Visualization for maritime situational awareness

Visualization
for
Maritime Situational Awareness

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College van Promoties, in het openbaar te verdedigen op maandag 23 november 2015 om 16:00 uur

door

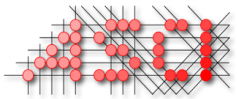
Roeland Johannus Scheepens

geboren te Cuckfield, Verenigd Koninkrijk

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter	
promotor:	prof.dr.ir. J.J. van Wijk
copromotor:	dr.ir. H.M.M. van de Wetering
leden:	prof.dr. G. Andrienko (City University London & Fraunhofer IAIS) prof.dr. J.A. Dykes (City University London) prof.dr.ir. B. de Vries dr. C. Hurter (French Civil Aviation University) prof.dr. N. Van der Weghe (Universiteit Gent)

This thesis was supported by the Dutch national program COMMIT. The research work was carried out as part of the Metis project under the responsibility of Embedded Systems Innovation by TNO with Thales Nederland B.V. as the carrying industrial partner.



This work was carried out in the ASCI graduate school. ASCI dissertation series number 335.

printing Gildeprint Drukkerijen, Enschede

cover design Walter Scheepens

ISBN 978-90-386-3931-4

This thesis was prepared using LaTeX typesetting. The cover was prepared using Adobe Photoshop and Adobe InDesign. The figures were prepared using Adobe Illustrator and Adobe Photoshop. The contents of the figures are taken from working prototypes written by Roeland Scheepens in C++ and NVIDIA Cg in the Microsoft Visual Studio Express IDE using OpenGL. Graphs were created using Mathematica and statistical result were obtained using IBM SPSS Statistics. Videos were prepared using Camtasia Studio.

No vessels were harmed in the making of this thesis.

A catalogue record is available from the Eindhoven University of Technology Library.

Preface

Having written 184 318 lines of C++ and 27 659 lines of shader code over the past 4 years, I can safely say that visualization is a hands-on research area and this is, in part, what has made it so enjoyable. I was initially attracted to the visualization group due to a great interest in computer graphics. Through the supervision of dr. Andrei Jalba, in a course aptly named *Additional Component Computer Graphics*, and later a talk with my future supervisor, dr. Huub van de Wetering, I finally decided to do my master project in the visualization group. During this project, which was formulated as a GPGPU (General-Purpose computing on Graphics Processing Units) assignment, my appreciation for pure visualization grew and lead to my interest in doing a PhD in the field. My decision to pursue a PhD in visualization is one I do not regret to this day.

In the following, I would like to thank all the people that have played a role in my PhD-life directly and indirectly, and those that played some other role in my life outside of my PhD. First and foremost, I would like to thank my promotor, prof.dr. Jarke J. van Wijk, for giving me the opportunity to do this PhD project. It was not until our first conference that we realized the full extent of this opportunity. Jack, thank you for being so accessible, and always being ready with new ideas and insights. Secondly, I would like to thank my supervisor, dr. Huub van de Wetering for his excellent daily supervision. Huub, thank you for your solid mathematical support, proofreading page after page of material without missing a single detail, and your many anecdotes and stories. I also owe my gratitude to the other members of my doctoral committee, prof.dr. Gennady Andrienko, prof.dr. Jason Dykes, prof.dr. Bauke de Vries, dr. Christophe Hurter, and prof.dr. Nico Van de Weghe, for agreeing to be my opponents in the oral defense, and for their valuable feedback that further improved my thesis. Also, I would like to thank my *predecessor*, Niels Willems, with whom I collaborated and who paved the way for me in the Poseidon project, for introducing me to the art of paper writing, and introducing me to many researchers who would later play a role in my research.

Part of what made my time as a PhD so enjoyable were the excellent colleagues in the visualization group and beyond. I would like to thank my colleagues dr. Kasper Dinkla, Stef van den Elzen, and Mickeal Verschoor for many years of coffee and interesting discussions. Also, I would like to thank all the other people at the visualization department who made my stay pleasant during my master and my PhD: dr. Andrei Jalba, dr. Danny Holten, dr. Jing Li, dr. Michel Westenberg, and Meivan Cheng. And I wish my colleagues that are currently pursuing a PhD the best of luck in their studies: Alberto Corvo, Bram Cappers, Martijn van Dortmont, Paul van der Corput, and Renata Raidou. Additionally, I would like to thank my colleagues from the Algorithms department for interesting discussions and/or being travel companions: Aleksandar Markovic, Ali Mehrabi Davoodabadi, Arthur

Goethem, prof.dr. Bettina Speckmann, dr. Herman Haverkort, dr. Kevin Buchin, dr. Kevin Verbeek, dr. Maike Buchin, prof.dr. Mark de Berg, Maximilian Konzack, Quirijn Bouts, and dr. Wouter Meulemans. And of course thanks to all other researchers I met and collaborated with over the years.

I did most of my work in cooperation with other researchers and partners in the Metis project. I would like to thank the people I collaborated with in the project for their efforts towards making Metis a success: André Bonhof (Thales), Jesper Hoeksema (VU), Marina Velikova (RUN), Peter Novak (TUD), Piërre America (TNO-ESI), Piërre van de Laar (TNO-ESI), Steffen Michels (RUN), and Teun Hendriks (TNO-ESI). With special thanks to Bas Huijbrechts (TNO-ESI) and David Watts (TNO-ESI) for running the project, and to Tom Regelink (TNO-ESI/Altran) for being the glue that made the technical demos a reality.

Furthermore, I would like to thank everyone not mentioned before who touched my life in some significant way over the past four years (in alphabetic order): Ana Teixeira, Anouk Katzenbauer, Carla Xavier, Carlijn Dohmen, David van Heijnsbergen, Declan Cockburn, Eileen Pereira, Emídio Carvalho (Bem hajas por tudo e, especialmente, QSF), Emmy de Jong, Luísa Benta (Bem hajas por tudo e desejote paz onde quer que estejas), Paddy Carr, Paula Lopes (a minha professora), Páuraig Babbington, and Sarah Thomas. And of course, the guys from my band, Horns: Cor Jolie, Patrick Veen, and Seth van Ringelenstijn.

Natuurlijk kunnen mijn ouders, Ans en Walter, en mijn twee broers, Walter en Bertram, hier niet ontbreken. Pa, ma, bedankt dat jullie mij altijd bijgestaan hebben, financieel en anderszins, en altijd in mij geloofd hebben ondanks de beren op de weg. Walter, bedankt voor het ontwerpen van de kaft van dit proefschrift en bedankt dat je mij geïntroduceerd en bijgestaan hebt met programmeren op vroege leeftijd. Bertram, bedankt voor je humor en dat je altijd weet welke film ik wil zien.

E Finalmente eu queria dizer algumas coisas à minha namorada Brasileira, Juliana Noleto, em Português. Estivemos longe por muito tempo, mas, a vida queira, isso vá mudar em breve. Obrigado pela tua paciência e amizade, obrigado por deixares-me conhecer o Brasil, obrigado por aceitares que eu não sempre tenho tempo por causa de escrever a minha tese, e obrigado por seres.

Obrigado à família Noleto pela sua hospitalidade. Gostei muito de ficar na vossa casa e vou voltar ao Brasil em breve. Bem hajam!

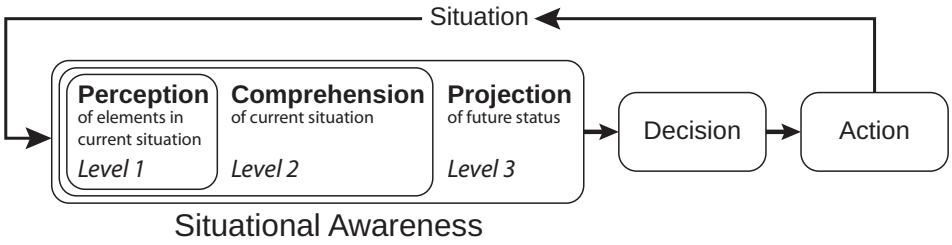
E obrigado à Sandra Gama e ao Emídio Carvalho por corrigirem o meu sumário Português.

Roeland Scheepens
Eindhoven, August 2015

Contents

1

Introduction



In this chapter we introduce the Maritime safety and security domain and the concept of situational awareness. We next introduce how visualization fits into this and explain the context of the project in which the work described in this thesis has been developed. Finally, we present our research goal and give an overview of the work discussed in this thesis.

"What is it doing? Why is it doing that? What will it do next?" - [69]

1

1.1 Maritime Safety and Security Domain

The Maritime Safety and Security Domain (MSSD) is concerned with protecting shipping, infrastructure at sea, economic assets such as oil or gas platforms, and environmental assets such as nature reserves against a broad range of threats [191, 180]. These threats include illegal activities such as smuggling drugs or arms, environmental degradation, illegal fishing, human trafficking, piracy, or terrorism [180, 45, 24]. The responsibility of safeguarding the maritime domain against such threats lies with international initiatives [155], national navies, law enforcement agencies such as the coast guard [190, 191, 192], and customs and border protection agencies. In this thesis we focus specifically on the context of law enforcement. We do not deal with Vessel Traffic Services (VTS) or the warfare domain.

