or not) is extracted from natural language descriptions and stored in a directed graph called qualitative place map where the vertices represent places and edges are labelled with different categories of qualitative relations between places that might be helpful in localization and navigation. Based on qualitative place maps, an interactive localization algorithm has been described that can locate agents based on eliminating landmarks that cannot satisfy the observations provided by agents; and a reliable navigation system has been introduced which (different from previous methods) also takes number of landmarks and qualitative distances into consideration. In the experiments, we demonstrate the usefulness of our localization algorithm as well as the improvement in reliable route calculation compared to previous navigation systems. More importantly, those previous navigation systems might not be useful in cases where the coordinates of places are not given. With our qualitative localization algorithm, some previous GPS-based navigation systems can also be used in GPS-denied environments by performing localization based on the observations from the agent or user.

In the future, we plan to combine our system with more sophisticated natural language processing techniques and extend its practical usefulness by applying it in more real-world scenarios.

Integrating Qualitative Spatial Relations with Learning-based Methods

“Change is the only constant in life.” – Heraclitus¹

In order to be trusted by humans, Artificial Intelligence agents should be able to describe rationales behind their decisions. One such application is human action recognition in critical or sensitive scenarios, where trustworthy and explainable action recognizers are expected. For example, reliable pedestrian action recognition is essential for self-driving cars and explanations for real-time decision making are critical for investigations if an accident happens. In this regard, learning-based approaches, despite their popularity and accuracy, are disadvantageous due to their limited interpretability.

This chapter presents a novel neuro-symbolic approach that recognizes actions from videos with human-understandable explanations. Specifically, we first propose to represent videos symbolically by qualitative spatial relations between objects called *qualitative spatial object relation chains*. An object relation chain is a chain of object relations to represent a series of spatial relation changes between objects. We further develop a *neural saliency estimator* to capture the correlation between such object relation chains and the occurrence of actions. Given an unseen video, this neural saliency estimator is able to tell which object relation chains are more important for the action recognized. We evaluate our approach on two real-life video datasets, with respect to recognition accuracy and the quality of generated action explanations. Experiments show that our approach achieves superior performance on both aspects to previous symbolic approaches, thus facilitating trustworthy intelligent decision making. Our approach can be used to augment state of the art learning approaches with explainabilities.

¹Greek philosopher.

5.1 Introduction

While learning-based approaches are very popular [LeCun et al., 2015], criticisms remain over their reliability and explainability in making high-stake decisions [Rudin, 2019]. Video action recognition (i.e. recognize which action occurs in a video) is one such typical task. For example, failures in recognizing pedestrian actions by self-driving cars may lead to critical safety issues and those failures could be caused by incorrect reasoning. Most state-of-the-art action recognition approaches [Lin et al., 2019; Tran et al., 2019] cannot be easily understood or justified by humans as they use neural networks in a black-box fashion [Rudin, 2019].

We argue that a video action recognizer is trustworthy or at least its trustworthiness is measurable if it is able to justify its decisions in a human-understandable way. Such justifications can be the existence of certain objects, object relations, or changes in object relations [Zhuo et al., 2019]. For example, in soccer matches, *offside* occurs if there is a particular change in positional relations between ball and players.

We will investigate action recognition and explanation by qualitative spatial relations; it is not difficult to extend our approach and relevant discussions to non-qualitative or non-spatial relations. A spatial relation is a relation that is relevant to the spatial properties of objects. For example, connectivity, direction, size, distance, moving direction or moving speed. Spatial relations between objects are essential in many real-world actions and one typical example is shown in Figure 5.1a. A relation is qualitative if it is a symbol with specific meaning (e.g. “close”) rather than a numeric value (e.g. “0.1 meter away from”). We choose qualitative spatial relations for two reasons: first, qualitative spatial relations are intuitive and close to human cognition and thus being human-understandable [Chen et al., 2015]; second, it is possible to detect such relations without involving much human labor [Gatsoulis et al., 2016].

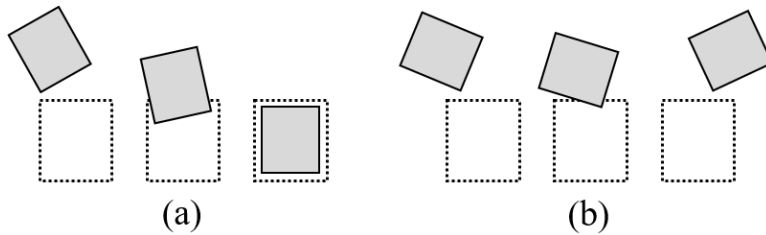


Figure 5.1: (a) Essential spatial relations when performing the action of *car parking*: at first, the car (grey rectangle) is disconnected from the parking space (dotted rectangle) and then they overlap; in the end, the car is included in the parking space. (b) An accidental interaction between the car and the parking space.

Qualitative spatial relations are formally defined in qualitative spatial calculi [Renz and Nebel, 2007; Cohn and Renz, 2008]. One of the most widely applied qualitative spatial calculi is Region Connection Calculus (RCC) [Randell et al., 1992b], which consists of topological relations (or connectivity relations) between regions (e.g. disjoint or overlapping). The *car parking* example in Figure 5.1a can be described as a chain of objects and their relations: $o_1-o_2-dc-po-pp$, where o_1 stands for

the car (grey rectangle), o_2 stands for the parking space (dotted rectangle), and dc , po and pp are RCC relations that indicate two regions are disconnected, partially overlapping, and included respectively.

In this chapter, we aim to achieve *explainable action recognition* by exploring critical qualitative spatial relations or relation changes between objects. Our motivation is to provide explanations for some of the results from “black-box” approaches. Our main contributions are threefold: (1) we introduce a novel symbolic action representation based on qualitative spatial object relation chains (or object relation chains in short); based on this representation, (2) we develop a *neural saliency estimator* that measures the correlation between object relation chains and actions in different contexts; and (3) we generate *action explanations* by identifying salient object relation chains. Based on these innovations, our action recognizer achieves better performance than previous symbolic approaches in both recognition accuracy and the quality of generated action explanations.

The rest of this chapter is organized as follows: in the next section, we introduce related work; in Section 3, we describe the problem to solve; in Section 4, we introduce our approach in detail; in Section 5, we evaluate our approach by experiments; and in the last section, we conclude the chapter and discuss future work.

5.2 Related Work

In this section, we give a brief introduction to previous attempts to apply qualitative spatial calculi to recognize or explain actions.

5.2.1 Qualitative Spatial Action Description

In this subsection, we will introduce previous work that utilizes qualitative spatial relations to describe actions. In [Dubba et al., 2015], actions like airplane arrival or departure are represented as changes in the RCC relations between airplanes, vehicles, and ground zones. Such representations are then used to recognize actions in unseen videos [Dubba et al., 2015]. In [Tayyub et al., 2014], the occurrence of various object relation chains like o_1-o_2-dc , $o_2-o_3-dc-po$, $o_1-o_3-dc-po-dc$ has been counted and used as features of machine learning approaches to recognize daily activities. A similar but more sophisticated representation strategy is applied in [Sridhar et al., 2010; Duckworth et al., 2019] to achieve unsupervised action clustering. In [Young and Hawes, 2015], four qualitative spatial calculi RCC, *STAR*, QDC, and QTC have been applied in [Young and Hawes, 2015] to describe the qualitative movement states of agents in a RoboCup soccer simulator. Their qualitative spatial relation based system is able to learn soccer player actions like kick or dash.

In most previous work, the set of candidate qualitative spatial relations is decided by human knowledge. Namely, most previous researchers manually determine which set of qualitative spatial calculi is chosen to describe actions. An experimental study has been conducted in [Sridhar et al., 2011] where the performance of applying different categories of qualitative spatial relations in five aspects in three differ-

ent tasks event classification, clustering, and detection has been compared. These five categories are: topology, direction, relative size, relative speed, and relative trajectories. According to their experimental results, the best performance is always achieved by a combination of all the five categories. Three categories of representations qualitative temporal, qualitative spatial, and quantitative spatial have been applied in [Tayyub et al., 2014] to recognized daily activities. The best recognition accuracy is also achieved by combining all three categories of representations based on the results of their ablation study.

A wide range of qualitative spatial calculi has been implemented in a set of published tools like SparQ [Dylla et al., 2006], the Qualitative Algebra Toolkit (QAT) [Condotta et al., 2006], Generic Qualitative Reasoner (GQR) [Gantner et al., 2008], and QSRLib [Gatsoulis et al., 2016]. Based on these tools, quantitative descriptions of objects can be automatically transformed into qualitative spatial relations between objects.

5.2.2 Qualitative Action Recognition Explanation

Given a set of positive and negative examples (namely videos that are relevant or irrelevant to a specific action), [Dubba et al., 2015] has proposed an inductive logic programming framework that learns which object relation chains explain actions like *airplane arrival* or *airplane departure*. Or, in their terminology, a pipeline of methods is proposed to learn relational models for actions. Their solution involves manual annotation effort in labelling data, which can be problematic when dealing with large-scale data: *Deictic Supervision* requires a human to label roughly where and when actions occur in videos; a predefined object hierarchy is required; and some of the parameters (e.g. the weights of positive and negative examples) are decided manually. Their algorithm is also very sensitive to the initial positive example selected and is very likely to provide multiple predictions for a single video, which might lead to many false positive errors in action recognition.

In [Zhuo et al., 2019], actions are qualitatively (or symbolically) described as state transitions. A state can be an object in a certain status (e.g. an open microwave) or a pair of object in a particular relationship (e.g. hand holding cup). The qualitative action descriptions are used as “prior knowledge” in action recognition and explanation. Specifically, first an object detector and state detector are trained; second, these two detectors are applied to obtain the state transitions of each object or object pair in each video; and third, an action is detected if a state transition satisfies the description of an action and this state transition is used as the explanation for the recognized action. They re-annotate the CAD120 dataset [Koppula et al., 2013] and experiment on multi-action-label videos from this new dataset. The experimental results show their approach can detect most actions (high recall) but also cause many false positive recognition errors (low precision).

5.2.3 Attention for Explanation

Attention is a mechanism that learns to assign greater weights to more important features [Vaswani et al., 2017] and has become popular and widely applied in deep learning and computer vision in recent years. Attention mechanism has been proved to be useful in improving the performance of recognition models [Selvaraju et al., 2017; Jetley et al., 2018; Fukui et al., 2019]. The intuition of attention is similar to human cognition in the sense that it helps recognition models find which parts of input data are more important and pay more “attention” on these deciding parts. Technically, by learning, an attention layer is able to assign greater weights to features that are critical in accurate recognition. By visualizing which features are more “attended”, attention is able to improve the explainability of image classification neural networks [Selvaraju et al., 2017; Fukui et al., 2019].

5.3 Problem Statement

In this chapter, we focus on the problem of *explainable action recognition*. Assume we are given a set of actions $\mathcal{A}$ and a set of videos $\mathcal{V}$. Given an unseen video $v \in \mathcal{V}$ as input, an explainable action recognizer outputs an action label $a \in \mathcal{A}$ and an explanation $\mathcal{X}$. The action label a denotes “action a occurs in video v ”, which is supported by the explanation $\mathcal{X}$. Note that, to simplify the following discussions, we assume there is only one action occurs in each video, while it is not difficult to adapt our approach to multi-action videos. If there are multiple action labels in a video, sophisticated action localization models [Kalogeiton et al., 2017; Köpüklü et al., 2019] can be utilized to locate each individual action and then our approach can be used to explain located intervals labeled with single actions.

Distinct from [Selvaraju et al., 2017] and [Fukui et al., 2019], whose explainability is achieved by visualizing numeric weights of features, explanations generated by our approach are in the form of salient object relation chains. In natural language, the relationship between a recognized action and the set of generated salient object relation chains is “this action is recognized because we believe this set of observed object relation chains is caused by this action”. Generating such explanations is challenging because it requires an agent to reason about the causality between actions and qualitative spatial relation changes with noisy observations. Specifically, given a video v with an action label a , some of the objects detected in the video might be irrelevant to action a ; and some of the object relation chains in v might be caused by accidental interactions (Figure 5.1b) between objects [Sridhar et al., 2011b; Sokeh et al., 2014] or actions other than a . It is also possible that some of the object relation chains are noisy or corrupted because they involve mislabelled objects or relations [Dubba et al., 2015]. Videos will be labeled with *null* if no obvious human-recognizable action occurs, which might make this problem even more challenging.

5.4 Proposed Approach

As shown in Figure 5.2, our proposed approach consists of three main components: *input*, *saliency estimation*, and *outputs*. In each of the three components, there is an innovative module that plays a key role: *object relation chain extractor* in input, *neural saliency estimator* in saliency estimation, and *explanation generator* in outputs.

We divide each video into several clips [Bleser et al., 2015] and manipulate the object relation chains in each video clip separately. The object relation chain extractor detects the set of object relation chains in each video clip. Then each set of extracted object relation chains is k-hot encoded into a vector so that it can be manipulated by a neural network. Each dimension of the k-hot vector stands for an object relation chain, and its value is 1 if its corresponding object relation chain appears in a video clip (0 otherwise). These k-hot encoded vectors constitute a matrix, which is the input of our neural saliency estimator. The output is a saliency score matrix whose values indicate which object relation chains are more salient in each video clip; and a saliency weighted object relation chain matrix so that the saliency scores of object relation chains can be utilized to boost performance in action recognition. With the saliency weighted object relation chain matrix as input, action in the video is recognized by several fully connected layers and their corresponding activation functions. The recognized action is justified by the explanations generated by our explanation generator. Details will be described in the following.

5.4.1 Object Relation Chain Extraction

Assume we are given a set of qualitative spatial relations $\mathcal{R}$ that are defined in a certain set of qualitative spatial calculi, and a set of object categories $\mathcal{O}$. Each object relation chain c is a $2 + n$ tuple $o_1-o_2-r_1-r_2-...-r_n$, where $o_1, o_2 \in \mathcal{O}$, $r_1, r_2, ..., r_n \in \mathcal{R}$, $n \in \mathbb{Z}^+$, and $r_i \neq r_j$ if $j = i + 1$. It is actually a subset of the representation introduced in [Duckworth et al., 2019] that focuses on a consecutive set of changes in the relation between the same object pair.