In the maritime safety and security domain, operators stationed on board of law enforcement vessels or in land-based stations monitor some area of interest for suspicious behaviour and anomalies. These areas of interest typically contain tens to hundreds of vessels. The operators can have a large number of responsibilities and objectives, such as managing emergency situations, enforcing the law, protecting economic interests, and controlling the border [191]. These responsibilities are defined by the mission the operator is on and the law enforcement agency they are working for. When a vessel is behaving suspiciously or is breaking some law, the operator can respond in a number of ways. The operator can contact the suspect vessel, the operator can fine the vessel, send a law enforcement vessel to intercept, board, and search the vessel, or the operator can mark the vessel as suspect so it can be held offshore for inspection [24]. These actions may, however, be costly, or come at the cost of human life, and resources are limited. This implies that operators need to be confident in their decision to act.

1.2 Situational Awareness

Situational Awareness[1] (SA) forms the basis for decision making. It is defined by Endsley [69] as:

> *The **perception** and **comprehension** of the current situation and the **projection** of the current situation into the future.*

The three levels of SA; *perception*, *comprehension*, and *projection into the future*, can be seen as hierarchical phases–see Figure 1.1. First, decision makers need to be able to perceive the status, attributes, and dynamics of relevant elements in the

[1] Also commonly known as Situation Awareness

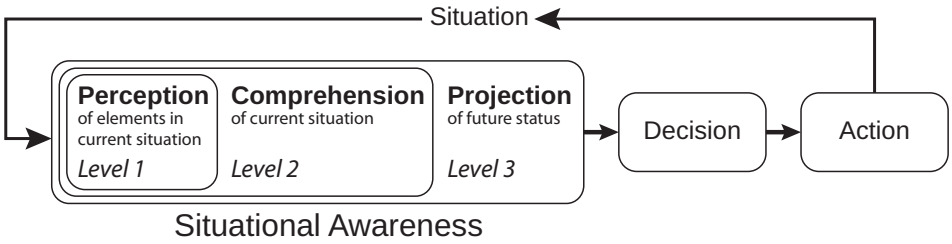


Figure 1.1: The model of situational awareness in decision making according to Endsley [69].

environment, or in other words, decision makers need to be aware of the elements in their area of interest. Given these relevant elements, decision makers need to be able to understand why elements are relevant and what their relationships are. Finally, decision makers need to be able to project the future actions of these elements into the future to decide what to do. This is, however, not a strictly linear process, but a hierarchical one [70]. All three levels are required to acquire and maintain situational awareness and to come to a decision that leads to an action [72, 69].

In the maritime domain, situational awareness helps operators to take appropriate actions in a constantly evolving situation [98]. It is vital for the early detection of suspicious situations and anomalies [160], that may evolve into a threat [153]. A suspicious situation typically involves multiple and diverse actors, mostly vessels, that may interact in complex ways. Information can be gathered on these actors to gain more insight on their identity and intent. This information can either be broadcast by the actors themselves through the Automatic Identification System (AIS) [111], gathered from multiple sensors with different properties, such as radar systems, cameras, or human observation, or actively gathered from databases or the internet by the operator [81]. This information can be further enriched using additional sources or reasoning [219, 191]. However, in current practice, operators manually ascertain the identity of vessels of interest [60] and the gathered information is fused and analyzed manually [167, 191].

Current situational awareness systems in the maritime safety and security domain provide tools for operators for focus and decision-support using configurable operational pictures [98]–see Figure 1.2. The human operators, however, still mainly do the reasoning to understand the situation themselves and provide the domain knowledge [81]. On the information retrieved about vessels, some automated reasoning can be applied that aids in comprehending the current situation [23]. Traditionally, automated reasoning or analysis is performed mostly on kinematic attributes of the vessels [173]. Decisions and reasoning on the intents of the vessels, however, are based on intuition and operational experience [45].

As more and more additional information about vessels is obtained through sensors and human observation in increasing quantity [204], fusing attributes automatically is gaining more interest [79]. Also, many efforts have been made to automate reasoning on the intents of vessels based on kinematic properties,

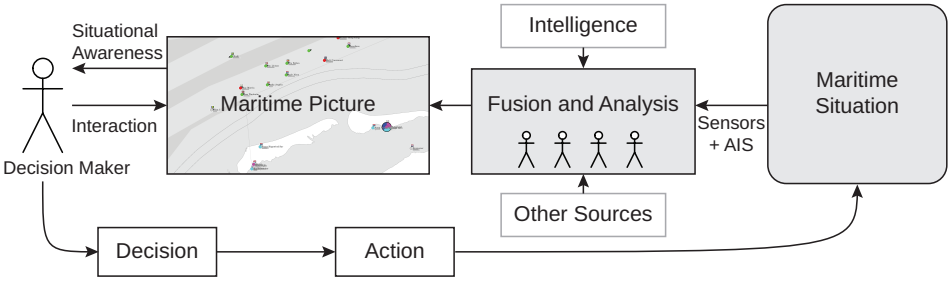


Figure 1.2: The decision-maker in the context of the maritime safety and security domain.

such as movement patterns, by learning and/or detecting anomalies [23, 157, 159, 125], or by finding suspicious behavior by applying domain-based rules or knowledge [60, 153]. Automated reasoning based on non-kinematic attributes, such as information about the vessel's crew, owner, cargo, and operator, is still an emerging field of research [137, 98, 172]. Automated reasoning requires the definition of domain rules or normal behavioral models [160]. To build such models, tooling is required to analyze and understand the maritime domain and the data available therein.

1.3 Visualization

The three main levels of situational awareness; perception, understanding, and projection into the future, can be automated to some extent as discussed in the previous section. The human operator, however, is still the decision maker. Situational awareness can be seen as the mental model of this decision maker. Additionally, to make use of the vast domain knowledge and operational experience of human operators, situational awareness systems involve the human operator in the loop [98]. This means the situational awareness system should present the current situation, its understanding of the current situation, and its projection into the future to the human operator in an understandable way. To this end, we use visualization. Visualization is a powerful tool for increasing situational awareness and reducing information overload [163, 126, 160].

Currently, (mainly kinematic) information on vessels present in the situation is *summarized* by visualizing it on a situational picture representing some geographical area of interest, often called the maritime picture [79]. Changes in the situation happen on screen slowly and subtly, *i.e.*, the visualization appears static. It is hard, however, for operators to maintain situational awareness from such an overview visualization [144]. With the rise of automated methods, some work has been done in terms of improving the maritime picture [133], or directly using visualization to help detect anomalies [161, 132, 179]. Also, some work has been done in using visualization for risk assessment [140]. In general, however, little work has been done on applying visualization to improve situational awareness

in the maritime domain. This is in part due to the relatively recent emergence of automated methods in the practice of the maritime safety and security domain.

1.4 Metis

The research described in this thesis has been a part of the Metis project [67] under the responsibility of the Embedded Systems Innovation (ESI) by TNO with Thales Nederland B.V. as the carrying industrial partner. In short, the aim of the Metis project is to enable next generation situational awareness systems for public safety and security with a case study in the maritime safety and security domain [98]. The Metis project can be roughly described by the following four research questions:

1. How to gather relevant information from unstructured sources such as news or Twitter messages [94]?
2. How to fuse information from distributed and heterogeneous sources, assess its quality, and use this to reason on the identity and intent of vessels [137, 139, 138]?
3. How to reconfigure a reasoning engine to spare sparse system resources, or to increase the quality of information [146, 147]?
4. How to present the above to a human operator to gain trust and situational awareness [175, 173, 172]?

The trend is that more and more additional information from additional sources is becoming available and is being used by operators [93]. Information is now not only gathered from trusted sources such as radar, satellites, cameras, intelligence reports, or internal databases, but can also be retrieved from external sources with varying reliability and trustworthiness. Much of the information required to gain situational awareness can be gathered from the latter sources [93]. These sources can vary from trusted databases such as IHS Fairplay [76], to less trusted online databases containing observations by enthusiasts, unstructured news sources [94], or even Twitter messages. The information from these heterogeneous sources is then fused [137] to best possibly predict the true values of the attributes in the information [206].

A typical area of interest may contain a large number of vessels. Since for each vessel a large volume of such information can be gathered and human operators are becoming an increasingly scarce resource [98], automated methods are required to aid the human operator in finding suspicious (or anomalous) vessels. Using traditional analyzers that reason on kinematic attributes and the fused attributes, a reasoning engine can reason on the intent of a vessel [139, 138]. Since there is a potentially large number of vessels to be reasoned on, the reasoning engine needs to be reconfigured where required based on available system resources and mission requirements [146]. Also, the reasoning engine may need to be reconfigured to use more sources if the quality of information is insufficient. If the

operator is to have any benefit of the automated methods mentioned above, their results need to be presented to the operators in an understandable way. Moreover, when using automated methods, the trust of the operator in the competence of the system becomes an issue. For example, to be able to make a decision with confidence, the operator needs to know not only that the vessel is considered suspicious, but also why the vessel is considered suspicious. This is where visualization plays a vital role by presenting the information in an understandable way [142, 57, 188, 175, 172].