The set of all possible object relation chains is definite given $\mathcal{O}, \mathcal{R}$, and n_r^* , where n_r^* is the maximum number of relations an object relation chain can have. An object relation extractor is a function $\mathcal{E}$ that maps a video clip into the set of object relation chains in it. I.e., $\mathcal{E}(v)$ denotes all object relation chains that can be observed in a video clip v . It's not difficult to implement an object relation chain extractor since state-of-the-art computer vision algorithms [Wang et al., 2019; Li et al., 2019; Tan et al., 2020] have already been able to detect and track object locations in a very accurate manner. Those object locations are mainly in the form of minimal bounding boxes, and qualitative spatial relations between objects are estimated by those between their corresponding bounding boxes. Relations between bounding boxes can be automatically calculated because most qualitative spatial relations have geometric definitions [Gatsoulis et al., 2016]. Given the qualitative spatial relations in each frame of a video clip, object relation chains are extracted by looking for pairs of objects that are in different qualitative spatial relations in two consecutive frames.

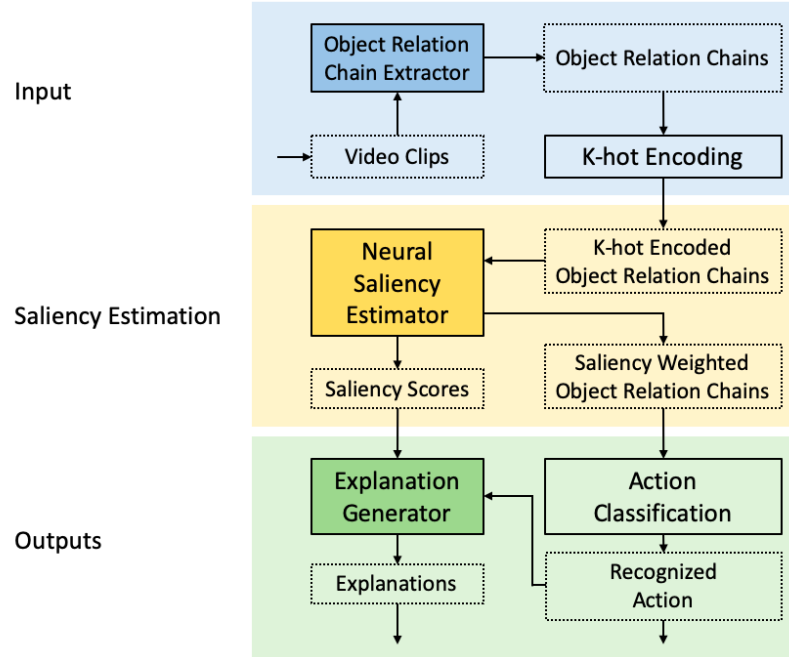


Figure 5.2: An overview of our approach: the input is a set of video clips and the outputs are a recognized action and its explanations. Our approach consists of three components: input, saliency estimation, and outputs. Solid rectangles are modules and dotted rectangles are data in different forms. Our innovative modules in each component are object relation chain extractor (blue), neural saliency estimator (yellow), and explanation generator (green).

5.4.2 Neural Encoding of Object Relation Chains

Given a video clip ν , we propose to represent its extracted object relation chains $\mathcal{E}(\nu)$ as a k -hot vector, $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$, where N is the total number of distinct object relation chains given $\mathcal{O}, \mathcal{R}$, and n_r^* ; $x_i = 1$ iff the corresponding object relation chain is observed in ν , otherwise, $x_i = 0$. A video that is divided into n_ν clips $\{\nu_1, \nu_2, \dots, \nu_{n_\nu}\}$ is represented as a $n_\nu \times N$ matrix $M^\nu = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_\nu}]$, where $\mathbf{x}_i$ is the corresponding k -hot vector of ν_i . In this way, we represent object relation chains in a video as a real-valued fixed-size matrix, which enables learning neural models on symbolic video representations. Our representation is compact because N is the number of all object relation chains rather than the number of all possible chains given a specific maximum length.

5.4.3 Saliency Estimation and Action Recognition

As aforementioned, agents usually have a noisy observation of the target action in the presence of interfering or incorrect object relation chains. In order to achieve robust recognition in a noisy environment, we start from the observation that *the semantics of an object relation chain is context-dependent*. Namely, an agent has to jointly

consider the co-occurrence of multiple object relation chains, in order to understand actions correctly in context. One such example is shown in Figure 5.1: let o_1 be *car* and o_2 be *parking space*, if it occurs alone, the object relation chain $o_1-o_2-dc-po$ is very likely caused by an accidental interaction (Figure 5.1b) while its co-occurrence with another object relation chain $o_1-o_2-po-pp$ is a strong indicator of *car parking* (Figure 5.1a).

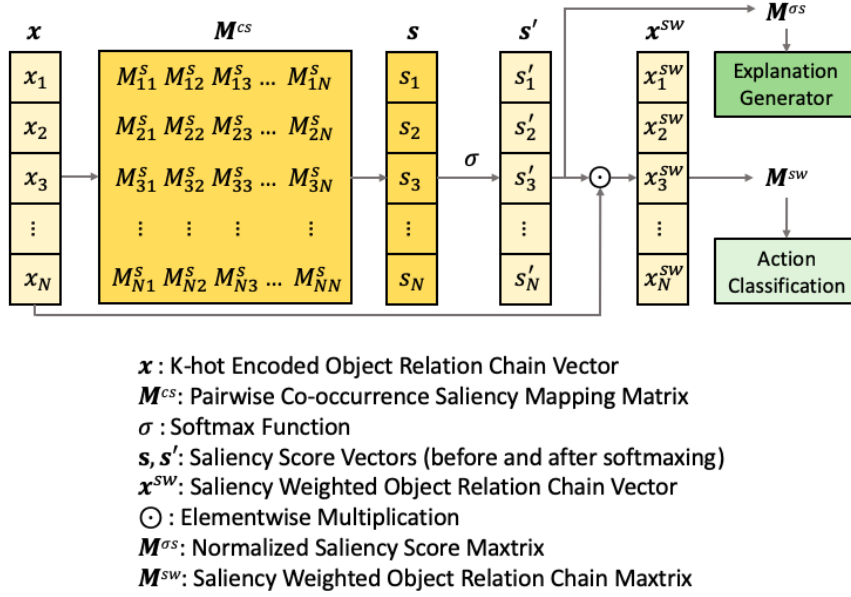


Figure 5.3: Our neural saliency estimator will be applied to each object relation chain vector in $\mathbf{M}^v$: each k-hot encoded $\mathbf{x}$ is mapped into a saliency score vector $\mathbf{s}$ by our pairwise co-occurrence saliency mapping matrix $\mathbf{M}^{cs}$. Then $\mathbf{s}$ is normalized by the softmax function. The normalized saliency score vector $\mathbf{s}'$ is used to constitute the normalized saliency score matrix $\mathbf{M}^{\sigma s}$ and produce the saliency weighted object relation chain vector $\mathbf{x}^{sw}$ by elementwise multiplied with $\mathbf{x}$. The saliency weighted object relation chain matrix $\mathbf{M}^{sw}$ consists of $\mathbf{x}^{sw}$ from each video clip and it is the input of our action classifier.

Following this observation, inspired by the attention mechanism applied in [Selvaraju et al., 2017; Jetley et al., 2018; Fukui et al., 2019], we propose an innovative *neural saliency estimator*, which measures the importance of observed object relation chains in action recognition, while taking into account other co-occurring object relation chains. It is actually a neural network layer that takes the encoding of observed object relation chains in a video clip as input and learns to re-weight them by considering exhaustively all the pair-wise co-occurrence of these chains.

As discussed above, the saliency of an object relation chain is relevant to all co-existing chains. This relevance is described in Equation 5.1, where $s_i \in \mathbf{s}$ is the saliency score of its corresponding object relation chain and a weighted sum of all co-existing chains. A higher M_{ij}^s indicates the occurrence of object relation chains i and j is highly correlated while lower values show such co-occurrence is likely

accidental thus being noise for recognition.

$$s_i = \sum_{j=1}^N M_{ij}^s x_j \quad (5.1)$$

So, as shown in Figure 5.3, given N distinct object relation chains, the k-hot encoded object relation chain vector x is mapped into the saliency score vector s by multiplying with a $N \times N$ matrix M^{cs} . Namely, $s = M^{cs}x$. The matrix M^{cs} is named as the *pairwise co-occurrence saliency mapping matrix*. Then the saliency score vector s will be exponentially normalized into s' by the softmax function σ as in Equation (5.2), where $s_i \in s$ and $s'_i \in s'$.

$$s'_i = \sigma(s_i) = \frac{e^{s_i}}{\sum_{j=1}^N e^{s_j}} \quad (5.2)$$

In the last step, the k-hot encoded object relation chain vector x will be re-weighted by the normalized saliency scores. As in Equation (5.3), each element x_i^{sw} in the saliency weighted object relation chain vector x^{sw} is the result of its correspondence in the original object relation chain vector x_i multiplied by the corresponding normalized saliency score s'_i .

$$x_i^{sw} = s'_i x_i \quad (5.3)$$

Given a $n_v \times N$ object relation chain matrix M^v , the outputs are a $n_v \times N$ normalized saliency score matrix M^{cs} and a $n_v \times N$ saliency weighted object relation chain matrix M^{sw} . The former will be used for explanation generation while the latter will be the input of action classification. We use the saliency scores to generate action explanations. In terms of action recognition, we apply multiple fully connected layers to acquire the final action prediction.

Our neural-based saliency estimator and action classifier are trained together by minimizing the *cross entropy loss* in action classification, which is based on the assumption that a better saliency estimator leads to a more accurate action classifier. This assumption will be proved by experiments. Like normal neural networks, our hyper-parameters are *learning rate*, *number of epochs*, and *batch size*, whose values are decided by a thorough grid search and the goal of the grid search is to find the combination of values that achieves the best classification accuracy.

5.4.4 Explanation Generation

Our explanation generator produces action explanations in the form of salient object relation chains (i.e. object relation chains with the highest saliency scores). Based on such an explanation generator, our action recognizer is explainable with salient object relation chains as our explanations. Such explanations are human-understandable because, for example, they can be easily translated into natural language in the form of “this action is performed in the video because we believe this set of observed object relation chains is caused by it” and critically, the spatial relations composing

these chains are intuitively understandable [Shariff et al., 1998].

Given a video v , let M^{σ_s} be the output of our neural saliency estimator and a be the output of the action classifier. Remember, each element in M^{σ_s} corresponds to the saliency score of a certain object relation chain in a specific video clip. So with M^{σ_s} , we can construct a function $\mathcal{S}$ that maps an object relation chain c , a video clip v , and an action a into a saliency score. For example, $\mathcal{S}(c, v, a) \in [0, 1]$ is a real number that indicates how salient c is if the action performed in v is predicted as a .

Our explanation for a certain action recognition is generated by the rule defined in (5.4), where v is a video clip, and $\mathcal{S}_k(\mathcal{E}(v), v, a)$ denotes the k th highest saliency score among all the saliency scores of object relation chains in $\mathcal{E}(v)$. Namely, we will select the top- k salient object relation chains (i.e. object relation chains with top- k highest saliency scores) in $\mathcal{E}(v)$ as our explanation for a single video clip. The explanation for the whole video is an orderly combination of the explanations for its video clips.

$$c \in \mathcal{X}(v, a) \Leftarrow \mathcal{S}(c, v, a) \geq \mathcal{S}_k(\mathcal{E}(v), v, a) \quad (5.4)$$

Our approach can also be applied to enhance the explainability of state-of-the-art action recognition models. If the prediction from our approach is the same as that from those approaches, the generated explanation describes the reasoning of the action recognition process. If not, their prediction is preferred and our explanation helps further examination.

5.4.5 Specialized Hyper-parameters

To simplify the above discussions, we implicitly assume that the choice of qualitative spatial relations $\mathcal{R}$, the number of video clips n_v , and the maximum allowed length of object relation chains n_r^* are given. Actually, these three are important and specialized hyper-parameters for our approach and their values will be decided by explorative experiments.

Our approach allows different choices of qualitative spatial calculi. We follow the custom in previous work that manually determines a small set of candidate qualitative spatial calculi. The final choice will be decided by which qualitative spatial calculus or set of qualitative spatial calculus can achieve the best performance in action recognition. Implementing an automatic qualitative spatial calculi deciding algorithm is also important when dealing with large-scale data but will be left as future work.

By dividing each video into an appropriate number of clips, object relation chains in the same video clip are more likely to be semantically relevant; and it is possible to find salient object relation chains in different phases of an action. Essential spatial changes of an action might only appear in certain time intervals of a video. An appropriate number of clips helps us locate those critical time intervals.

The maximum allowed length of object relation chains should also affect the performance of our approach in action recognition and explanation. If the maximum length is too short, actions that involve long chain of spatial relation changes cannot

be well described; if it is too long, our approach will be not friendly to short actions and the number of possible object relation chains will increase accordingly, which might harm the efficiency of our approach.

5.5 Experiments and Evaluation

In this section, we conduct experiments on two datasets to evaluate our approach by comparing it with two other baseline symbolic explainable action recognition approaches.

5.5.1 Baseline Approaches

In this subsection, we introduce two baseline approaches, *intuitive thresholding* and *intuitive maximum* that are also able to generate explanations for actions. They originate from the inductive and abductive reasoning algorithm in [Dubba et al., 2015] and we adapt it for our usage as introduced below.

For each action $a \in \mathcal{A}$, we refer to the set of videos labeled with a as its *positive examples* $P(a)$ and the set of videos not labeled with a as its *negative examples* $N(a)$. Formally, for each $a \in \mathcal{A}$, $P(a) = \{v | \mathcal{L}(v) = a, v \in \mathcal{V}\}$ and $N(a) = \{v | \mathcal{L}(v) \neq a, v \in \mathcal{V}\}$. These two approaches are based on the same *intuitive* assumption: the more frequent an object relation chain exists in the positive examples of an action and the less frequent it is in the negative examples of the same action, the more likely this object relation chain is caused by the action [Dubba et al., 2015]. Based on this assumption, let $P(c, a) = \{v | c \in \mathcal{E}(v), v \in P(a)\}$ be the set of a 's positive examples that contain c and $N(c, a) = \{v | c \in \mathcal{E}(v), v \in N(a)\}$ be the set of a 's negative examples that contain c , the saliency score of an object relation chain c for an action a is denoted as $\mathcal{S}_{it}(c, a)$ and defined in Equation (5.5), where *it* is short for intuitive and η is a hyper-parameter to adjust the relative importance of these two ratios.

$$\mathcal{S}_{it}(c, a) = \frac{|P(c, a)|}{|P(a)|} - \eta \frac{|N(c, a)|}{|N(a)|} \quad (5.5)$$

These two approaches share the same saliency estimator as defined in (5.5) but are different in explanation generation and action recognition. *Intuitive thresholding* is based on the assumption that an object relation chain is caused by an action if the saliency score of an object relation chain for an action is higher than a certain threshold. Formally, let τ be the threshold and $\mathcal{D}(a)$ be the set of causal object relation chains for a , $\mathcal{D}(a)$ is constructed according to the assumption that $c \in \mathcal{D}(a) \Leftarrow \mathcal{S}_{it}(c, a) > \tau$. Given an unseen video v , the set of v 's action labels $\mathcal{L}^\tau(v)$ is decided by the principle that $a \in \mathcal{L}^\tau(v)$ iff there exists c that satisfies both $c \in \mathcal{E}(v)$ and $c \in \mathcal{D}(a)$. Namely, a video v can be labelled with an action a if there is an object relation chain in v that is also in a 's set of causal object relation chains $\mathcal{D}(a)$. An action explanation is a set that consists of such causal object relation chains.

One obvious drawback of the first approach is: it is very likely an unseen video

is labelled with multiple actions, which is a property inherited from the approach in [Dubba et al., 2015]. To overcome this problem, similar to our approach, the second baseline approach *intuitive maximum* focuses on object relation chains with top 1 saliency scores: $\mathcal{L}(v) = a \Leftarrow c \in \mathcal{E}(v), a \in \mathcal{A}$, and $\mathcal{S}_{it}(c, a) = \max(\{\mathcal{S}_{it}(c_*, a_*) | c_* \in \mathcal{E}(v), a_* \in \mathcal{A}\})$. In this case, an action explanation consists of only one object relation chain (i.e. the one with the maximum saliency score and used to determine the recognized action). We report results of both approaches using optimized hyper-parameters for action recognition from grid search.

5.5.2 Datasets

We use two datasets in our experiments: CAD120 [Koppula et al., 2013] and CAD120++ [Zhuo et al., 2019]. CAD120++ is a relabelled version of CAD120. As shown in Table 5.1, CAD120 consists of 124 videos that record 10 actions (e.g. *making cereal* or *taking medicine*) performed by four different people. 12 categories of objects have been labelled including *hand*, *bowl*, *cup* and so on. In contrast, there are 551 videos and 10 different action labels in CAD120++. Actions (e.g. *pick* or *place*) in CAD120++ are much simpler than those in CAD120 in the sense that average number of frames per video is much smaller: 55.7 v.s. 525.3. Only CAD120++ provides ground truth action explanations, which makes it possible to evaluate the explanations generated by those three approaches.

Table 5.1: CAD120 and CAD120++ datasets. **#videos**: number of videos. **#acts**: number of action categories. **#objs**: number of object categories. **#frames**: number of frames per video. **g.t.**: whether ground truth action definitions are provided.

Dataset	#videos	#acts	#objs	#frames	g.t.
CAD120	124	10	12	525.3	No
CAD120++	551	10	12	55.7	Yes

Action definitions provided by [Zhuo et al., 2019] are shown in Table 5.2. In CAD120++, 7 out of 10 actions are defined in the form of changes in relations between objects: there is no definition for *null* and the definitions for *open* and *close* rely on changes in object attributes (e.g. one of the given explanations for *close* is that the attribute of *microwave* has changed from *open* to *closed*).

5.5.3 Hyper-parameter Finetuning

We follow the convention as in [Tayyub et al., 2014; Zhuo et al., 2019] to divide each dataset into 4 folds based on which person performs the action (as mentioned earlier, there are 4 different actors). Namely, we apply 4-fold cross validation and results are averaged across all the 4 folds. The first baseline *intuitive thresholding* predict multiple action labels for each video while the second baseline *intuitive maximum* and our approach are designed for single-label action recognition. Their performance in action recognition is evaluated by the average accuracy of the 4-folds in classification.

Table 5.2: Action definitions in CAD120++. No definition for *null*; *close* and *open* are defined as attribute transitions of a specific object; and the other seven actions are defined as relation transitions between a pair of subject and object. To save space, **m-box** is short for “medicine-box” and **micr** is short for “microwave”.

Action	state transition	subject	object
null	-	-	-
close	open $\rightarrow$ closed	-	m-box, micr, bottle
open	closed $\rightarrow$ open	-	m-box, micr, bottle
pick	not_hold $\rightarrow$ hold	hand	box, m-box, bowl, cup, book, cloth, remote, apple, bottle, plate
			box, m-box, bowl, cup, book, cloth, remote, apple, bottle, plate
place	hold $\rightarrow$ not_hold	hand	cup, book, cloth, remote, apple, bottle, plate
drink	apart $\rightarrow$ contact	head	cup, bottle
eat	apart $\rightarrow$ contact	head	apple, bowl
micr	separate $\rightarrow$ contain	micr	cup, box, bowl
take	contain $\rightarrow$ separate	micro	cup, box, bowl
clean	separate $\rightarrow$ contain	micro	cloth

All hyper-parameters are decided are a grid search of values that achieve the best performance in accuracy. There are two hyper-parameters for the two baseline methods: η , which decides the relative importance of positive examples over negative examples and τ , which is the threshold that decides whether an object relation chain is salient for an action. In the grid search, we vary the value of η between 0 and 2, with an increment of 0.01; and the value of τ between 0 and 1, with an increment of 0.1.

For our approach, the normal neural network training parameters are decided as follows: learning rates is chosen from the set of $[0.1, 0.01, 0.001, 0.0001, 0.00001]$, the candidate set of number of epochs is between 10 to 200 with an increment of 10; and possible values of batch size are in $[16, 32, 64]$. We decide the best combination of values of these three parameters by exploring the combination that achieves the best recognition accuracy.

We determine the values of the three specialized hyper-parameters (i.e. choice of qualitative spatial calculi, number of video clips, and maximum length of object relation chains) by running experiments with one parameter changeable and the other two fixed. First, we decide which set of qualitative spatial relations to extract from videos given both the number of splits and the maximum length of object relation chains as 2. We manually select five qualitative spatial calculi according to human knowledge: RCC-2, RCC-4, RCC-8, QDC-4, and CDR. Note that we assume the ground truth action definitions are unknown when deciding the set of qualitative spatial calculi. We implement the five chosen qualitative spatial calculi based

on QSRLib [Gatsoulis et al., 2016]. As shown in Table 5.3, we choose RCC-4 as the only qualitative spatial calculus to describe actions in CAD120++ because its performance is the best out of other RCC variations and other calculi like QDC-4 and CDR when it is individually used in action recognition; and its performance drops when combined with QDC-4 or CDR. Experimental results in CAD120 are similar: RCC-4 alone achieves the best performance in action recognition. Coarser representations (e.g. RCC-4) sometimes achieves better performance than finer-grained representations (e.g. RCC-8), which might be because irrelevant relations make it harder to learn more accurate explanations.

Table 5.3: Best action recognition accuracies given different qualitative spatial calculi are applied to describe actions in CAD120++ videos. Both the default number of splits and the default maximum length of object relation chains are set as 2.

rcc2	rcc4	rcc8	qdc4	cdr	rcc4&qdc4	rcc4&cdr
63.5	70.2	62.8	37.4	36.8	45.2	44.1

Second, given the default choice of qualitative spatial calculi as RCC-4 and the default maximum length of object relation chains as 2, we have also conducted experiments to decide the best value of the number of clips each video should be divided into. As in Table 5.4, in CAD120++, the best accuracy is achieved by setting the number of clips as 2 while in CAD120, the best choice of number of clips is 5.

Table 5.4: Action recognition accuracies given various number of clips a video is divided into. The default choice of qualitative spatial calculi is RCC-4 and the default maximum length of object relation chains is 2.

Dataset	Number of clips				
	1	2	3	4	5
CAD120++	67.9	70.2	66.4	61.7	62.0
CAD120	83.1	87.1	92.7	93.6	95.2

Third, given RCC-4 as the chosen qualitative spatial calculus and the determined different number of clips for CAD120++ and CAD120, we have confirmed the best maximum length of object relation chains are 2 for both datasets according to the experimental results shown in Table 5.5.

5.5.4 Explanation Evaluation Metric

Ground truth action explanations in the form of object relation chains are generated by replacing natural language relations (e.g. *not holding*, *holding*, *apart*, or *contacting*) in CAD120++ explanations with RCC-4 relations. For example, one of the ground truth explanations of *pick* is (*hand,box,not holding,holding*), which is translated into an

Table 5.5: Action recognition accuracies given different maximum length of object relation chains. The default choice of qualitative spatial calculi is RCC-4 and the default number of clips is 2 for CAD120++, and 5 for CAD120.

Dataset	Maximum chain length		
	2	3	4
CAD120++	70.2	68.9	69.5
CAD120	95.2	83.7	85.2

object relation chain *hand-box-dc-po*. So, for each of the 7 actions, there is a ground truth explanation in the form of a set of object relation chains $\mathcal{X}_{gt}(a)$.

Saliency estimation accuracy (or the quality of generated explanations) is evaluated by *top k rate* ($k \in \{1, 2, 5\}$). Given a video v and its predicted action a , $\mathcal{X}(v, a)$ is a top-k explanation if there is at least one of its top-k object relation chains is in $\mathcal{X}_{gt}(a)$. Top k rate is the rate of recognized videos with top-k explanations out of all recognized videos.

5.5.5 Experimental Results and Analysis

Given the fine-tuned hyper-parameters, a series of experiments has been conducted on CAD120++ and CAD120 and the results are listed in Table 5.6. The two baseline approaches and our approach have been evaluated in Experiment 1, 2, and 3 respectively. One additional experiment has been conducted to further analyze the relationship between saliency estimation accuracy and action recognition accuracy and the results are shown in Figure 5.4.

Action Recognition Accuracy As in Table 5.6, action recognition accuracy is evaluated and our approach achieves much better performance on both datasets when compared to the two baseline approaches. This is mainly because the proposed neural saliency estimator captures the co-occurrence of object relation chains. In this way, our approach is less vulnerable to interfering or incorrect object relation chains thus achieving more robust and accurate recognition. These results also support that our approach is data efficient and works well on small-scale datasets like CAD120.

Quality of Explanation Generation As in Table 5.6, quality of generated explanations (or saliency estimation accuracy) is evaluated by “top1/2/5 rates”. The top1/2/5 rates of our approach are much better than those of the two baseline approaches. Our explanation generator learns to model the relation between object relation chains and actions, which makes it more adaptive in diverse scenarios.

Impact of Saliency Estimation on Recognition As shown in Figure 5.4, during the training process, action recognition accuracy and top1 rates are positively correlated. This indicates that recognition benefits from improved saliency estimation, which further supports the assumption that our neural saliency estimator plays a key role in action recognition.

Qualitative Analysis As shown in Figure 5.5, the ground truth object relation chain

Table 5.6: Results of experiments on CAD120++ and CAD120. **No.:** experiment indexes. **approach:** the approach to action recognition and explanation generation. **intui_thres** is short for intuitive thresholding; **intui_max** is short for intuitive maximum; and **ours** stands for our approach. **accuracy** is the metric for action recognition while **top1/2/5 rates** is to evaluate generated explanations. Results in accuracy and top1/2/5 rates are in percentage (%) and best results are highlighted in bold.

No.	approach	CAD120++		CAD120
		top1/2/5 rates	accuracy	accuracy
1	intui_thres	53.6/57.6/67.2	53.9	67.3
2	intui_max	29.8/33.8/35.4	54.3	71.0
3	ours	61.6/73.8/84.1	70.2	95.2

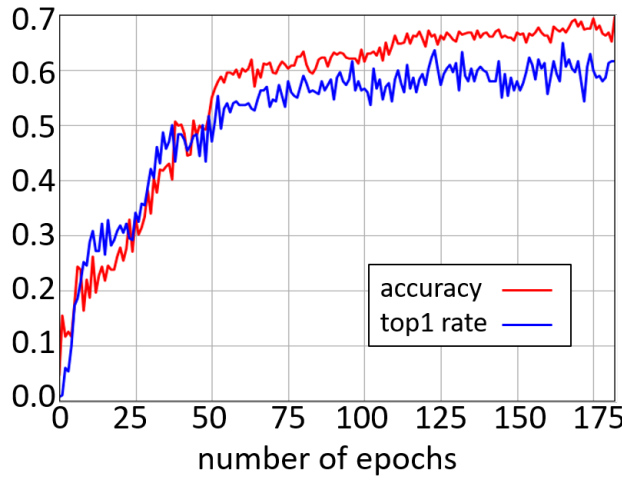


Figure 5.4: Changes in action recognition accuracy and top1 rates as the number of epochs increases. Results are from training our approach on CAD120++ with different number of epochs. The red and blue lines show the changes in action recognition accuracy and top1 rate respectively.

microwave-cloth-po-ppi is identified with the highest saliency score. As a result, the action *clean* in the top video is correctly recognized. As a failure case analysis, in the bottom video, our approach mistakenly estimates interactions between hand and microwave (i.e. the top two object relation chains) as salient object relation chains, which causes an incorrect prediction.