1.5 Research Question

The main research question covered by this thesis is as follows:

> *How can operators be supported in attaining situational awareness using interactive visualization?*

In this thesis, we answer this question in four parts of which the last three parts correspond with the three levels of Situational awareness as defined by Endsley [69]–see Figure 1.3:

- **i** **Analysis**: How can we provide tools to *analyze* and summarize patterns, enabling domain experts to find critical areas and to verify what normal or anomalous behavior is? Automated methods for situational awareness, especially in the comprehension and projection level, often require expert rules, domain knowledge, or normal models [160] supplied by domain experts. The above tools can aid analysts in building and verifying these expert rules, domain knowledge, and normal models.
- **ii** **Perception**: How can we help the operator in *perceiving* a situation? The operator needs to be able to perceive the status, attributes, and dynamics of relevant elements in the environment [69]. For example, which of the hundreds of vessels on screen require attention from the operator and which do not?
- **iii** **Comprehension**: How can we help the operator in *comprehending* a situation? Perceiving relevant elements alone is not enough. The operator must also be enabled to understand why elements, their relationships, and events are relevant. For example, why is a vessel suspected of smuggling?
- **iv** **Projection**: How can we help the operator in *projecting* a current situation into the future? To be able to make decisions, the operator must be able to project how the situation is going to evolve into the future. For example, is the vessel suspected of piracy going to attack a nearby merchant vessel?

We discuss these parts in more detail in the overview section.

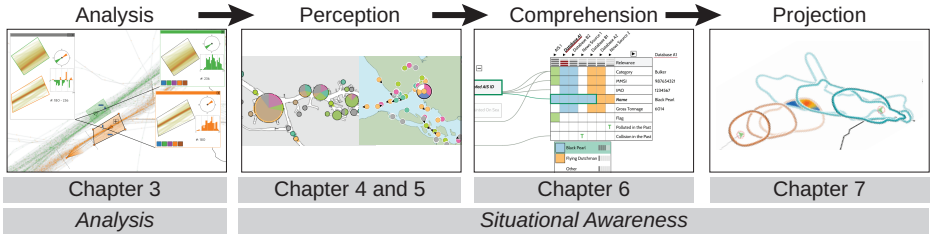


Figure 1.3: An overview of the contents of this thesis and their relationships.

1.6 Approach

Each part of the research question is answered by designing visualization techniques and iteratively developing these in interactive prototypes based on our own experience and input from domain experts and potential users. These prototypes are then validated using domain experts and potential users, mostly from the maritime domain, in both formal and informal studies. Also, we aim to show that our techniques are generic enough to be applied in other domains as well.

All work in this thesis has been fully implemented from scratch in stand-alone applications by the author, Roeland Scheepens, with the exception of the reasoning engine and its abstraction as discussed in Chapter 6, which has been implemented by Steffen Michels of the Radboud Universiteit, a partner in the Metis project, as part of his research. All applications have been implemented in C++ using Qt for a part of the user interface. The visualizations have been implemented using native OpenGL and Cg shader code. To increase efficiency, the algorithms have been either parallelized on the Central Processing Unit (CPU) or implemented on the Graphics Processing Unit (GPU) using shaders. Unfortunately, the implemented work cannot be made publicly available.

1.7 Overview

In the remainder of this thesis we aim to answer the research question stated in Section 1.5. We begin by presenting some background for the work in this thesis in Chapter 2. Our main contributions, the visualization and interaction techniques we developed, are described in Chapters 3 through 7–see Figure 1.3.

In Chapter 3 we discuss a way to gain more domain knowledge by analyzing movement patterns of moving objects [170]. We do this by visualizing traffic flows and provide interaction tools to support their exploration. The user can intuitively select and filter traffic flows from an overview visualization. The dynamic behaviors of selected flows may then be shown in annotation windows in which they can be interactively explored and compared. We demonstrate the effectiveness of our method through a number of use cases in the air traffic domain and the maritime safety and security domain.

In Chapter 4 we show how to support perception in situational awareness by reducing the cognitive overload and visual overlap caused by dense populations in the maritime picture [175]. In moving object visualization in general, objects and their attributes are commonly represented by glyphs on a geographic map. In areas on the map densely populated by these objects, visual clutter and occlusion of glyphs occur. We propose a method to solve this problem and through a user study we find that, for a set of representative tasks, our method does not perform significantly worse than competitive visualizations with respect to correctness. Furthermore, it performs significantly better for density comparison tasks in high density data sets. We also find that the participants of the user study have a preference for our method.

In Chapter 5 we show how to support perception in situational awareness using a specialized multi-variate glyph designed in cooperation with domain experts [174]. The glyph is designed to help the operator to perceive quickly which vessels are relevant and require more attention. Starting from maritime domain requirements, a number of design parameters and feasible choices are determined. We determine the best choices by showing the glyphs in a sandbox environment and allowing the domain experts to vary the parameters.

In Chapter 6 we show how to support comprehension in situational awareness by visually explaining conclusions of a reasoning engine that raises an alarm if a certain situation is reached [172]. We offer an improvement over current visualization methods, where only a list of evidence is shown. Two groups of domain and operational experts are used to evaluate our system by testing a number of use cases in the maritime domain based on real data. Experts can easily follow the reasoning structure, and can quickly understand and find complicated patterns and relate them to real-world situations in an evidence matrix.

In Chapter 7 we show how to support projection into the future of a current situation using a visualization method for the interactive exploration of predicted positions of moving objects, in particular, ocean-faring vessels [173]. Users, investigating and exploring the possible development of a situation, can see where a vessel will be in the near future according to a given prediction model. Through a number of real-world use cases and a discussion with users, we show our methods can be used in monitoring traffic for collision avoidance, and detecting illegal activities, like piracy or smuggling. By applying our methods to pedestrian movements, we show that our methods can also be applied to a different domain.

Finally, we discuss the presented visualization techniques, reflect on our research questions, and discuss some open research questions and possible future work in Chapter 8.

1.8 Related Publications

The main content chapters of this thesis are based on the publications listed in this section. We divide the publications into three categories: Primary, secondary and tertiary papers. The primary papers are all first-authored by Roeland Scheepens

and each serve as core material for a specific chapter. The secondary papers are papers co-authored by Roeland Scheepens, cover a wider subject within the Metis project, and serve as context for this thesis. The tertiary papers are papers first-authored by Roeland Scheepens in collaboration with Niels Willems [215] in the Poseidon project, the predecessor of the Metis project, and provide additional context and background for the work discussed in this thesis.

1.8.1 Primary Publications

- R. Scheepens, C. Hurter, H. van de Wetering, and J.J. van Wijk. **Visualization, Selection, and Analysis of Traffic Flows**. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):[To appear], 2015. [170].
  This publication serves as core material for Chapter 3.

- R. Scheepens, H. van de Wetering, and J.J. van Wijk. **Non-overlapping aggregated multivariate glyphs for moving objects**. *In Proceedings of the 7th IEEE Pacific Visualization Symposium (PacificVis 2014)*, pages 17–24, March 2014. [175].
  This publication serves as core material for Chapter 4.

- R. Scheepens, H. van de Wetering, and J.J. van Wijk. **Design of Glyphs for Uncertain Maritime Data [Poster]**. *Eurographics Conference on Visualization (EuroVis)*, June 2014. [174].
  This publication serves as core material for Chapter 5.

- R. Scheepens, H. van de Wetering, and J.J. van Wijk. **Rationale visualization for decision support [Poster] (honorable mention)**. *In IEEE Information Visualization Conference*, Paris, France, 2014. [171].
  This publication serves as core material for Chapter 6.

- R. Scheepens, S. Michels, H. van de Wetering, and J.J. van Wijk. **Rationale visualization for safety and security**. *Computer Graphics Forum*, 34(3):191–200, 2015. [172].
  This publication serves as core material for Chapter 6.

- R. Scheepens, H. van de Wetering, and J.J. van Wijk. **Contour based visualization of vessel movement predictions**. *International Journal of Geographical Information Science*, 28(5):891–909, 2014. [173].
  This publication serves as core material for Chapter 7.

1.8.2 Secondary Publications

- B. Huijbrechts, M. Velikova, R. Scheepens, and S. Michels. **Metis: An integrated reference architecture for addressing uncertainty in decision-support systems**. *Procedia Computer Science*, 44:476–485, 2015. [98].

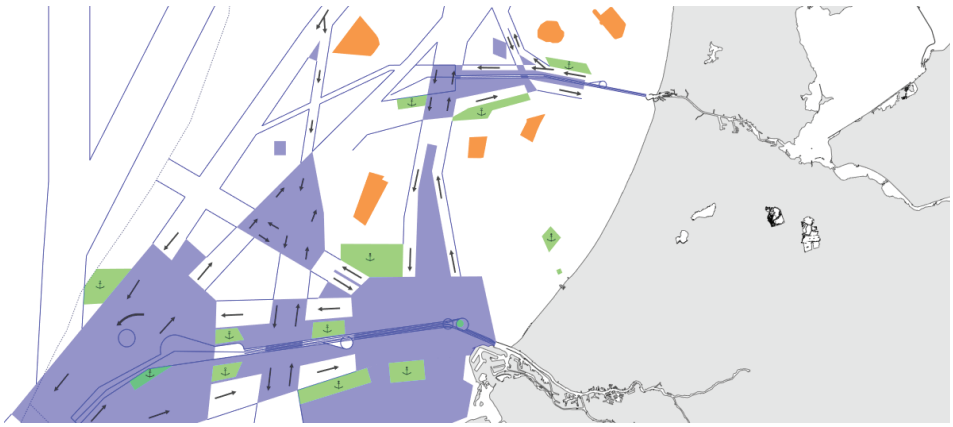
1

1.8.3 Tertiary Publications

- R. Scheepens, N. Willems, H van de Wetering, and J.J. van Wijk. **Interactive visualization of multivariate trajectory data with density maps**. *In Proceedings of the 4th IEEE Pacific Visualization Symposium (PacificVis 2011)*, pages 147–154, 2011. [178].
- R. Scheepens, N. Willems, H van de Wetering, and J.J. van Wijk. **Composite density maps for multivariate trajectories**. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2518-–2527, dec. 2011. [176].
- R. Scheepens, N. Willems, H van de Wetering, and J.J. van Wijk. **Interactive density maps for moving objects**. *IEEE Computer Graphics and Applications*, 32(1):56-–66, jan.-feb. 2012. [179].
- N. Willems, R. Scheepens, H. van de Wetering, and J.J. van Wijk. **Visualization of vessel traffic**. *Situation Awareness with Systems of Systems*, pages 73–87. Springer, 2013. [216].

2

Background



In this chapter we give background information on the concepts introduced in Chapter 1. We start by describing the Maritime domain in more detail where we give an impression of what goes on in one of the busiest maritime areas in the world, the North-Sea and what information can typically be gathered from traffic in such an area. We next discuss a number of threats and anomalous or suspicious behaviour an operator may be looking for. Following this, we give more background on situational awareness and especially the human factors involved in it. Finally, we give more background on visualization where we mainly focus on perception.

2.1 Maritime Domain

In 2014 the world merchant fleet consisted of approximately 88 000 ships in total with a total of 1.192 billion deadweight tons (DWT)[1] [198]. In 2011, 8.74 billion tons of goods were transported by sea [198]. It is expected that the number of merchant vessels will roughly stay the same, but that their size will increase [192].

In our use cases we mainly look at the Dutch Exclusive Economic Zone (EEZ) –see Figure 2.1, an area of approximately 58 000 km^2 (not including the Dutch territorial waters of 7000 km^2), which is a part of the North-Sea and one of the busiest maritime areas in the world [192]. The port of Rotterdam, one of the largest ports in the world, is also located in this area. We discuss the North-Sea and specifically the Dutch EEZ in some more detail [192]. The North-Sea is used for a large number of human activities including fishing, energy production by offshore wind farms, shipping (both international routes and routes to and from ports), recreation, sand extraction, petroleum and gas extraction at 162 platforms, defense exercises in 5 zones with a total of 4200 km^2, and many more. Additionally, the North-Sea contains infrastructures such as 4500 km of pipelines and 3300 km of cables, habitats for marine mammals, breeding grounds for fish, and underwater cultural heritage such as sunken ships and remnants of ancient settlements [192]. To maintain a healthy marine ecosystem and to maintain a sustainable food supply through fishing, vulnerable areas are protected and designated as nature conservation areas of 11 374 km^2 in total, where fishing is restricted or prohibited. Generally, fishing is allowed everywhere, except in nature conservation areas and areas closed for energy production.

All these activities are increasing and some of them are at odds with one another, such as fishing versus conservation. To allow all these activities to take place safely side by side, regulation and enforcement of these regulations is required. To enhance safety and security at open sea, such as the Dutch EEZ, a number of international treaties are in effect in international waters, under the United Nations Convention on the Law of the Sea (UNCLOS) [199]. These are treaties such as the International Convention for the Prevention of Pollution From Ships (MARPOL) [107] and the International Convention for the Safety of Life at Sea (SOLAS) [110], which governs safety standards for ocean-faring vessels. The responsibility of enforcing these treaties lies with international initiatives [155], national navies, law enforcement agencies such as the coast guard [190, 191], and customs and border protection agencies. In the Dutch EEZ the coastguard is responsible for incident and disaster management and enforcement of legislation relating to the marine environment, traffic safety, and fishing [192, 191].

To increase navigational safety, Traffic Separation Schemes (TSS) are in effect in some areas to regulate traffic and avoid collisions [108]. This is done by separating traffic flows in opposite directions. Vessel traffic is separated into lanes using a *separation zone*, which vessels are not allowed to use. Other zones are

[1] DWT is a measure for the weight a ship can carry. This includes not only cargo, but also crew, passengers, fuel, and provisions.

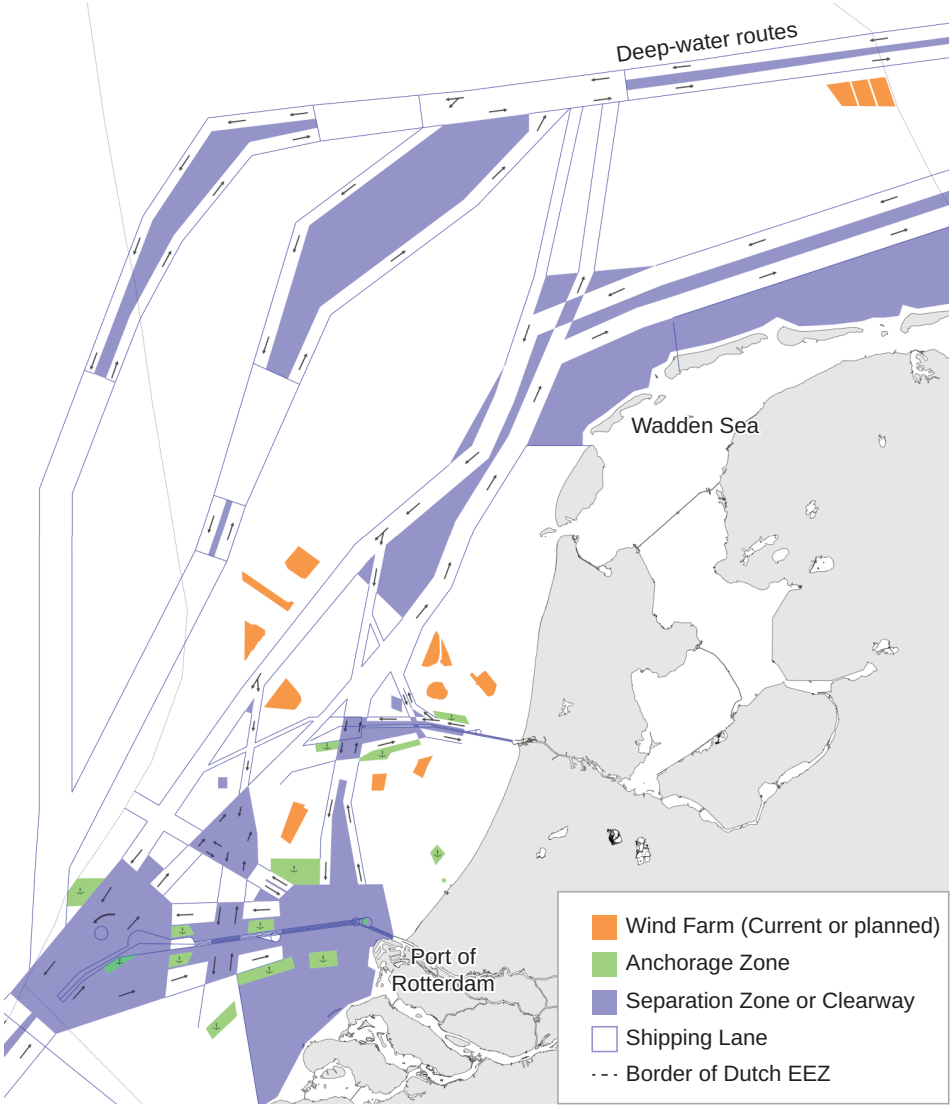


Figure 2.1: The Dutch EEZ, the TSS, and locations of wind farms. All data used to generate this image are freely available from the Dutch National Geo-Register (DGR) [154].

clearways, which may contain no obstacles, and *anchorage zones* where vessels may go at anchor. TSS are usually in effect in busy shipping areas, such as the North-Sea, around capes, or in confined areas. They are established by the International Maritime Organization (IMO) under the SOLAS convention [110]. In Figure 2.1 we show the TSS in effect in the Dutch EEZ as of August 2013. The shipping routes in the Dutch EEZ take up a total of 3600 km^2 [192]. Additional rules may apply. For example, tankers in the Dutch EEZ are required to follow the deep-water routes further out from the coast, to protect the *Wadden Sea*, a particularly sensitive area [192], from potential deliberate or undeliberate spills.

2.1.1 AIS

The Automatic Identification System (AIS) is a system by which vessels broadcast information about themselves such as their location, kinematic attributes, identity and voyage related information. It is intended to more easily avoid collisions and to enable littoral states to identify vessels operating near their coasts [190]. Under the SOLAS convention [110], it is mandatory for all ships of 300 gross tonnage and above on international voyages, all passenger ships, mobile offshore drilling units, and all international port facilities to use AIS [109, 111]. Cargo ships not on international voyages are also required to carry AIS if their gross tonnage is 500 or greater.