5.5.6 Large-scale Dataset

We have also applied our approach in Charades [Sigurdsson et al., 2016], a large-scale dataset that consists of 9848 videos. The best mean average precision (mAP) is 15.4, which is not comparable to state-of-the-art neural network models. However, our approach is fast and light-weight because it only uses a small number of symbolic

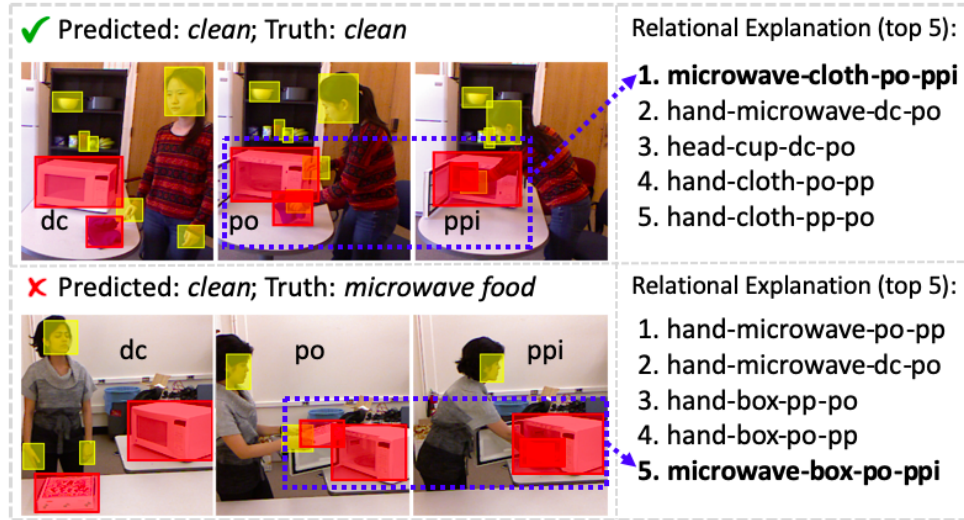


Figure 5.5: A correctly recognized action (top) and a wrongly recognized action (bottom). Right: top 5 salient object relation chains in the explanation. Ground truth object relation chains are in bold. Objects in red are involved in ground truth object relation chains while irrelevant objects are in yellow.

features (i.e. object relation chains) and our goal is to provide extra explainability for SOTA models, not replacing them.

5.6 Conclusion and Future Work

In this chapter, we have proposed an accurate neuro-symbolic approach that achieves explainable action recognition. Our main innovations are (1) object relation chain, a simple but effective symbolic action video representation; (2) a neural saliency estimator that estimates the importance of object relation chains in action recognition; and (3) an explanation generator that generates action explanations in the form of salient object relation chains.

We believe that our approach makes an important contribution towards solving a more general problem: explainable classification by finding critical symbolic features. Our pipeline is to first extract symbolic features and then estimate their importance in classification, which can also be applied to many other similar problems that require trustworthy and explainable agents. Our approach is data efficient and performs well on accuracy and explainability in small-scale datasets. While in large-scale datasets our approach might not be as accurate as state-of-the-art learning approaches in action recognition, our approach can be used to augment these approaches and to provide them with explainabilities. In the future, we aim to make more contributions on this general problem by combining symbolic features with numeric features and utilizing large-scale knowledge databases.

Determining Integrative Relational Spatial Representation

“The largest square has no corners. The greatest vessel takes the longest to finish. Great music has the faintest notes. The great form is without shape.” – Lao Tzu

Most real-world events (e.g. car parking) include essential spatial changes, which can be modeled by changes in qualitative spatial relations between entities. Such qualitative spatial representation is close to human cognition and thus human understandable. It has been applied to help solving many practical problems. However, in almost all previous applications, people manually decide which qualitative spatial model(s) to use without any model selection procedure. In this chapter, we propose and study model selection strategies that can be used to determine the best qualitative spatial representations to recognize a particular event.

6.1 Introduction

Models proposed by researchers in the field of Qualitative Spatial Representation and Reasoning [Renz and Nebel, 2007] are often known as calculus or algebra. We name them as q-models in this chapter because here we are only interested at their representation ability while calculi are more often used in discussing theoretical properties. As introduced in Chapter 1, their practical usefulness consists of at least two parts: representation ability (the ability to represent some of the spatial changes that are essential in describing certain events) and reasoning power (deriving new knowledge based on information at hand). In this chapter, we focus on the first part and aim to explore solutions for a long-standing but rarely discussed problem: model selection-based qualitative spatial representation determination (or q-representation determination). Reasoning over the determined representations is also important but will be left as future work.

As introduced in Chapter 2 and Chapter 5, essential spatial changes in real-world events (e.g. moving a car into a parking space or kicking a ball away) can be represented by a set of entities and changes in their qualitative spatial relations. These

qualitative spatial relations are defined in different q-models, which (as the basis of q-representation) have been applied in many practical problems.

However, it has not been discussed whether there exist better choices of q-model(s) in scenarios discussed in previous chapters. In fact, almost all researchers or practitioners manually decide which q-model(s) to be used for representing a certain event though some of them have done ablation studies to check which combination of q-models performs the best [Sridhar et al., 2011a]. Such manually selected model(s) might be problematic in event representation: they could be incomplete (i.e. not all relevant spatial changes have been covered); ineffective (e.g. spatial changes covered are not deciding); or even be irrelevant in describing an event.

In this chapter, we aim to explore solutions for the problem of model selection-based q-representation determination. We will first introduce principles for building the set of candidate q-models; second describe three model selection strategies and discuss their advantages and disadvantages; and third combine these strategies with a simple scoring technique and provide a preliminary evaluation on their performance in achieving automatic q-representation determination in a simulated scenario.

6.2 Related Work

There is an experimental study in [Sridhar et al., 2011a] where the performance of applying different combinations of five q-models in three different tasks event classification, clustering, and detection has been compared. According to their experimental results, the best performance is always achieved by a combination of all the five q-models. Three categories of representations qualitative temporal, qualitative spatial, and quantitative spatial have been applied in [Tayyub et al., 2014] to recognized daily activities. The best recognition accuracy is also achieved by combining all three representations based on the results of their ablation study.

To the best of our knowledge, there is almost no other previous work that is relevant to model selection in q-representation determination. Regarding to single model-based q-representation determination, [Dubba et al., 2015] have proposed an Inductive Logic Programming-based (or ILP-based) framework and they have discussed how to learn the q-representation of a spatial event (or relational event model in their terminology) given a set of positive and negative examples. Positive examples of an event are video clips where the event exists while negative examples are those irrelevant to the event. Q-representations are preferable if they can cover most positive examples and deny most negative examples. The selected event models are in the forms of RCC-3 relation chains between object pairs and are applied to recognize events in unseen videos.

Next, we will provide a brief overview of several representative qualitative spatial calculi to show that qualitative spatial relations are able to describe spatial relations in different aspects; involving multiple forms of objects; and in various granularities.

6.2.1 Spatial Aspects

There are different taxonomies of spatial aspects in [Renz and Nebel, 2007; Dylla et al., 2017]. Based on both of them, the taxonomy that will be used in our following discussions is $\{dir, dis, mot, siz, top\}$ where *dir* is for direction (e.g. north or left), *dis* is for distance (e.g. near or far), *mot* is for motion (e.g. leaving or still), *siz* is for size (e.g. smaller or larger), and *top* is for topology (e.g. connected or disjoint). Time is regarded as a special and implicit aspect because it has to be considered in almost all spatial events (as long as there are changes in spatial relations).

Most q-models only cover one aspect: the *STAR* Calculus [Renz and Mitra, 2004] focuses on qualitative direction relations between points with arbitrary granularity; the Qualitative Distance Calculus (QDC) [Clementini et al., 1997] is also designed with arbitrary granularity and is used to describe qualitative distance relations; the Qualitative Trajectory Calculus [Van de Weghe et al., 2006] models the moving direction and speed relations between objects as in Figure 6.1a; Point Algebra (PA) [Van Beek, 1992] (which consists of three relations $>$, $<$ and $=$) can be used to describe the relative size of objects; RCC-m consists of only topological relations.

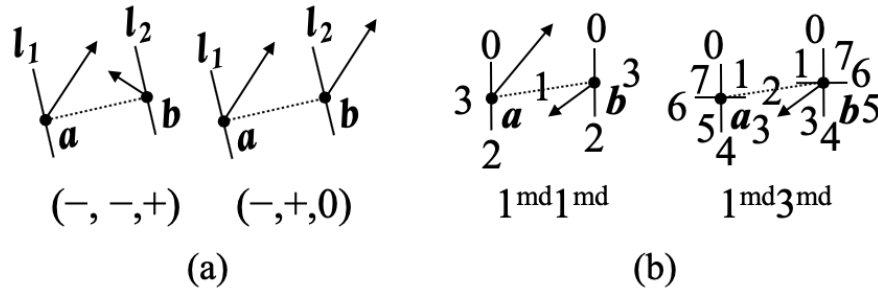


Figure 6.1: (a) Two examples of QTC_B relations: (left) the first $-$ means a is moving towards b (or it is moving to the region that is right to Line l_1); the second $-$ indicates b is also moving towards a (or it is moving to the region that is left to Line l_2); the last $+$ means a 's speed is greater than that of b ; (right) the first $-$ means a is moving towards b ; the second $+$ shows b is moving away from a (or it is moving to the region that is right to Line l_2); the last 0 means they have the same speed. (b) $QAMD_1$ (left): QTC_B is integrated with $STAR_1$. The moving direction of an object can be described with 0, 1, 2 or 3; in contrast, $QAMD_2$ (right) is finer-grained and describes moving directions with 8 discrete values..

There are also q-models that cover multiple aspects. For example, the Ternary Point Configuration Calculus (TPCC) [Moratz and Ragni, 2008] is a combination of direction and distance relations and CORE-9 [Cohn et al., 2012] is relevant to multiple spatial aspects: direction, topology, etc.

6.2.2 Object Forms and Relation arities

Q-models that are relevant to the same aspect can vary because of different abstract forms of objects and relation arities. For example, the *STAR* Calculus consists of

binary direction relations between points; Cardinal Direction Relations (CDR) [Skadopoulos and Koubarakis, 2004] focuses on binary direction relations between regions; while the Double Cross Calculus (DCC) [Freksa, 1992c] deals with ternary direction relations between points.

6.2.3 Granularities

Some q-models have different variations because they are designed to be flexible with different granularities. For example, the granularities of $\mathcal{STAR}_m$ and $\mathcal{QDC}_m$ can be different positive integers. A special case is the RCC family: there are at least five variations of RCC- m with different granularities. Namely, m can be 2, 3, 4, 5, or 8. Coarser variations are generated from finer variations by mapping multiple finer relations into a single coarser relation or dropping some of the finer relations. For example, two RCC-3 relations po and pp are mapped into a RCC-2 relation c (which indicates two regions are connected) [Gatsoulis et al., 2016].

The granularity of a q-model can be decided by users or learnt by clustering metric relations, which is similar to the case in [Tayyub et al., 2014] where numeric time durations are clustered into short, equal and long. For example, in [Galata et al., 2002], numeric direction, velocity, and distance relations between cars are clustered into qualitative spatial relations, which is the basis of describing high-level interactive behaviors between cars. Similar technique is also utilized in [Bleser et al., 2015].

6.3 Candidate Q-models

In this section, we will introduce our principles for building a set of candidate q-models and based on which in the next section several model selection strategies will be proposed. We assume there is no domain or context knowledge to achieve an automatic and general solution for the problem of model selection-based q-representation determination. To deal with the agnostic problems that might come with our basic assumption, the set of candidate q-models should be inclusive. Our default set of candidate q-models is built according to the following principles while there might be alternatives based on different principles:

(1) At least one q-model for each aspect should be selected to ensure all aspects are covered. It is not easy to determine which aspect or which combination of aspects is relevant for a certain event without expert knowledge. So a safe choice is to cover all the five aspects mentioned in Section 6.2.

Q-models that are relevant to multiple aspects are not considered. This is for three reasons: first, in theory they can be replaced by multiple single-aspect models; second, these multi-aspect models are less intuitive and thus less human understandable (e.g. it is easy to describe single-aspect spatial changes to people in natural language while less straightforward to explain multi-aspect spatial changes); and third, multi-aspect models are more cumbersome and insensitive when used to describe actions that are only relevant to partial aspects of those models.

(2) If applicable, q-models with different forms of objects and relation arities should be covered. For example, $\mathcal{STAR}$ and CDR are included in the default set to cover direction relations between both points and regions. We focus on binary relations in this chapter while discussions on relations with other arities are left as future work.

(3) For q-models with various granularities, different variations should be considered. The selection of model granularities can be critical in certain cases. For example, it is possible that one critical spatial change can be described by qualitative spatial relations with finer granularity but is not observable when coarser variation of the q-model is applied. Finer q-models can also be disadvantageous because they might record too many irrelevant qualitative spatial relation changes, which is misleading in q-representation determination. Another problem to consider in realistic scenarios is coarser distinctions are more reliable to be recognized than finer ones due to noise in input data [Gatsoulis et al., 2016]. There are also relations rarely occur in real-world data (e.g. it is very unlikely that two regions are exactly the same) [Dubba et al., 2015]. Based on this fact, the set of candidate granularities for RCC-m is $\{2, 3, 4\}$ so that eq is excluded. For q-models with arbitrary granularity, we may have to empirically define a candidate set of granularities (since it is of course not feasible to choose from an infinite number of variations). For example, candidate granularities for $\mathcal{STAR}_m$ is $\{1, 2, 4\}$. An alternative choice could be to determine the granularity but unsupervised clustering algorithms as in [Galata et al., 2002; Tayyub et al., 2014].

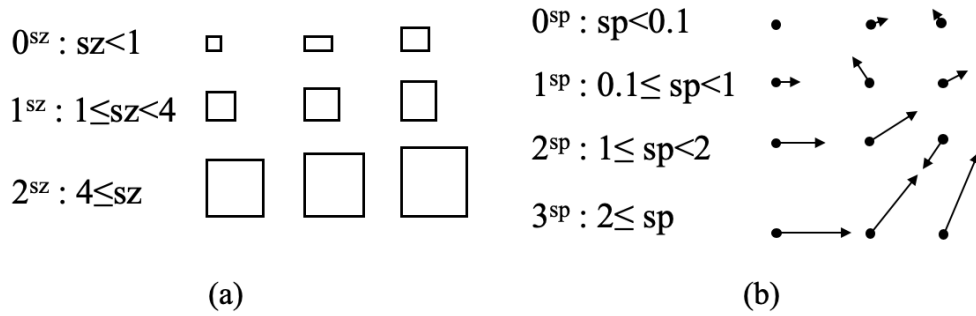


Figure 6.2: (a) One possible definition of QASz₃: rectangles are divided into three distinct groups according to their sizes (or areas) and their qualitative sizes will be denoted as 0^{sz} , 1^{sz} , and 2^{sz} respectively. QASz₃ relations are in the forms of $i^{sz}j^{sz}$ where $i, j \in \{0, 1, 2\}$. (b) QASp₄: qualitative speeds of moving objects (arrow orientation indicates the moving direction and arrow length indicates magnitude of moving speed) are described with 4 discrete values 0^{sp} , 1^{sp} , 2^{sp} and 3^{sp} regardless of their moving directions. QASp₄ relations are in the forms of $i^{sp}j^{sp}$ where $i, j \in \{0, 1, 2, 3\}$.

(4) If applicable, both absolute and relative relations should be covered. Absolute relations are defined according to certain criteria while relative relations are defined by relative comparisons. According to this principle, besides qualitative relative size

relations (i.e. whether the size of one object is larger, smaller, or equal to that of the other object), we believe it is also necessary to consider qualitative absolute sizes according to the fact that people might describe an object as *small*, *large* or *huge* without comparing to the size of any specific object. Therefore, a new q-model named **QASz_m** is proposed, which consists of m^2 size relations and m is the granularity. To align with other qualitative spatial relations discussed in this chapter, QASz_m relations are also binary and in the forms of $i^{sz}j^{sz}(o_1, o_2)$ where $i, j \in \{0, 1, \dots, m-1\}$. It indicates the qualitative size of o_1 is i^{sz} and that of o_2 is j^{sz} (an example of the criteria for different qualitative sizes is in Figure 6.2a). Such distinctions of qualitative sizes can also be manually decided by users or learnt by clustering.

We also propose **QASp_m** to describe qualitative absolute speeds of moving objects since QTC_B only records qualitative relative speeds (i.e. whether the speed of one object is greater, small or equal to that of another object). Similarly, QASp_m relations are in the forms of $i^{sp}j^{sp}(o_1, o_2)$ where $i, j \in \{0, 1, \dots, m-1\}$, which means the qualitative speed of o_1 is i^{sp} and that of o_2 is j^{sp} (an example of the criteria for different qualitative speeds is in Figure 6.2b). The moving-or-stationary model proposed in [Gatsoulis et al., 2016] (which only distinguishes whether an object is stationary or moving) can be regarded as a specific case of QASp₂.

The third q-model to be proposed is **QAMD_m**, which consists of qualitative absolute moving direction relations with arbitrary granularity since QTC_B only describes relative moving directions in a relative manner with fixed granularity. QAMD_m is an integration of QTC_B and STAR_m as in Figure 6.1b and its relations are also in the forms of $i^{md}j^{md}(o_1, o_2)$ where $i, j \in \{s, 0, 1, \dots, m-1\}$. Moving directions are described as STAR_m relations $\{0, 1, \dots, m-1\}$ and s is short for still which means the object is not moving.

Our default set of candidate q-models is shown in a hierarchical manner in Figure 6.3 so that the inter-connections of q-models can be clearly displayed. This hierarchical tree consists of four levels: Level 0 (or the top level) is the root of the hierarchical tree (denoted as Q-models); Level 1 is the set of spatial aspects that an event might be relevant to; Level 2 is the set of single-aspect q-models; and Level 3 (or the bottom level) is the set of candidate granularities for some of the q-models. Each leaf node represents a certain candidate q-model and there are 21 q-models in total.

6.4 Model Selection Strategies

6.4.1 Naive

Given the default set of candidate q-models C_d (as shown in Figure 6.3 and $|C_d| = 21$), one naive solution for model selection in q-representation determination is to examine all of the q-model combinations. Assume there is a scoring function $S : (E, C) \rightarrow \mathbb{R}$ where E is a set of positive and negative examples for a certain event e and $C \subseteq C_d$. $S(E, C)$ indicates the best score of q-representations for e can achieve by using relations from q-models in C (more details in Section 5). This naive strategy can always find the optimal combination of q-models (or the combination with the

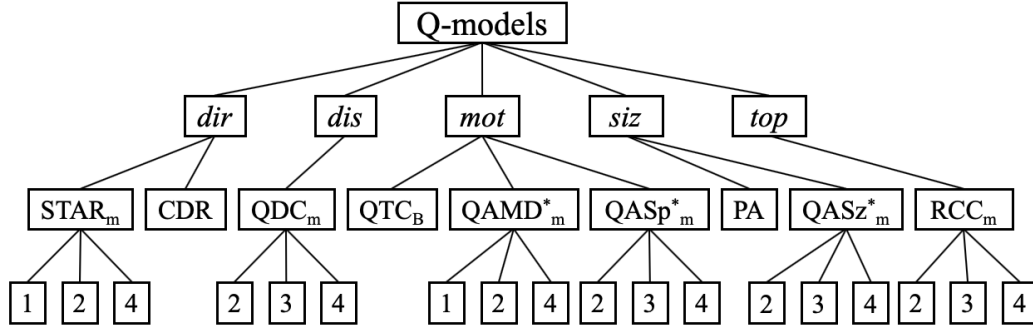


Figure 6.3: The default hierarchical tree of candidate q-models. Models marked with * are first formally proposed in this chapter.

maximum score). However, it is inefficient because the computational complexity is exponential (i.e. $(2^{21} - 1) * t_s \approx 2 * 10^6 * t_s$ where t_s is the average time spent on each scoring).

The number of iterations can be largely reduced by considering the fact that it is unnecessary to combine different variations of the same q-models (e.g. $STAR_1$ and $STAR_2$) and setting an upper limit to the number of q-models selected (since it is not reasonable to have a q-representation consisting of relations from 21 q-models). But still the number is larger than 1000 if the maximum number of q-models in a combination is 3. So it is only suitable in cases where the scoring process is extremely fast.

Algorithm 5 Random Climbing

```

1: procedure RANDOMCLIMB( $C_d, n_{it}$ )
2:    $C_{max} = \text{RandomSubset}(C_d)$ 
3:   for  $i$  in  $n_{it}$  do
4:      $C_+ = \text{RandomAdd}(C_{max})$ 
5:      $C_- = \text{RandomRemove}(C_{max})$ 
6:      $C_r = \text{RandomReplace}(C_{max})$ 
7:      $C_{max} = \text{MaxScore}(C_{max}, C_+, C_-, C_r)$ 
8:   return  $C_{max}$ 

```

6.4.2 Random Climbing

The second strategy for model selection (Algorithm 5) is inspired by the greedy hill-climbing algorithm described in [Tang et al., 2013]. The basic idea is: first randomly select a subset of C_d (Line 2); then check whether the score can be improved by randomly adding a new q-model, removing a q-model, or replacing a q-model with another q-model (Line 4-7); after a certain number of iterations (Line 3), return the current best q-model combination (Line 8). Another stopping criterion could be that the maximum score remains unchanged from a certain number of iterations [Tang et al., 2013]. Another necessary threshold is the maximum number of q-models

allowed in C_{max} to avoid over-complicated q-representations (similar to what have been discussed at the end of the previous subsection). By using this strategy, we can control the number of iterations. However, it cannot guarantee to return the optimal set of q-models (even if the number of iterations is greater than that of Naive).

6.4.3 Fast Selection

In this subsection we will propose a fast model selection strategy, which is based on several assumptions (while in contrast the first two strategies are assumption-free). These assumptions are helpful in accelerating the model selection process but it is very difficult to theoretically prove their correctness. Instead, the performance of this strategy will be evaluated by experiments in Section 6.

The following assumptions are the basis of our fast selection strategy described in Algorithm 6: (1) An event can be relevant to one or multiple spatial aspects and whether one aspect is relevant can be tested by checking some of the q-models in this aspect. The representative q-model for an aspect can be either the coarsest (i.e. the one that consists of the fewest number of relations), the finest (i.e. the one that consists of the greatest number of relations) or a random one. For example, we can decide whether *dir* is relevant by checking the score of one of the *dir* q-models or their combinations with q-models from other aspects.

(2) For *mot* and *siz*, absolute and relative relations might be complementary. For example, we should examine not only PA and QAS_{*m*} separately but also the combination of them.

(3) Q-models in the same aspect but covering different forms of objects are exclusive. For example, we have to choose one from STAR_{*m*} and CDR.

(4) Q-models from different aspects are independent in deciding their best combination. For example, given a set of q-models $C_1 = \{c_{11}, c_{12}, c_{13}\}$ in one aspect and a second set of q-models $C_2 = \{c_{21}, c_{22}, c_{23}\}$ in another aspect, the optimal combination of models from these two aspects can be found in $3 * 3$ iterations by using the naive strategy. However, based on this assumption, we can find the best combination in $3 + 3$ iterations: first find the best q-model in C_1 by checking the scores of $\{c_{11}, c_{21}\}$, $\{c_{12}, c_{21}\}$, and $\{c_{13}, c_{21}\}$; then find the best combination by checking the scores of $\{c_{12}, c_{21}\}$, $\{c_{12}, c_{22}\}$, and $\{c_{12}, c_{23}\}$ if the best q-model in C_1 is c_{12} .

The fast selection strategy (Algorithm 6) consists of two main steps. Based on assumption (1), first we determine which spatial aspects are relevant by checking all the combinations of the five initial q-models (one from each aspect). As mentioned above, these q-models can be either the coarsest (i.e. $\{STAR_1, QDC_2, QASp_2, QASz_2, RCC_2\}$), the finest (i.e. $\{CDR, QDC_4, QAMD_4, QASz_4, RCC_4\}$), or randomly selected from each aspect (Line 2). $\wp(C)$ returns the power set of C (Line 3). The best combination C_{max} is the one in $C_{allcomb}$ with maximum score. Aspects with q-models in the best combination C_{max} are regarded as relevant, which are denoted as $Aspects(C_{max})$ (Line 5).

Second we go through all relevant aspects (Line 5-13) to find the best combination of models. Based on Assumption (2) and (3), for *dir*, *dis* or *top* only one q-model can

Algorithm 6 Fast Selection

```

1: procedure FASTSELECT( $C$ )
2:    $C_{init} = \text{Coarsest}(C)/\text{Finest}(C)/\text{Random}(C)$ 
3:    $C_{allcomb} = \wp(C_{init}) \setminus \emptyset$ 
4:    $C_{max} = \text{MaxScore}(C_{allcomb})$ 
5:   for  $aspt \in \text{Aspects}(C_{max})$  do
6:      $c = \text{QModel}(aspt, C_{max})$ 
7:      $C_{aspt} = \text{AllModelsOrCombs}(aspt, C)$ 
8:     for  $C_{tmp}$  in  $C_{aspt}$  do
9:        $C_{tmp} = C_{tmp} \cup (C_{max} \setminus c)$ 
10:      if  $S(C_{tmp}) > S(C_{max})$  then
11:         $C_{max} = C_{tmp}$ 
12:   return  $C_{max}$ 

```

be included in a combination while for *mot* or *siz* certain model combinations are allowed. Based on Assumption (4), when checking the models or model combinations in each aspect (Line 7-8), the choice of models or model combinations in other aspects remain unchanged (Line 9). $\text{QModel}(aspt, C_{max})$ returns the q-model from *aspt* in C_{max} (Line 6). The so far best combination is updated if the score of the newly checked combination is higher (Line 10-11).

In the first step, we will have to go through $2^5 - 1 = 31$ iterations to find relevant aspect(s) (if the maximum number of q-models that can be selected is 5). The minimum number of iterations required in the second steps is 2 if only *dis* or *top* is relevant while the maximum number is 43 if all the five aspects are relevant (numbers of iterations required for *dir*, *dis*, *mot*, *siz* and *top* are 3, 2, 30, 6 and 2 respectively). So in the worse case, the fast selection strategy just has to go through 74 iterations, which should be much faster than two previous strategies. However, its effectiveness depends on the correctness of the underlying assumptions and might be helpful only in certain cases.

Algorithm 7 Q-representation Determination

```

1: procedure BESTQRWITHSCORE( $E, C$ )
2:    $E = E_+ \cup E_-$ 
3:   for  $\tau \in \mathcal{T}_{E_+}(C)$  do
4:     for  $e \in E_+$  do
5:       if  $\tau \in \mathcal{T}_e$  then
6:          $N_+^\tau = N_+^\tau + 1$ 
7:     for  $e \in E_-$  do
8:       if  $\tau \in \mathcal{T}_e$  then
9:          $N_-^\tau = N_-^\tau + 1$ 
10:     $S(\tau) = \alpha \frac{N_+^\tau}{|E_+|} - \beta \frac{N_-^\tau}{|E_-|} - \gamma |\tau|$ 
11:     $\tau_{best} = \text{TWithMaxScore}(\mathcal{T}_{E_+})$ 
12:   return  $\tau_{best}, S(\tau_{best})$ 

```

6.5 Q-representation Determination

To fast evaluate the above mentioned strategies, in this section we will provide a fast and straightforward scoring algorithm to determine the best q-representation for a certain event (and also assign a score to it). The ILP-based framework in [Dubba et al., 2015] is not applied here because it has to traverse a large search space and thus is not suitable in our case where a large number of iterations is required (which is inevitable when model selection is considered).

A q-representation of an event e is actually a chain of spatial relations $\tau_e = \langle r_1 r_2 \dots r_{|\tau|} \rangle (o_i, o_j)$ [Tayyub et al., 2014] to indicate the essential changes in spatial relations between two objects in this event given a certain set of q-models $C = \{c_1, c_2, \dots, c_{|C|}\}$. $r \in c_1 \times c_2 \times \dots \times c_{|C|}$ (i.e. the Cartesian product of all $c \in C$) and $o_i, o_j \in O_e$ (i.e. the set of all objects that can be recognized in e). Our simplified assumption here is that in an event only two objects are relevant (in practice there might be more, so the q-representation should be a set of chains of spatial relations between different objects). For example, the car parking example (Figure 5.1a) can be well described by a q-representation $\langle (dc, 1^{sp}0^{sp})-(po, 1^{sp}0^{sp})-(in, 0^{sp}0^{sp}) \rangle (o_c, o_{ps})$ where $C = \{RCC-3, QASp_2\}$.

Given a set of q-models C and a set of positive (+) and negative (−) examples for a certain event E , our q-representation determination algorithm is described in Algorithm 7. $\mathcal{T}_{E_+}(C)$ (Line 3) is the set of all unique chains of spatial relations between two objects that can be extracted in the set of positive examples E_+ given a set of q-models C . The score of each τ in $\mathcal{T}_{E_+}(C)$ is calculated at Line 10, where N_+^τ is the number of positive examples τ covers, $|E_+|$ is the number of positive examples, N_-^τ is the number of negative examples it covers, $|E_-|$ is the number of negative examples, $|\tau|$ is the length of τ (or the number of “-” plus one) and α, β, γ are three configurable parameters to balance of the importance of these factors [Dubba et al., 2015]. $|\tau|$ is considered here because long chains of spatial changes are not preferable. The computational complexity of this scoring function is linear to the number of positive and negative events. We may also get rid of chains that are too short (e.g. $|\tau| = 2$) to reduce accidental spatial relation changes.