AIS is a distributed self-organizing system. Each AIS carrying vessel regularly broadcasts messages containing dynamic, voyage related or static attributes using a very high frequency (VHF) transmitter [190]. Vessels are required to have onboard sensors that record information such as GPS position, heading and rate of turn. Dynamic information such as vessel position, speed, heading, course, and rate of turn are broadcast every few seconds based on the vessel's speed and rate of turn. The dynamic information is broadcast every three minutes if the vessel is at anchor. Static and voyage related information such as the vessel's identifiers; IMO number and MMSI, name, ship and cargo type, vessel size, and voyage destination and estimated time of arrival are broadcast every six minutes. The static and voyage related information is entered by the operator of the vessel and the dynamic information is gathered by the vessel's own sensors.

AIS is, however, not reliable for vessel monitoring as not all vessels are required to carry AIS [24] and, under certain circumstances, vessels are allowed to turn off their AIS [110]. Additionally, the coverage of AIS is limited [190]. Even though there are offshore receivers available, mostly on mining platforms, there is no full coverage, especially in remote areas. Also, the usage of AIS is not properly enforced [24].

Trajectories

Even though AIS is mainly intended for collision avoidance, the emergence of AIS has greatly stimulated research in the maritime domain [92]. AIS data is increasingly being recorded for, *e.g.*, analysis and visualization of vessel movements [179, 176]. Figure 2.2 shows an overview of vessel traffic over two days in the Dutch EEZ visualized using a density map [179].

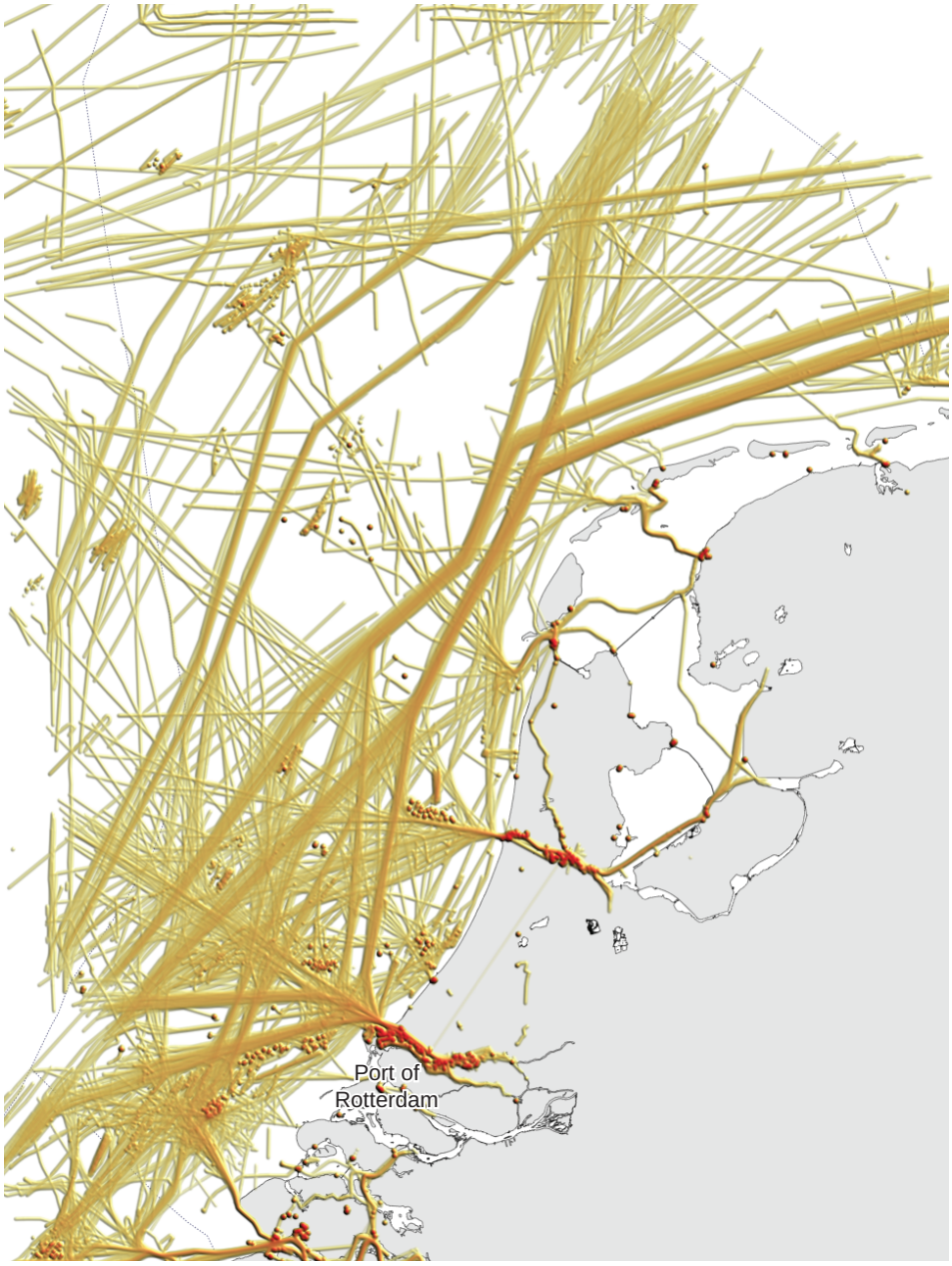


Figure 2.2: An overview of AIS trajectories over two days in the Dutch EEZ, visualized using a density map [179].

In this thesis, we describe a trajectory of a moving object o by a sequence of states $\boldsymbol{\alpha}^{\mathbf{o}}$. A state is a tuple containing constant attributes such as object id, type and area or size, and dynamic attributes such as time t, position $\mathbf{p}_o$, velocity v_o, and heading h_o. It may also contain additional attributes such as the altitude of an aircraft a_o, or the type of a vessel τ_o. In the case of vessels, this trajectory is derived from AIS.

2.1.2 Suspicious Behavior and Threats

What is considered suspicious behavior depends on numerous factors, such as location of the vessel or the type of vessel. For example, a fishing vessel loitering at sea is expected behavior, while a cargo vessel loitering at sea is considered suspicious. Such suspicious, or anomalous, behavior may evolve into a threat. In this section we discuss a number of these potential threats and suspicious behavior that may be indicative of these threats: smuggling, environmental threats, plunder, and piracy.

In general, a vessel in violation or engaged in criminal activity will attempt to hide both its activity and its identity using various strategies. Vessels may choose to turn off their AIS. This can, however, be easily detected using other sensors such as radar and the human eye. Therefore, a common strategy for vessels engaged in illegal activity or in violation is to *spoof* their AIS message by changing information in an attempt to hide their activity, their identity, or even their true location [24, 93, 98]. This is much harder to detect as it requires identity management and comparison with other sensors, sources, and previous readings. Other strategies may be employed to hide identity such as repainting the hull of the vessel at sea to make the vessel appear different to the human eye and cameras.

Smuggling

To evade the authorities, smugglers often use multiple vessels to smuggle their illegal goods. For instance, smaller vessels are often used to transfer goods from larger vessels to the shore to circumvent port authorities. This usually involves a rendezvous of two or more vessels at sea. A rendezvous is therefore considered suspicious and may be indicative of smuggling. A rendezvous may involve vessels stopping at the same location, vessels moving parallel for a period of time, a vessel stopping for a period of time shortly after another vessel has moved through the area, or two vessels moving towards each other and making a turn upon encounter [203, 60].

Not every rendezvous, however, is anomalous and requires attention from an operator. For example, a rendezvous between a tanker and a pilot vessel near a harbor is a common occurrence. In these cases a pilot boards the tankers from a pilot vessel to guide the large vessels into narrow harbor waters. To better determine whether a rendezvous requires attention, additional information may be needed, such as information about the owners, the crew and their relations, the history of the vessel, and relevant financial transactions [24].

In [93] and [98] a number of real smuggling cases are discussed.

Environmental Threats and Plunder

Vessels are banned from discharging any type of waste except food waste under the MARPOL [107] convention. Vessels, however, occasionally still illegally discharge waste. Apart from temporarily trying to hide their identity, *e.g.*, by changing AIS attributes, while they are discharging, violating vessels may move off course, discharge, and return to their original course. Other environmental hazards are posed by tankers that, instead of using the required route, move closer to ecologically sensitive areas, such as the Wadden Sea, to cut costs.

In certain areas, fishing is restricted or prohibited for certain species or methods [192]. Violations and illegal fishing, however, are fairly common [191]. Indicative of such violations may be vessels loitering in areas where fishing is restricted. On the other hand, loitering vessels that appear to behave as fishing vessels may also be involved in other illegal activity such as plundering national heritage sites, for instance sunken ships, for metal or valuable artifacts. If a loitering vessel is not, in reality, a fishing vessel, it may be indicative of plunder.

Piracy

In the last decade, maritime piracy has increased, especially in the Gulf of Aden and off the Somali coast [153, 149, 87]. Piracy is defined by the International Maritime Bureau (IMB) as: "*An act of boarding or attempting to board any ship with the intent to commit theft or any other crime and with the intent or capability to use force in the furtherance of that act*" [105]. Pirates typically attack merchant vessels with several small, fast boats for valuable cargo or to hold the crew and passengers for ransom. These small boats do not have a large action radius and are therefore supported by a larger *mother ship* that tracks and follows potential victims using GPS from AIS [149]. The pirate mother ships hide amongst the local commercial vessels [153].