6.6 Demonstrative Evaluation

In this section, we will conduct a demonstrative evaluation on our proposed strategies by combining them with the simple scoring algorithm described in the previous section. Our evaluation is demonstrative and tentative because it is based on a small-scale set of simulated events and several simplifying assumptions. It is left as future work that evaluating these strategies on large-scale real-world datasets with more realistic assumptions.

6.6.1 Features of Determined Q-Representations

The first set of experiments aims to analyze the determined q-representations. There are 4 events and 5 positive examples for each of them. Each example consists of a series of spatial changes on three objects. The positive examples of each event conform to its ground-truth q-representation (manually generated). Namely, objects in these examples satisfy the constraints implied by the spatial relation changes in the q-representation.

When determining the q-representation for a certain event, its ground-truth q-representation is of course unknown and the positive examples of other events are used as negative examples. It is also unknown which two of the three objects in an example are relevant. To reduce the disturbance of accidental interactions between objects, we just neglect chains of spatial relations with length less than 3 (i.e. $|\tau| < 3$). Since all examples are clean and tidy (which is almost impossible in real-world data), it is very easy to find a q-representation that can cover all positive examples and reject all negative examples (while only the simplest q-representation is listed). The ground-truth and determined representations are listed as follows (objects are omitted). *gt* means “ground truth” and *de* indicates “determined”. e_i is the i th event.

$$\begin{aligned} \tau_{e_1}^{gt} &= \langle in-po-dc \rangle \text{ with } C = \{RCC-3\}; \\ \tau_{e_1}^{de} &= \langle in-po-dc \rangle \text{ with } C = \{RCC-3\}; \\ \tau_{e_2}^{gt} &= \langle (dc, 0^{sz}1^{sz})-(c, 1^{sz}0^{sz})-(dc, 0^{sz}0^{sz}) \rangle \text{ with } C = \{RCC-2, QASz_2\}; \\ \tau_{e_2}^{de} &= \langle 0^{sz}1^{sz}-1^{sz}0^{sz}-0^{sz}0^{sz} \rangle \text{ with } C = \{QASz_2\}; \\ \tau_{e_3}^{gt} &= \langle (5, 2^{sp}0^{sp})-(3, 1^{sp}0^{sp})-(1, 0^{sp}0^{sp}) \rangle \text{ with } C = \{STAR_2, QASp_3\}; \\ \tau_{e_3}^{de} &= \langle 5-3-1 \rangle \text{ with } C = \{STAR_2\}; \\ \tau_{e_4}^{gt} &= \langle (0, 1^{md}1^{md}, >)-(1, 3^{md}3^{md}, <)-(0, 3^{md}3^{md}, <) \rangle \text{ with } C = \{QDC_2, QAMD_1, PA\}; \\ \tau_{e_4}^{de} &= \langle (0, >)-(1, <)-(0, <) \rangle \text{ with } C = \{QDC_3, PA\}. \end{aligned}$$

It can be observed from the above q-representations that $\tau_{e_1}^{de}$ is the same as $\tau_{e_1}^{gt}$; $\tau_{e_2}^{de}$ and $\tau_{e_3}^{de}$ can be regarded as “subsets” of their corresponding ground truths; $\tau_{e_4}^{de}$ is also more simple than $\tau_{e_4}^{gt}$ by not considering $QAMD_1$ and replacing QDC_2 with its finer-grained version QDC_3 . We can conclude that (at least in these cases) determined q-representations (1) can correctly describe their corresponding events; (2) prefer to be succinct; and (3) might be able to adjust its representation ability by changing the granularity of certain q-models.

6.6.2 Effectiveness and Efficiency

The second set of experiments is to evaluate the effectiveness and efficiency of our proposed model selection strategies. Positive and negative examples for a given event are randomly generated to mimic noisy and chaotic real-world data. The experimental results are listed in Table 1 and each score is the average of 10 repetitions. We can still determine q-representations from such noisy examples with decent scores. The maximum number of q-models in a combination is set to 3 (so the total number of iterations for Naive is about 1200). The number of maximum iterations for Random Climbing is set to 200. Of course the best score can always be achieved by

Naive given different number of positive and negative examples. It is noticeable that Fast Selection is better than Random Climbing in terms of both effectiveness and efficiency because it can always achieve the second best scores with the coarsest or random initial q-models though the gap between their scores becomes smaller as the number of examples increases; and its average number of iterations is much less than that of Random Climbing (i.e. 200).

Table 6.1: Scores achieved by different model selection strategies (Naive, Rdm, FastC, FastF, and FastR) given different number of positive (+) and negative (−) examples (as in Column 1). Rdm is short for Random Climbing. FastC means Fast Selection with the coarsest q-models from each aspect as the initial q-models; FastF is Fast Selection with the finest initial q-models; and FastR means Fast Selection with random initial q-models. The average numbers of iterations are in brackets for the three Fast Selection strategies.

+/-	Naive	Rdm	FastC	FastF	FastR
5,5	0.64	0.5	0.6 (25.3)	0.5(18.6)	0.48 (19)
5,10	0.58	0.44	0.52 (25.9)	0.48 (18.8)	0.46 (26.7)
10,10	0.44	0.37	0.37 (28.6)	0.34 (18)	0.38 (22.4)
10,15	0.42	0.35	0.4 (18.8)	0.35 (18.9)	0.37 (19.3)
20,20	0.34	0.32	0.31 (22.8)	0.29 (18)	0.33 (19.7)

6.7 Conclusion and Future Work

To conclude, in this chapter, we have described a set of candidate q-models; and then discussed and evaluated three different model selection strategies for q-representation determination. As aforementioned, the next step is to evaluate these selection strategies on large-scale real-world datasets with more realistic assumptions.

In the future, we may work on extending our proposed candidate set of q-models by considering q-models that are defined in a probabilistic manner; checking the overlaps between the representation ability of different q-models; or as mentioned in the introduction trying to solve the reasoning problems over the determined representations. In a wider range, we might also try to consider non-spatial changes in event representation determination; or to derive reasonable natural language descriptions from the determined q-representations.

Conclusion

“Keep your eyes on the stars, but don’t forget that your feet are necessarily on the earth.” – Theodore Roosevelt¹

In Section 1.3, we proposed four research problems: qualitative spatial relation integration, integrative qualitative localization and navigation, integrating QSRR with deep learning, and integrative qualitative spatial representation selection. In this thesis, we have discussed these problems and their solutions in detail. In this chapter, we conclude this thesis by summarizing our key contributions and impact (Figure 7.1); and then discuss future work.

7.1 Contributions and Impact

Figure 7.1: Summary of key contributions and impact of Chapter 3-6.

	Key Contributions	Impact
Chapter 3	Qualitative Direction Relation Integration	Qualitative Spatial Relation Integration
Chapter 4	Integrative Qualitative Localization and Navigation	Human-like Localization and Navigation
Chapter 5	Integrating QSRR with Deep Learning	Trustworthy AI Agents
Chapter 6	Integrative Spatial Representation Selection	Automatic Spatial Representation Determination

First, we have presented the first formalism that is able to integrate direction relations across different frames of reference. In the past, direction relations described in different frames of reference were detached but now they are inter-connected based on our proposed formalism. Relations in one frame of reference can also be translated into those in another. We have integrated direction relations by explicitly modeling

¹In a speech in Chicago, Illinois, 1900.

the frames of reference used in direction relation descriptions and analyzing across-frames-of-reference transformations. This methodology is also applicable for other categories of spatial or spatio-temporal relations. For example, in most cases, qualitative size relations and qualitative distance relations are also described in implicit frames of reference.

Second, we have provided a novel system that integrates spatial and non-spatial relations in qualitative localization and navigation. This system is applicable in GPS-denied environments because it is based on place knowledge and interactions with people. Our system is not designed to replace main-stream quantitative localization and navigation systems. Instead, it is to be combined with those systems to implement more human-like localization and navigation. GPS localization-based navigation systems can also be used in GPS-denied environments with the assistance of our qualitative localization mechanism. The success of our integrative system encourages future human-like localization and navigation systems to integrate more symbolic and thus human understandable knowledge in the future.

Third, we have proposed an accurate neuro-symbolic approach that integrates qualitative spatial relations with deep neural networks to achieve explainable action recognition. Symbolic structured features are more explainable and human-understandable than numeric unstructured features. We have extracted symbolic features from videos and estimated their importance in classification. This strategy also applies to building trustworthy agents in other scenarios. We believe our first trial on such integration enlightens future attempts on integrating large-scale knowledge databases with learning-based approaches and on building more trustworthy AI agents.

Last, to the best of our knowledge, we are the first to discuss the problem of automatic qualitative spatial representation determination. We are also the first to propose possible solutions for this problem. We have proposed three determination strategies and discuss their advantages and disadvantages. We have implemented these strategies and conducted tentative experiments to show their feasibility and performance on a small set of events. Our exploration on this problem provide a basis of and will boost future research on more sophisticated strategies for automatic qualitative spatial presentation determination.

7.2 Future Work

Our ultimate goal is to fulfill practically useful spatial intelligence by applying an integrative methodology on QSRR and techniques from other related domains. We expect this thesis inspires and motivates future research on this ultimate goal. Next, we plan to work on unsolved problems in the aforementioned four higher-level research domains and other domains where QSRR can be applied.

7.2.1 Integrative Qualitative Distance Relations

Similar to direction relations, in realistic scenarios, distance relations are also described in the same or different frames of reference. For example, [Richter et al., 2013] point out six levels of granularities for place location descriptions: furniture, room, building, street, district, city, and country, which are also six levels of frames of reference for distance relation description. There are already discussions on the problem of integrating qualitative distance relations across different frames of reference in [Hernandez et al., 1995], which claim that there are three categories of frames of reference for distance deciding: intrinsic, extrinsic, and deictic. Intrinsic frames of reference are the inherent characteristics of reference objects (e.g. the size of a house); extrinsic frames of reference are external factors like means of travel (e.g. by foot or by car); and deictic frames of reference are external viewpoints.

However, unlike direction relations [Hua et al., 2018], there is still no formal solution on the problem of representing and reasoning about distance relations defined in different frames of reference. We believe the prerequisite of across-frame-of-reference distance relation reasoning is also the connections between the underlying frames of reference. Different from direction relations, in across-frame-of-reference distance relation reasoning, the absorption rule [Hernandez et al., 1995] is not neglectable: in composition, the distance relation defined in a small-scale frame of reference will be absorbed by that defined in a much larger-scale frame of reference.

7.2.2 Human-like Localization and Navigation

[Hua et al., 2019] is our first step onto human-like localization and navigation. We experiment our system on simulated data while its performance on large-scale datasets or in real-world scenarios remains unknown to us. We believe it is necessary to explore the following two problems to achieve authentic human-like localization and navigation: first, how QSRR can benefit high-quality spatial knowledge extraction from different sources and data fusion in different modalities (e.g. natural language descriptions v.s. OpenStreetMap data); and second, how to apply QSRR in producing more human-like route instructions. In Section 2.3.2, we have introduced previous research on extracting qualitative spatial representation from videos and automatic action description generation. It is worth checking whether techniques developed for video actions can be transferred to the domain of extracting and producing spatial knowledge in localization and navigation.

7.2.3 Trustworthy AI Agents

Building robust or trustworthy AI agents [Marcus, 2020] is an extremely broad topic and our focus is the trustworthiness brought by finding critical or determining symbolic knowledge in decision making. For example, finding critical symbolic features in action recognition. In Chapter 5, we applied a simple 1d attention layer to find critical object relation chains while in recent years there are numerous developments

in more sophisticated and effective attention mechanisms. For example, the “transformer” model proposed in [Vaswani et al., 2017]. It remains an open problem how symbolic knowledge can be utilized together with such effective attention mechanisms and whether the transformer model is more effective in finding critical symbolic knowledge.

There is also research on learning explanatory rules by inductive logic programming [Evans and Grefenstette, 2018]. We may also check whether it is feasible to apply their approach to learn explanations in the forms of qualitative spatial relations.

7.2.4 Spatial Knowledge Graph

Knowledge graph [Ehrlinger and Wöß, 2016], as a typical knowledge representation paradigm, draws more and more attention in recent years and has been widely applied in various domains. For example, search engines and question-answering services². Knowledge graph and qualitative spatial representation share common basis: object pairs and their relations. It would be interesting to explore whether qualitative spatial reasoning techniques can be applied in knowledge graph and whether the successful applications of knowledge graph can enlighten future QSRR applications.

7.2.5 Automatic Qualitative Spatial Representation Determination

The fast selection strategy proposed in Chapter 6 is based on several simplifying assumptions and evaluated in small-scale simulated data. In the future, we plan to search for more theoretical supports on these assumptions and find an appropriate large-scale dataset for a more convincing evaluation. We may also attempt to develop a more adaptive learning-based qualitative representation determination strategy.

7.3 Summary

In this thesis, we focus on integrative qualitative spatial representation and reasoning. We present our exploratory solutions on four specific research questions under this topic. We hope this thesis enlightens future research on improving the practical usefulness of qualitative spatial representation and reasoning.

²https://en.wikipedia.org/wiki/Knowledge_graph

Bibliography

- ALLEN, J. F., 1990. Maintaining knowledge about temporal intervals. In *Readings in qualitative reasoning about physical systems*, 361–372. Elsevier. (cited on pages 2, 16, and 20)
- BLEI, D. M.; NG, A. Y.; AND JORDAN, M. I., 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3 (2003), 993–1022. (cited on page 20)
- BLESER, G.; DAMEN, D.; BEHERA, A.; HENDEBY, G.; MURA, K.; MIEZAL, M.; GEE, A.; PETERSEN, N.; MAÇÃES, G.; DOMINGUES, H.; ET AL., 2015. Cognitive learning, monitoring and assistance of industrial workflows using egocentric sensor networks. *PloS one*, 10, 6 (2015), e0127769. (cited on pages 20, 70, and 86)
- BONIARDI, F.; BEHZADIAN, B.; BURGARD, W.; AND TIPALDI, G. D., 2015. Robot navigation in hand-drawn sketched maps. In *2015 European conference on mobile robots (ECMR)*, 1–6. IEEE. (cited on page 47)
- CHEN, H.; VASARDANI, M.; WINTER, S.; AND TOMKO, M., 2018. A graph database model for knowledge extracted from place descriptions. *ISPRS International Journal of Geo-Information*, 7, 6 (2018), 221. (cited on pages 19 and 47)
- CHEN, J.; COHN, A. G.; LIU, D.; WANG, S.; OUYANG, J.; AND YU, Q., 2015. A survey of qualitative spatial representations. *The Knowledge Engineering Review*, 30, 1 (2015), 106–136. (cited on pages xvii, 2, 9, 10, and 66)
- CHRONIS, G. AND SKUBIC, M., 2004. Robot navigation using qualitative landmark states from sketched route maps. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, vol. 2, 1530–1535. IEEE. (cited on page 47)
- CLEMENTINI, E., 2013. Directional relations and frames of reference. *GeoInformatica*, 17, 2 (2013), 235–255. (cited on pages 14 and 25)
- CLEMENTINI, E. AND BILLEN, R., 2006. Modeling and computing ternary projective relations between regions. *IEEE Transactions on Knowledge and Data Engineering*, 18, 6 (2006), 799–814. (cited on pages 23 and 26)
- CLEMENTINI, E.; DI FELICE, P.; AND HERNÁNDEZ, D., 1997. Qualitative representation of positional information. *Artificial intelligence*, 95, 2 (1997), 317–356. (cited on pages 12, 20, and 85)

- COHN, A. G.; BENNETT, B.; GOODAY, J.; AND GOTTS, N. M., 1997. Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 1, 3 (1997), 275–316. (cited on page 10)
- COHN, A. G.; LI, S.; LIU, W.; AND RENZ, J., 2014. Reasoning about topological and cardinal direction relations between 2-dimensional spatial objects. *Journal of Artificial Intelligence Research*, 51 (2014), 493–532. (cited on page 14)
- COHN, A. G. AND RENZ, J., 2008. Qualitative spatial representation and reasoning. In *Handbook of Knowledge Representation*, 551–596. Elsevier. doi:10.1016/S1574-6526(07)03013-1. [https://doi.org/10.1016/S1574-6526\(07\)03013-1](https://doi.org/10.1016/S1574-6526(07)03013-1). (cited on pages 2, 9, 14, 23, 47, and 66)
- COHN, A. G.; RENZ, J.; AND SRIDHAR, M., 2012. Thinking inside the box: A comprehensive spatial representation for video analysis. In *Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*. (cited on pages 14 and 85)
- CONDOTTA, J.-F.; SAADE, M.; AND LIGOZAT, G., 2006. A generic toolkit for n-ary qualitative temporal and spatial calculi. In *Thirteenth International Symposium on Temporal Representation and Reasoning (TIME'06)*, 78–86. IEEE. (cited on pages 15, 19, and 68)
- DANIEL, M.-P. AND DENIS, M., 1998. Spatial descriptions as navigational aids: A cognitive analysis of route directions. *Kognitionswissenschaft*, 7, 1 (1998), 45–52. (cited on page 47)
- DANIELE, A. F.; BANSAL, M.; AND WALTER, M. R., 2017. Navigational instruction generation as inverse reinforcement learning with neural machine translation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 109–118. IEEE. (cited on pages 47 and 58)
- DIJKSTRA, E. W., 1959. A note on two problems in connexion with graphs. *Numerische mathematik*, 1, 1 (1959), 269–271. (cited on pages 44, 45, and 56)
- DUBBA, K. S.; COHN, A. G.; HOGG, D. C.; BHATT, M.; AND DYLLA, F., 2015. Learning relational event models from video. *Journal of Artificial Intelligence Research*, 53 (2015), 41–90. (cited on pages 21, 67, 68, 69, 75, 76, 84, 87, and 92)
- DUCKHAM, M. AND KULIK, L., 2003. “simplest” paths: automated route selection for navigation. In *International Conference on Spatial Information Theory*, 169–185. Springer. (cited on pages 44, 45, and 46)
- DUCKHAM, M.; WINTER, S.; AND ROBINSON, M., 2010. Including landmarks in routing instructions. *Journal of Location Based Services*, 4, 1 (2010), 28–52. (cited on pages 46 and 47)

- DUCKWORTH, P.; HOGG, D. C.; AND COHN, A. G., 2019. Unsupervised human activity analysis for intelligent mobile robots. *Artificial Intelligence*, 270 (2019), 67–92. (cited on pages 20, 67, and 70)
- DYLLA, F., 2009. Qualitative spatial reasoning for navigating agents. *Behaviour Monitoring and Interpretation: Ambient Assisted Living*, B. Gottfried and H. Aghajan, Eds. IOS Press, (2009), 98–128. (cited on pages xviii, 3, 17, 18, 19, and 47)
- DYLLA, F.; FROMMBERGER, L.; WALLGRÜN, J. O.; AND WOLTER, D., 2006. Sparq: A toolbox for qualitative spatial representation and reasoning. In *Qualitative Constraint Calculi: Application and Integration, Workshop at KI*, 79–90. (cited on pages 19 and 68)
- DYLLA, F.; LEE, J. H.; MOSSAKOWSKI, T.; SCHNEIDER, T.; DELDEN, A. V.; VEN, J. V. D.; AND WOLTER, D., 2017. A survey of qualitative spatial and temporal calculi: algebraic and computational properties. *ACM Computing Surveys (CSUR)*, 50, 1 (2017), 7. (cited on pages xvii, xviii, 2, 4, 6, 9, 14, 15, and 85)
- EGENHOFER, M. J. AND HERRING, J., 1990. Categorizing binary topological relations between regions, lines, and points in geographic databases. *The*, 9, 94-1 (1990), 76. (cited on page 10)
- EHRLINGER, L. AND WÖSS, W., 2016. Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48 (2016), 1–4. (cited on page 98)
- EVANS, R. AND GREFFENSTETTE, E., 2018. Learning explanatory rules from noisy data. *Journal of Artificial Intelligence Research*, 61 (2018), 1–64. (cited on page 98)
- FELLNER, I.; HUANG, H.; AND GARTNER, G., 2017. “turn left after the wc, and use the lift to go to the 2nd floor” generation of landmark-based route instructions for indoor navigation. *ISPRS International Journal of Geo-Information*, 6, 6 (2017), 183. (cited on page 46)
- FERNYHOUGH, J.; COHN, A. G.; AND HOGG, D. C., 2000. Constructing qualitative event models automatically from video input. *Image and Vision Computing*, 18, 2 (2000), 81–103. (cited on pages xviii and 19)
- FOGLIARONI, P.; WALLGRÜN, J. O.; CLEMENTINI, E.; TARQUINI, F.; AND WOLTER, D., 2009. A qualitative approach to localization and navigation based on visibility information. In *International Conference on Spatial Information Theory*, 312–329. Springer. (cited on pages 18 and 47)
- FRANK, A., 1998. Formal models for cognition—taxonomy of spatial location description and frames of reference. In *Spatial Cognition*, 293–312. Springer. (cited on pages 23 and 27)
- FRANK, A. U., 1991. Qualitative spatial reasoning with cardinal directions. In 7. *Österreichische Artificial-Intelligence-Tagung/Seventh Austrian Conference on Artificial Intelligence*, 157–167. Springer. (cited on pages xvii, 10, 11, 16, 23, 24, 25, and 26)

-
- FREKSA, C., 1992a. Temporal reasoning based on semi-intervals. *Artificial intelligence*, 54, 1-2 (1992), 199–227. (cited on page 17)
- FREKSA, C., 1992b. Using orientation information for qualitative spatial reasoning. In *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, International Conference GIS - From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning, Pisa, Italy, September 21-23, 1992, Proceedings*, 162–178. (cited on pages 11, 23, and 24)
- FREKSA, C., 1992c. Using orientation information for qualitative spatial reasoning. In *Theories and methods of spatio-temporal reasoning in geographic space*, 162–178. Springer. (cited on pages 16, 18, 47, 49, and 86)
- FUKUI, H.; HIRAKAWA, T.; YAMASHITA, T.; AND FUJIYOSHI, H., 2019. Attention branch network: Learning of attention mechanism for visual explanation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10705–10714. (cited on pages 69 and 72)
- GALATA, A.; COHN, A.; MAGEE, D.; AND HOGG, D., 2002. Modeling interaction using learnt qualitative spatio-temporal relations and variable length markov models. In *ECAI*, 741–745. (cited on pages 20, 86, and 87)
- GANTNER, Z.; WESTPHAL, M.; AND WÖLFL, S., 2008. Gqr-a fast reasoner for binary qualitative constraint calculi. In *AAAI Workshop on Spatial and Temporal Reasoning*. AAAI Chicago (IL). (cited on page 68)
- GATSOUKIS, Y.; ALOMARI, M.; BURBRIDGE, C.; DONDRUP, C.; DUCKWORTH, P.; LIGHTBODY, P.; HANHEIDE, M.; HAWES, N.; HOGG, D.; COHN, A.; ET AL., 2016. Qsrlib: a software library for online acquisition of qualitative spatial relations from video. (2016). (cited on pages 10, 19, 66, 68, 70, 78, 86, 87, and 88)
- GEREVINI, A. AND RENZ, J., 2002a. Combining topological and size information for spatial reasoning. *Artificial Intelligence*, 137, 1-2 (2002), 1–42. (cited on pages 13 and 14)
- GEREVINI, A. AND RENZ, J., 2002b. Combining topological and size information for spatial reasoning. *Artif. Intell.*, 137, 1-2 (2002), 1–42. (cited on page 17)
- HAQUE, S.; KULIK, L.; AND KLIPPEL, A., 2006. Algorithms for reliable navigation and wayfinding. In *International Conference on Spatial Cognition*, 308–326. Springer. (cited on pages 45, 56, 57, 58, and 62)
- HERNANDEZ, D., 1994. *Qualitative representation of spatial knowledge*, vol. 804. Springer Science & Business Media. (cited on page 25)
- HERNANDEZ, D.; CLEMENTINI, E.; AND DI FELICE, P., 1995. Qualitative distances. In *International Conference on Spatial Information Theory*, 45–57. Springer. (cited on pages 50 and 97)

-
- HUA, H.; RENZ, J.; AND GE, X., 2018. Qualitative representation and reasoning over direction relations across different frames of reference. In *Sixteenth International Conference on Principles of Knowledge Representation and Reasoning*. (cited on pages 49, 53, and 97)
- HUA, H.; ZHANG, P.; AND RENZ, J., 2019. Qualitative place maps for landmark-based localization and navigation in gps-denied environments. In *SIGSPATIAL*, 23–32. ACM. (cited on page 97)
- ISLI, A. AND COHN, A. G., 2000. A new approach to cyclic ordering of 2d orientations using ternary relation algebras. *Artificial Intelligence*, 122, 1-2 (2000), 137–187. (cited on pages 15 and 16)
- ISLI, A.; HAARSLEV, V.; MÖLLER, R.; ET AL., 2001. *Combining cardinal direction relations and relative orientation relations in qualitative spatial reasoning*. Univ., Bibliothek des Fachbereichs Informatik. (cited on pages 14, 17, 23, 26, and 34)
- JETLEY, S.; LORD, N. A.; LEE, N.; AND TORR, P. H., 2018. Learn to pay attention. *arXiv preprint arXiv:1804.02391*, (2018). (cited on pages 69 and 72)
- KALOGITON, V.; WEINZAEPEL, P.; FERRARI, V.; AND SCHMID, C., 2017. Action tubelet detector for spatio-temporal action localization. In *Proceedings of the IEEE International Conference on Computer Vision*, 4405–4413. (cited on page 69)
- KLIPPEL, A.; TAPPE, H.; KULIK, L.; AND LEE, P. U., 2005. Wayfinding choremes a language for modeling conceptual route knowledge. *Journal of Visual Languages & Computing*, 16, 4 (2005), 311–329. (cited on pages 18 and 49)
- KOPPULA, H. S.; GUPTA, R.; AND SAXENA, A., 2013. Learning human activities and object affordances from rgb-d videos. *The International Journal of Robotics Research*, 32, 8 (2013), 951–970. (cited on pages 68 and 76)
- KÖPÜKLÜ, O.; WEI, X.; AND RIGOLL, G., 2019. You only watch once: A unified cnn architecture for real-time spatiotemporal action localization. *arXiv preprint arXiv:1911.06644*, (2019). (cited on page 69)
- KORDJAMSHIDI, P.; HOIS, J.; VAN OTTERLO, M.; AND MOENS, M.-F., 2013. Learning to interpret spatial natural language in terms of qualitative spatial relations. (2013). (cited on pages 3 and 19)
- KREUTZMANN, A. AND WOLTER, D., 2014. Qualitative spatial and temporal reasoning with and/or linear programming. In *ECAI 2014*, 495–500. IOS Press. (cited on page 25)
- KREUTZMANN, A.; WOLTER, D.; DYLLA, F.; AND LEE, J. H., 2013. Towards safe navigation by formalizing navigation rules. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 7, 2 (2013), 161–168. (cited on page 3)

-
- KRIEG-BRÜCKNER, B. AND SHI, H., 2006. Orientation calculi and route graphs: Towards semantic representations for route descriptions. In *International Conference on Geographic Information Science*, 234–250. Springer. (cited on pages xviii, 18, and 47)
- KYRIACOU, T.; BUGMANN, G.; AND LAURIA, S., 2005. Vision-based urban navigation procedures for verbally instructed robots. *Robotics and Autonomous Systems*, 51, 1 (2005), 69–80. (cited on page 47)
- LECUN, Y.; BENGIO, Y.; AND HINTON, G., 2015. Deep learning. *nature*, 521, 7553 (2015), 436–444. (cited on pages 5 and 66)
- LEE, J. H.; RENZ, J.; WOLTER, D.; ET AL., 2013. Starvars-effective reasoning about relative directions. In *IJCAI*, 976–982. (cited on pages 16, 23, 24, 25, 26, 27, and 28)
- LEVINSON, S. C., 1996. Frames of reference and molyneux’s question: Crosslinguistic evidence. *Language and space*, 109 (1996), 169. (cited on pages 16, 24, 27, and 37)
- LEVITT, T. S., 1990. Qualitative navigation for mobile robots. *Int. J. Artificial Intelligence*, 44 (1990), 305–360. (cited on page 47)
- LI, B.; WU, W.; WANG, Q.; ZHANG, F.; XING, J.; AND YAN, J., 2019. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In *CVPR*, 4282–4291. (cited on page 70)
- LI, S., 2007. Combining topological and directional information for spatial reasoning. In *IJCAI*, vol. 137, 435–440. (cited on page 14)
- LI, S. AND COHN, A. G., 2012. Reasoning with topological and directional spatial information. *Computational Intelligence*, 28, 4 (2012), 579–616. (cited on page 14)
- LI, S.; LIU, W.; AND WANG, S., 2013. Qualitative constraint satisfaction problems: An extended framework with landmarks. *Artificial Intelligence*, 201 (2013), 32–58. (cited on page 48)
- LIGOZAT, G., 1998. Reasoning about cardinal directions. *J. Vis. Lang. Comput.*, 9, 1 (1998), 23–44. (cited on page 16)
- LIGOZAT, G. F., 1993. Qualitative triangulation for spatial reasoning. In *European Conference on Spatial Information Theory*, 54–68. Springer. (cited on pages 11, 23, 24, 25, 26, and 40)
- LIN, J.; GAN, C.; AND HAN, S., 2019. Tsm: Temporal shift module for efficient video understanding. In *ICCV*, 7083–7093. (cited on page 66)
- LIU, F.; VASARDANI, M.; AND BALDWIN, T., 2014. Automatic identification of locative expressions from social media text: A comparative analysis. In *Proceedings of the 4th International Workshop on Location and the Web*, 9–16. ACM. (cited on pages 19, 47, and 51)