Several strategies have been employed to combat piracy in the region: International warships patrolling the region; grouping vessels into convoys escorted by warships; (electronic) surveillance using patrolling aircraft; and using sensors such as satellites, radar, and on board transponders [153, 149]. Due to the size of the area, patrolling warships cannot cover the entire area. Because of this and the limited action radius of the pirate vessels, evasive action is considered the best course of action for a merchant vessel to escape pirates [87], however, this works only in the case of an early warning given by operators monitoring the area. Early warning signs for an imminent pirate attack may include a vessel that engages on an intercept course with a merchant vessel, or the presence of a pirate mother ship discovered through intelligence reports or suspicious behavior.

2.2 Situational Awareness

Originating from the Human Factors field, Situational Awareness (SA) [69] has gained increasing attention in the past decades [72]. Especially in safety and security domains, such as the Maritime Safety and Security Domain, where missions have become more fuzzy, there is increasing uncertainty in information and

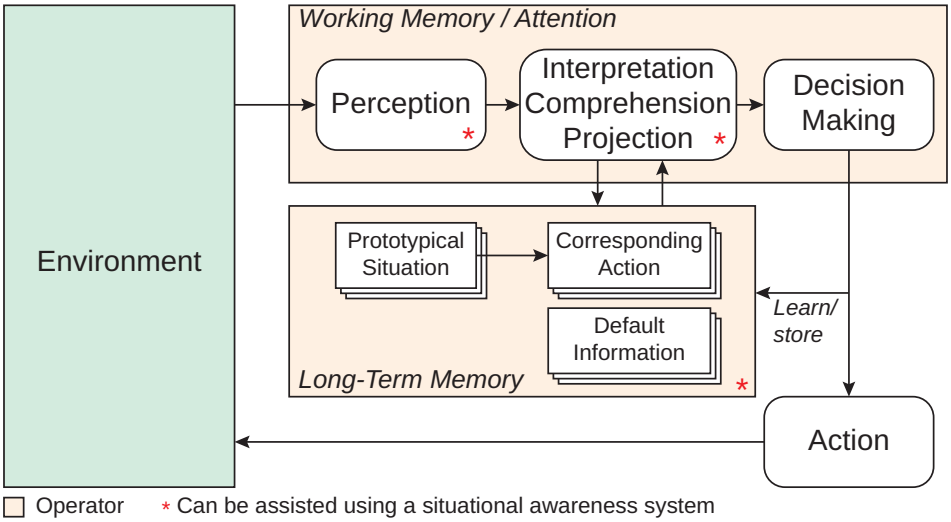


Figure 2.3: The mechanisms of situational awareness of an operator. Elements marked with a * can be (partly) augmented using an external situational awareness system. This figure is based on a figure by Endsley [69].

actors, and it is not clear exactly who the enemy is as opposed to in the traditional naval warfare domain [81]. The importance of a notion of situational awareness for crews of military aircraft goes back as far as World War I [69]. It is now considered important in varying fields such as air traffic control, large system operations, safety and security systems for, *e.g.*, police and fire fighters, tactical and strategic systems for the warfare domain, and the Maritime safety and security domain [69, 206, 81, 72]. Operator SA is a crucial construct for decision-making and performance in dynamic decision-making environments [69].

The psychology of perception plays a large role in SA, but it also depends in part on the psychology of attention–see Figure 2.3. According to Hoffman [95]: "Attention determines what is *in* immediate awareness and hence determines the contents in the stream of consciousness." SA can be seen as an extension to classical attention psychology, as the awareness not only of the dynamic surroundings, but also an awareness of the actions and intentions of others and the projection into the future of those actions [95]. It is described as a state of knowledge [69], not only of the current situation but also the past and the potential future of the environment and its relevant elements. The collective set of processes used to achieve, acquire, and maintain that state of knowledge (SA) is called *situation assessment* [70]. These processes take place in the human working memory, also known as short-term memory–see Figure 2.3. Attention also forms a major limiting factor for Situational Awareness as it is hard to perceive multiple elements in parallel. Where this attention is directed depends on how the information is presented and the operator's preconceptions, experience, and knowledge. It is therefore a challenge to maintain global situational awareness while focusing

attention on a set of elements for a subgoal. According to Endsley *et al.* [71], situational awareness systems should provide an overview for global SA across operator goals, while presenting the operator detailed information related to the immediate goals.

In [69] Endsley discusses how decision-makers compare the current situation to a set of prototypical situations in memory with a corresponding course of action using a form of pattern matching–see Figure 2.3. The process of acquiring and maintaining SA depends on a large number of factors: the innate abilities, training and experience of the decision maker; the objectives of a decision maker, *e.g.*, an operator on a warship patrolling for pirates has a different perspective of a situation than a coastguard operator monitoring for traffic violations; the system design, *i.e.*, how much of the needed information is presented and how is it presented; and the complexity of the situation, workload, and stress.

Within the SA framework, human operators work with default information, which is the information that can be assumed about an element if no specific information is available, *i.e.*, a kind of normal model based on operator experience and knowledge [69]–see Figure 2.3. For example, an operator perceives a suspect cargo vessel and, without knowing the capabilities of this specific vessel, can make certain assumptions about how fast this vessel can move. Also, the operators may have a certain level of confidence in information, which can influence the decisions the operator makes based on this information [145]. Both these constructs allow humans to achieve some degree of SA based on incomplete information.

When decision-makers work in teams, such as coastguard operators [191], an overall SA is achieved where each member has a specific set of SA elements they are responsible for [73]. Here it is not only necessary that each individual has the required SA for their specific tasks, but that relevant information is also successfully shared with team members that require it [70].

In most domains, instead of information being acquired directly by the human operator, information is sensed by an automated system that presents the information to the human operator. This, however, leads to potential loss of information at three stages [69]. First, these systems will acquire only information based on what the designer of the system understands is required and is technologically possible. Second, the system may not present all required information due to design and technological limitations. And lastly, the presented information may not be transferred to the human operator properly due to perceptual, attention, and memory constraints. The way in which the information is presented to the operator determines how much and how accurately information can be acquired and relates directly to the mental workload of the operator. In all stages of the design of situational awareness systems, these stages of potential loss of information should be taken into account. While reducing operator workload, automation may also reduce operator situational awareness if the operator is not kept sufficiently *in the loop* [37]. In Figure 2.3 the parts of the framework of SA that can be assisted using automated methods are marked.

2.2.1 Anomaly Detection

Since attention is a limiting factor to SA, it is important for the operator to direct attention to relevant elements in the environment. Elements that are anomalous, *i.e.*, elements that are or behave differently in some way, are relevant. Roy [167] gives a draft taxonomy of both kinematic and non-kinematic anomalies. An anomaly, however, is not necessarily a threat, nor is a threat necessarily an anomaly, but an anomaly may evolve into a threat. Therefore, these elements require attention to further investigate whether they pose any threat or not. Operators detect anomalies based on intuition and experience [81, 45], but can also be aided by visual aids [163] or automated methods [156], which are mainly based on kinematic attributes. Automated anomaly detection methods can be roughly divided into two categories: bottom-up, data driven approaches and top-down, rule driven approaches.

Data driven approaches can automatically find anomalies by learning what is normal [203]. Examples of such approaches can be found in [159, 156, 23, 157]. The major advantage of this approach is that it can continuously adapt to changing and evolving threats in the domain. It does, however, require data to be available and a training period to learn what is normal.

Rule-based approaches are based on existing knowledge about suspicious behaviour and apply rules to find anomalies [144, 203]. Examples of such approaches can be found in [60, 167, 153]. These approaches can be applied immediately without any learning period or available data and can find anomalies in a more reliable manner [203]. The disadvantage is that the rules require updating as illegal activity changes and evolves over time. Here, the major challenge is to extract and capture expert knowledge from operators and domain experts to create these rules [144].

2.3 Visualization

Visualization is not just about generating aesthetically pleasing images. It is an interactive, computer aided process to acquire meaningful insights from data using visual representations [49, 35, 213]. As such, visualization is not only used to communicate an idea, but it is also used as a tool to discover ideas [35]. In this thesis we mainly deal with exploratory visualization as described above, where interaction plays a vital role to enable exploration and discovery [27]. Communicative visualization techniques such as infographics [34], however, are purely used to communicate an idea.

Famous early examples of visualizations are Charles Minard's map of Napoleon's disastrous 1812 Russian campaign (see Figure 2.4) and Florence Nightingale's early use of the polar chart, which resembles a pie chart, to visualize number and causes of death during the Crimean War [197]. The rise of computer graphics and the emergence of powerful consumer hardware in the recent decades has given the field of visualization research a significant boost and has enabled

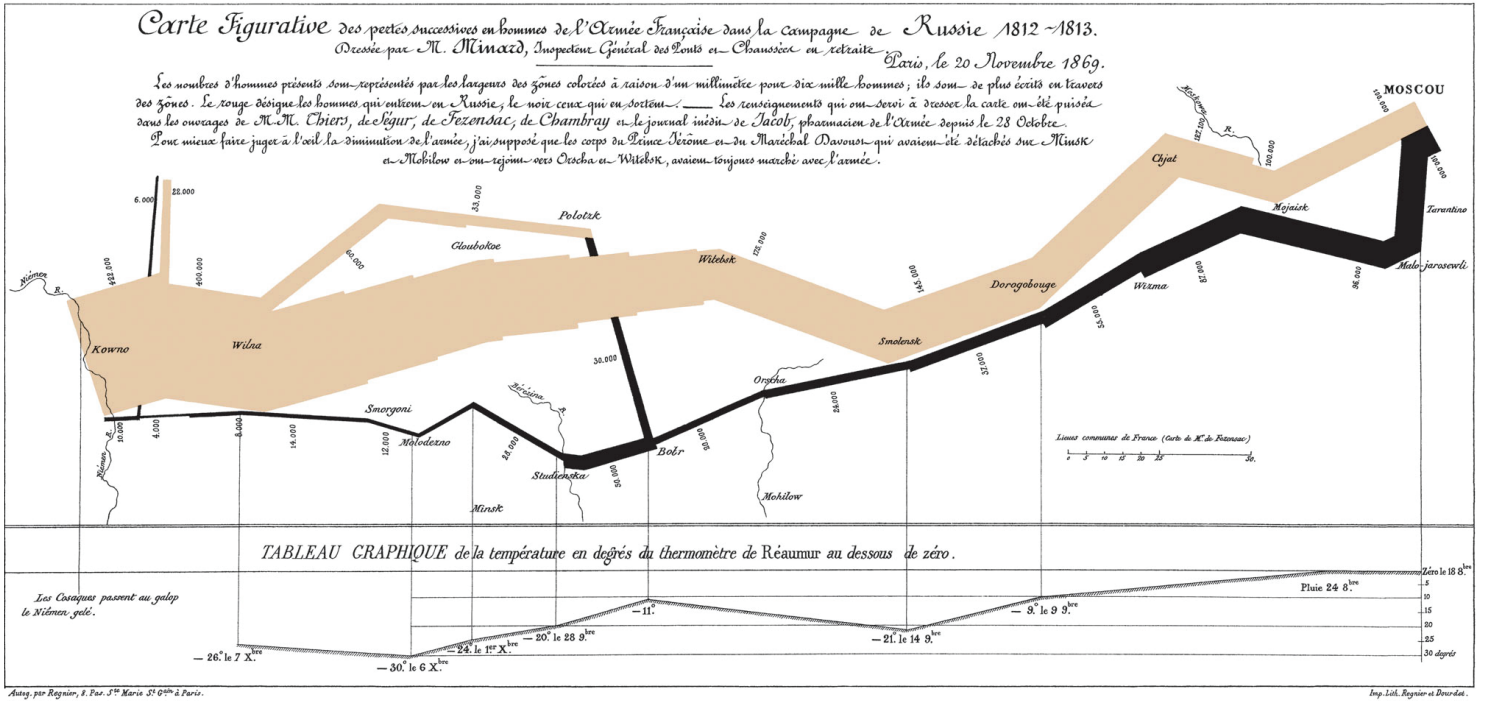


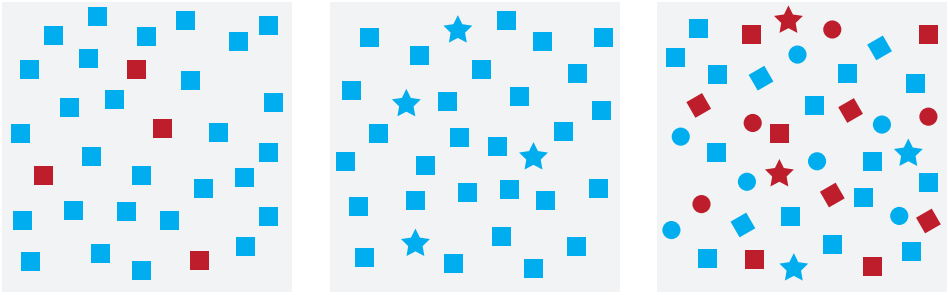
Figure 2.4: Charles Minard's figurative map of the successive losses of men of the French Army in the Russian campaign 1812-1813. Drawn up by M. Minard, Inspector General of Bridges and Roads in retirement in Paris, November 20, 1869. The number of men is represented by the width of the colored ribbons. The brown ribbons designate the men who enter into Russia, the black those who leave it. The spatial scale is drawn in *lieues communes de France* (common French league) which is approximately 4444 m. The lower portion of the graph shows the temperature on the army's return from Russia in degrees below freezing on the Réaumur scale. 1°Ré is approximately 1.25 °C.

Figure 2.5: A table of visual parameters. We show whether they are associative or selective, and whether they are poor, marginal, or good for representing nominal, ordinal, or quantitative data [166].

interactive visualizations.

The human visual system is very powerful in recognizing visual patterns such as clusters, anomalies, and trends, and can be leveraged to transfer information effectively and efficiently [205]. Visualization also enables the perception of emergent properties that were not expected and can even expose problems in the data such as errors in data acquisition or processing [213]. The use of external cognitive artifacts such as visual aids, significantly amplifies cognitive performance. Additionally, the way in which information is presented can determine how a decision is made or a problem is solved [69, 211]. It is therefore very important to consider human perception and its limitations while designing visualizations [211]. Visual perception and cognition take place in the human working memory.

According to Bertin [19], the following visual parameters can be identified: position, size, shape, color value, color hue, orientation, and texture–see Figure 2.5. Other visual parameters can be identified, such as [129, 128]: satu-



a. Find all red squares b. Find all stars c. Find all rotated red squares

Figure 2.6: In a and b only one visual parameter is varied (color and shape, respectively) and the visual elements can be found pre-attentively. In c three visual parameters are varied (color, shape, and orientation) and therefore finding the visual elements requires attentive search.

ration, arrangement, crispness, resolution, and transparency. The effectiveness and efficiency of these visual parameters depends on whether the input data is nominal, ordinal, or quantitative. A visual parameter is said to be selective if a change in this visual parameter alone in an element makes it easier to select the element from all other elements [38]. A visual parameter is said to be associative if a change in this visual variable is enough to perceive elements as a group [38]. In Figure 2.5 we show, for some of these variables, which visual variables are suited for which of the above classes of analysis tasks.

Due to limitations in human working memory and attention, not all visual stimuli can be perceived equally. The human visual system can process visual stimuli in two distinct stages [195]. Certain simple visual parameters such as color, size, shape, orientation, movement, spatial position, and texture can be processed pre-attentively in parallel across the visual field– See Figure 2.6ab. This processing typically takes well under a second and can be exploited in the design of visualizations. By encoding values using these simple visual parameters, a user of the visualization can process information quickly and in parallel. In the second stage, more complex objects with combinations of visual parameters can be identified. The second stage, however, requires focussed attention and visual fixation and as such can only be done serially. Therefore, this stage takes significantly more time–see Figure 2.6c.

There are more limitations of human perception and cognition that may influence the way a visualization is perceived. The human visual system does not store visual details across views to form a stable representation of a scene [183]. In fact, relatively little visual information is preserved. This leads to people missing visual changes even as they are happening, especially if they are not sufficiently salient or occur outside of focused attention. This phenomenon is known as *change blindness*. It needs to be taken into account in visualizations where the information displayed may change over time such as in a maritime operational picture. Also, the perception of individual colors may vary depending on the surrounding color and color contrasts [141]. This means not only selecting the

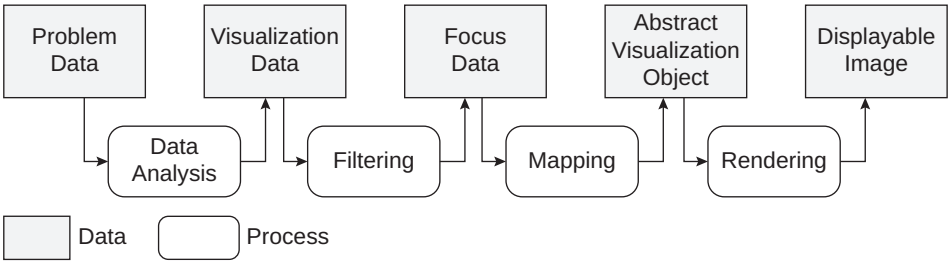


Figure 2.7: The visualization pipeline [56].

right colors is important in visualization design [214], but also effects colors have upon each other and how well colors can be perceived given some pre-existent background.

Generally, the visualization process follows the visualization pipeline [86, 56, 211] as shown in Figure 2.7. The problem data is first analyzed to form a model of the underlying data suitable to be visualized. This is usually a preprocessing step that does not directly involve the user. The user can then interactively filter the data to focus on data of interest relating to the user's task. Typically, a user will apply multiple such filters in any one session. This focus data is a subset of the visualization data and is mapped to abstract visualization objects that are then rendered onto the screen. The mapping is generally defined by the visualization designer. Shneiderman [181] argues, according to his visual information-seeking mantra: "*Overview first, zoom and filter, then details-on-demand*".

Visualization techniques can generally be categorized into two major fields [194]: *Scientific visualization* and *Information Visualization*. Scientific visualization deals with data that describe some physical process or object that has an inherent spatial or spatio-temporal component [211]. Examples are visualization of flows of gasses or liquids [135], medical imaging [28], and volume visualizations [115]. Information Visualization deals with abstract and non-spatial data [194]. Examples are visualizations of statistical data [46], document and text exploration [26, 158], and multivariate and dynamic networks [200, 15].