-
- LIU, W.; LI, S.; RENZ, J.; ET AL., 2009. Combining rcc-8 with qualitative direction calculi: Algorithms and complexity. In *IJCAI*, vol. 2009, 854–859. (cited on page 14)
- LÜCKE, D.; MOSSAKOWSKI, T.; AND MORATZ, R., 2011. Streets to the opra finding your destination with imprecise knowledge. In *Proc. IJCAI Workshop on Benchmarks and Applications of Spatial Reasoning*, 25–32. Citeseer. (cited on pages 18, 47, and 50)
- MARCUS, G., 2020. The next decade in ai: four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*, (2020). (cited on page 97)
- MICHON, P.-E. AND DENIS, M., 2001. When and why are visual landmarks used in giving directions? In *International Conference on Spatial Information Theory*, 292–305. Springer. (cited on pages 44 and 46)
- MORATZ, R., 2006. Representing relative direction as a binary relation of oriented points. In *ECAI*, vol. 6, 407–411. (cited on pages 16, 23, 24, 25, and 26)
- MORATZ, R.; DYLLA, F.; AND FROMMBERGER, L., 2005. A relative orientation algebra with adjustable granularity. In *Proceedings of the Workshop on Agents in Real-Time and Dynamic Environments (IJCAI 05)*, vol. 21, 22. (cited on pages 12 and 27)
- MORATZ, R. AND RAGNI, M., 2008. Qualitative spatial reasoning about relative point position. *Journal of Visual Languages & Computing*, 19, 1 (2008), 75–98. (cited on pages 12, 14, 20, 23, 24, 25, 26, 38, 39, and 85)
- MOSSAKOWSKI, T. AND MORATZ, R., 2012. Qualitative reasoning about relative direction of oriented points. *Artif. Intell.*, 180-181 (2012), 34–45. (cited on pages 26, 27, and 40)
- NEBEL, B. AND BÜRCKERT, H.-J., 1995. Reasoning about temporal relations: a maximal tractable subclass of allen’s interval algebra. *Journal of the ACM (JACM)*, 42, 1 (1995), 43–66. (cited on page 16)
- OSSWALD, S.; KRETZSCHMAR, H.; BURGARD, W.; AND STACHNISS, C., 2014. Learning to give route directions from human demonstrations. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 3303–3308. IEEE. (cited on pages 47 and 58)
- OUYANG, J.; FU, Q.; AND LIU, D., 2007. A model for representing topological relations between simple concave regions. In *International Conference on Computational Science*, 160–167. Springer. (cited on page 10)
- PERICO, D. H.; SANTOS, P. E.; AND BIANCHI, R. A., 2021. Guided navigation from multiple viewpoints using qualitative spatial reasoning. *Spatial Cognition & Computation*, 21, 2 (2021), 143–172. (cited on page 18)

-
- RANDELL, D. A.; COHN, A. G.; AND CUI, Z., 1992a. Computing transitivity tables: A challenge for automated theorem provers. In *International Conference on Automated Deduction*, 786–790. Springer. (cited on page 15)
- RANDELL, D. A.; CUI, Z.; AND COHN, A. G., 1992b. A spatial logic based on regions and connection. *KR*, 92 (1992), 165–176. (cited on pages xvii, 2, 9, 10, 20, 26, and 66)
- RENEGAR, J., 1992. On the computational complexity and geometry of the first-order theory of the reals. part i: Introduction. preliminaries. the geometry of semi-algebraic sets. the decision problem for the existential theory of the reals. *Journal of symbolic computation*, 13, 3 (1992), 255–299. (cited on page 39)
- RENZ, J., 1999. Maximal tractable fragments of the region connection calculus: A complete analysis. In *IJCAI*, 448–455. (cited on page 16)
- RENZ, J. AND MITRA, D., 2004. Qualitative direction calculi with arbitrary granularity. In *PRICAI*, vol. 3157, 65–74. (cited on pages 11, 16, 23, 24, 25, 26, 30, 49, and 85)
- RENZ, J. AND NEBEL, B., 1999. On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the region connection calculus. *Artificial Intelligence*, 108, 1-2 (1999), 69–123. (cited on pages 15 and 38)
- RENZ, J. AND NEBEL, B., 2007. Qualitative spatial reasoning using constraint calculi. In *Handbook of Spatial Logics*, 161–215. Springer. doi:10.1007/978-1-4020-5587-4_4. https://doi.org/10.1007/978-1-4020-5587-4_4. (cited on pages 1, 2, 3, 4, 9, 14, 15, 17, 38, 47, 66, 83, and 85)
- RICHTER, D.; WINTER, S.; RICHTER, K.-F.; AND STIRLING, L., 2013. Granularity of locations referred to by place descriptions. *Computers, Environment and Urban Systems*, 41 (2013), 88–99. (cited on pages 51 and 97)
- RICHTER, K.-F. AND DUCKHAM, M., 2008. Simplest instructions: Finding easy-to-describe routes for navigation. In *International Conference on Geographic Information Science*, 274–289. Springer. (cited on pages 44 and 46)
- RUDIN, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 5 (2019), 206–215. (cited on page 66)
- RUSSO, D.; ZLATANOVA, S.; AND CLEMENTINI, E., 2014. Route directions generation using visible landmarks. In *Proceedings of the sixth ACM SIGSPATIAL international workshop on indoor spatial awareness*, 1–8. ACM. (cited on page 46)
- SCIVOS, A. AND NEBEL, B., 2001. Double-crossing: Decidability and computational complexity of a qualitative calculus for navigation. In *Spatial Information Theory: Foundations of Geographic Information Science, International Conference, COSIT 2001, Morro Bay, CA, USA, September 19-23, 2001, Proceedings*, 431–446. (cited on pages 15, 31, 34, and 39)

-
- SCIVOS, A. AND NEBEL, B., 2004. The finest of its class: The natural point-based ternary calculus Ir for qualitative spatial reasoning. In *International Conference on Spatial Cognition*, 283–303. Springer. (cited on page 34)
- SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; AND BATRA, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626. (cited on pages 69 and 72)
- SERRÃO, M.; RODRIGUES, J. M.; AND DU BUF, J. H., 2014. Navigation framework using visual landmarks and a gis. *Procedia Computer Science*, 27 (2014), 28–37. (cited on page 46)
- SHAH, D. C. AND CAMPBELL, M. E., 2013. A qualitative path planner for robot navigation using human-provided maps. *The International Journal of Robotics Research*, 32, 13 (2013), 1517–1535. (cited on page 47)
- SHARIFE, A. R. B.; EGENHOFER, M. J.; AND MARK, D. M., 1998. Natural-language spatial relations between linear and areal objects: the topology and metric of english-language terms. *International journal of geographical information science*, 12, 3 (1998), 215–245. (cited on page 74)
- SIGURDSSON, G. A.; VAROL, G.; WANG, X.; FARHADI, A.; LAPTEV, I.; AND GUPTA, A., 2016. Hollywood in homes: Crowdsourcing data collection for activity understanding. In *European Conference on Computer Vision*, 510–526. Springer. (cited on page 80)
- SKIADOPOULOS, S. AND KOUBARAKIS, M., 2004. Composing cardinal direction relations. *Artificial Intelligence*, 152, 2 (2004), 143–171. (cited on pages 12 and 86)
- SOKEH, H. S.; GOULD, S.; AND RENZ, J., 2014. Determining interacting objects in human-centric activities via qualitative spatio-temporal reasoning. In *ACCV*, 550–563. Springer. (cited on page 69)
- SRIDHAR, M.; COHN, A. G.; AND HOGG, D. C., 2010. Unsupervised learning of event classes from video. In *AAAI*. (cited on pages 20, 21, and 67)
- SRIDHAR, M.; COHN, A. G.; AND HOGG, D. C., 2011a. Benchmarking qualitative spatial calculi for video activity analysis. In *Proceedings ijcai workshop benchmarks and applications of spatial reasoning*, 15–20. Leeds. (cited on pages xviii, 13, 20, 21, and 84)
- SRIDHAR, M.; COHN, A. G.; AND HOGG, D. C., 2011b. From video to rcc8: exploiting a distance based semantics to stabilise the interpretation of mereotopological relations. In *COSIT*, 110–125. Springer. (cited on pages 17 and 69)
- TAN, M.; PANG, R.; AND LE, Q. V., 2020. Efficientdet: Scalable and efficient object detection. In *CVPR*, 10781–10790. (cited on page 70)

-
- TANG, K.; YAO, B.; FEI-FEI, L.; AND KOLLER, D., 2013. Combining the right features for complex event recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, 2696–2703. (cited on page 89)
- TARJAN, R., 1972. Depth-first search and linear graph algorithms. *SIAM journal on computing*, 1, 2 (1972), 146–160. (cited on page 54)
- TAYYUB, J.; TAVANAI, A.; GATSOUKIS, Y.; COHN, A. G.; AND HOGG, D. C., 2014. Qualitative and quantitative spatio-temporal relations in daily living activity recognition. In *Asian Conference on Computer Vision*, 115–130. Springer. (cited on pages 20, 67, 76, 84, 86, 87, and 92)
- TENBRINK, T. AND KUHN, W., 2011. A model of spatial reference frames in language. *Spatial Information Theory*, (2011), 371–390. (cited on page 24)
- TRAN, D.; WANG, H.; TORRESANI, L.; AND FEISZLI, M., 2019. Video classification with channel-separated convolutional networks. In *ICCV*, 5552–5561. (cited on page 66)
- VAN BEEK, P., 1992. Reasoning about qualitative temporal information. *Artificial intelligence*, 58, 1-3 (1992), 297–326. (cited on pages 13, 15, and 85)
- VAN DE WEGHE, N.; COHN, A.; DE TRE, G.; AND DE MAEYER, P., 2006. A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems. *Control and Cybernetics*, 35, 1 (2006), 97–119. (cited on pages 13, 20, and 85)
- VASARDANI, M.; TIMPF, S.; WINTER, S.; AND TOMKO, M., 2013. From descriptions to depictions: A conceptual framework. In *International Conference on Spatial Information Theory*, 299–319. Springer. (cited on pages 19 and 47)
- VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; AND POLOSUKHIN, I., 2017. Attention is all you need. In *Advances in neural information processing systems*, 5998–6008. (cited on pages 69 and 98)
- VILAIN, M.; KAUTZ, H.; AND VAN BEEK, P., 1990. Constraint propagation algorithms for temporal reasoning: A revised report. In *Readings in qualitative reasoning about physical systems*, 373–381. Elsevier. (cited on page 16)
- WANG, Q.; ZHANG, L.; BERTINETTO, L.; HU, W.; AND TORR, P. H., 2019. Fast online object tracking and segmentation: A unifying approach. In *CVPR*, 1328–1338. (cited on page 70)
- WESTPHAL, M. AND RENZ, J., 2011. Evaluating and minimizing ambiguities in qualitative route instructions. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 171–180. ACM. (cited on pages 44 and 58)

-
- WESTPHAL, M.; WÖLFL, S.; AND GANTNER, Z., 2009. Gqr: A fast solver for binary qualitative constraint networks. In *AAAI Spring Symposium: Benchmarking of Qualitative Spatial and Temporal Reasoning Systems*, 51–52. (cited on page 19)
- WIJEWARDENA, G. G.; VASARDANI, M.; AND WINTER, S., 2016. Towards indoor localization and navigation independent of sensor based technologies. In *Eighth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*. ACM Press. (cited on page 45)
- WINTER, S., 2002. Modeling costs of turns in route planning. *GeoInformatica*, 6, 4 (2002), 345–361. (cited on page 46)
- WINTER, S.; RICHTER, K.-F.; SHI, M.; AND GAN, H.-S., 2011. Get me out of here: Collaborative evacuation based on local knowledge. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, 35–42. ACM. (cited on page 45)
- WINTER, S.; TOMKO, M.; VASARDANI, M.; RICHTER, K.-F.; KHOSHELHAM, K.; AND KALANTARI, M., 2019. Infrastructure-independent indoor localization and navigation. *ACM Computing Surveys*, 52, 3 (2019), 61:1–61:24. (cited on page 45)
- WOLTER, D. AND LEE, J. H., 2010. Qualitative reasoning with directional relations. *Artif. Intell.*, 174, 18 (2010), 1498–1507. (cited on page 16)
- WOLTER, D. AND WALLGRÜN, J. O., 2013. Qualitative spatial reasoning for applications: New challenges and the sparq toolbox. In *Geographic Information Systems: Concepts, Methodologies, Tools, and Applications*, 1639–1664. IGI Global. (cited on pages 16, 18, and 42)
- YOUNG, J. AND HAWES, N., 2015. Learning by observation using qualitative spatial relations. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, 745–751. International Foundation for Autonomous Agents and Multiagent Systems. (cited on pages 21 and 67)
- ZHUO, T.; CHENG, Z.; ZHANG, P.; WONG, Y.; AND KANKANHALLI, M., 2019. Explainable video action reasoning via prior knowledge and state transitions. In *Proceedings of the 27th ACM International Conference on Multimedia*, 521–529. (cited on pages 66, 68, and 76)
- ZIMMERMANN, K. AND FREKSA, C., 1996. Qualitative spatial reasoning using orientation, distance, and path knowledge. *Appl. Intell.*, 6, 1 (1996), 49–58. (cited on pages 15 and 31)