We also distinguish the field of *geographic visualization*. These are visualization techniques for geospatial analysis that deal with data with a geographic component. These techniques are usually applied in a Geographic Information System (GIS) and displayed on a map. Examples are visualizations of large amounts of trajectories using density maps (see Figure 2.2) or other types of aggregation methods [179, 215, 176, 10, 3], or visualizations of movement patterns using flow maps [30, 83], and many more [58, 4]. The visualizations presented within this thesis fall within the areas of information visualization and geographic visualization.

2.3.1 Glyphs

The use of glyphs is a visualization technique that is often applied to visualize multivariate attributes of multiple data records. They are placed in display space on a position defined by the contents of the data record. In geographic visualizations the position of a glyph on a map is typically predetermined by the geographic location associated with the data record. In information visualization in general, the position may depend on some mapping from non-spatial attributes to the display space.

As Borgo *et al.* [25] define it, a *glyph* is a small and independent visual object that encodes attributes of a multivariate data record. Typically, the multivariate attributes of the data record are mapped to the visual parameters of the glyphs such as its position, shape, size, color, texture, orientation, aspect ratio or curvature [210], of which color [42] and size [127] are considered the most dominant and attract the most attention from the user in a pre-attentive search. Visual elements attracting attention in a pre-attentive search is referred to as the *pop-out effect* [91]. Visual parameters of glyphs are assumed to have the following order with respect to degree of the pop-out effect [210, 91, 25, 131, 19, 82]: color > size > shape > orientation > texture.

Additionally, a number of dynamic visual parameters can be identified to describe animation or changes of visual representations over time. These parameters are [128]: moment, duration, frequency, order, rate of change, and synchronisation. Using dynamic visual parameters has been shown to be beneficial for depicting dynamic phenomena [116, 118] or dynamic attributes in glyphs.

Glyphs are often used to convey multiple attributes via multiple visual parameters. The composition of these visual parameters may affect how well individual visual parameters can be perceived–see Figure 2.6. Maguire *et al.* [131] propose a systematic process for glyph design using perceptual guidelines. The ordering of discriminative capacity of the visual parameters discussed above is linked to a conceptual hierarchy in the data. As a case study it is applied to visualize workflows of biological experiments. This work illustrates the need to consider visual perception in successful glyph design [210].

2.3.2 Visualization and Situational Awareness

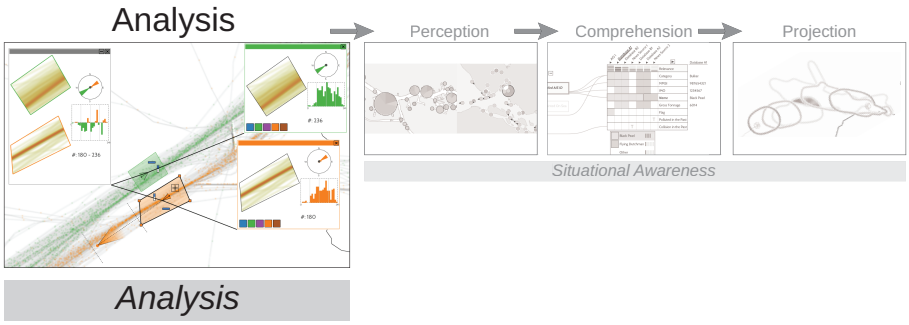
Due to the processing power of the human visual system and its unique ability in recognizing visual patterns, visualization is a powerful tool to aid an operator in acquiring and maintaining situational awareness.

Riveiro *et al.* [163, 161, 162] suggest to use visualization to enable the operator to detect and find anomalous behaviour. Anomalous behaviour is hard to define and requires human expert knowledge to detect. Therefore, the authors propose to interactively visualize expert rules and normal behavioral models derived from the data. In a similar sense, Scheepens *et al.* [178, 179, 177] visualize density maps of normal movement behavior. When current movements are visualized over this context, anomalous movement behavior in spatial locations and kinematic attributes can be visually detected.

2

3

Visualization, Selection, and Analysis of Traffic Flows



How can we provide tools to analyze and summarize patterns, enabling domain experts to find critical areas and to verify what standard or anomalous behavior is?

Automated methods for situational awareness, especially in the comprehension and projection level, often require expert rules, domain knowledge, or normal models supplied by domain experts. In this chapter we present how to aid a domain analyst in gaining insight into traffic flows of vessels or aircraft using interactive visualization. We show an overview of the traffic using a density map. The directions of traffic flows are visualized using an animated particle system on top of the density map. The user can extract traffic flows using a novel selection widget that allows for the intuitive selection of an area, and filtering on a range of directions and any additional attributes. Using simple, visual set expressions, the user can construct more complicated selections. The dynamic behaviors of selected flows may then be shown in annotation windows in which they can be interactively explored and compared.

The contents of this chapter have in part previously appeared in [170].

A video on the contents of this chapter can be found at: https://youtu.be/tfRV2bhmQsY.

3.1 Introduction

Moving objects such as cars, vessels, aircraft, or pedestrians do not move at random, but collectively form patterns. These collective patterns, or dynamic collective behavior [9], may be formed in two ways [54]: By groups that share some functional relationship, such as groups of animals travelling together, or by cohorts, which have some other factor in common, such as aircraft that have the same destination, or vessels that use the same shipping lane. In this chapter, we are interested in analyzing the latter patterns, which we call *traffic flows*. A traffic flow is represented by a set of trajectories. This gives rise to a number of challenges: How to visualize an overview of all trajectories such that a user can easily find traffic flows of interest; how to select these traffic flows; how to analyze their dynamics, *i.e.*, their behaviour over time; and how to compare multiple flows.

Analyzing trajectories of moving objects is of great interest to extract temporal patterns and to understand past events. Air traffic management deals with moving objects, namely aircraft, which follow flight routes through sectors (volumes) that subdivide the airspace under its control. For improved flight safety and flight route optimization, flight routes and sectors may evolve over time. Defining these optimizations properly is complex, and relies on a deep understanding of the dynamics of the air traffic. To facilitate this, we investigate how to increase the understanding and aid the analysis of flow dynamics using an interactive visual approach.

Recorded data may have poor semantics with only the location of the moving objects available at a given time, but additional information can be derived with simple algorithms, for example: speed, acceleration, direction. After this additional information processing, trajectories can be analyzed to extract relevant insights. Interactive visualization systems can leverage human visual analytical skills to get these insights. Since trajectories of moving objects are directed, the direction information is highly relevant and important to display. Most existing systems use arrows to show direction [3], leading to cluttered views that hinder data exploration. Other solutions investigated the use of color gradients [100]; this requires specific visual mappings, and subsequent color blending gives issues when gradients overlap. Some systems employ animated textures to show direction [22], requiring a minimum trajectory width, which is not suitable for large data sets with many entangled traffic flows. In this chapter, we propose the use of particles. Particle systems are an under exploited visualization technique that show the spatial extent and direction of a trajectory through moving particles and can provide strong cues with little clutter. Even if trajectories overlap, particle movements remain visible and indicate trajectory directions. Moreover, combined with particle density indicating trajectory densities, traffic flows become visible. As a drawback, a particle system requires animation with a high frame rate and interactive response time. The latter requirements are especially challenging for large moving object data sets.

Our overall contribution is to provide an integrated set of visualization and interaction techniques for investigating the dynamics of traffic flows. In our final

design, we use a density map of the trajectories combined with moving particles. We developed a new multi-dimensional selection widget for traffic flows. This interaction technique allows geographical selection of traffic flows, possibly filtered on direction and additional dimensions such as altitude or velocity. Such interaction helps to better highlight subsets of the traffic flows and to perform qualitative visual comparison in terms of direction and density. Furthermore, we provide additional visualizations to show and compare selections and their dynamics. The presented work illustrates how a particle system can be a great asset to explore a multi-dimensional data set of moving objects and how we developed visualization and interaction techniques to take full advantage of them. To support our claim, we present a number of case studies where our approach helps to retrieve insight from different kinds of moving objects data sets: vessels and aircraft.

In Section 3.2 we discuss related approaches. In Section 3.3 we explore our problem description by discussing input data, the target users and their requirements, and in Section 3.4, we discuss a set of typical tasks. Following this, we discuss our approach in Sections 3.5 (Visualization), 3.6 (Selection), and 3.7 (Analysis). We demonstrate and validate our approach according to a number of use cases using aircraft and vessel data in Section 3.8. And finally, we discuss our work, draw conclusions and discuss future work in the final section.

3.2 Related Work

Much prior work already exists regarding spatio-temporal data exploration [14]. In the following, we focus on work related to moving object visualization and exploration and outline our improvements.

Andrienko and Andrienko [2] investigate visualizing and comparing the variation in spatially distributed time-series data, which they call behavior. The distribution of these behaviors is displayed on a map using glyphs that show the time graph of the time-varying attribute for each area. This technique is, however, based on time-varying data that is tied to a predefined area, whereas we allow the user to dynamically select traffic flows.

FromDaDy [104] is a tool that allows the exploration of trajectory data through brushing, and picking and dropping the selections into juxtaposed views. We improve on this by introducing a novel selection widget that allows the user to intuitively select traffic flows more precisely based on direction, and potentially other attributes. Tominski *et al.* [193] visualize time varying attributes of trajectories by stacking trajectory bands that encode attribute values by color in 3D space.

Andrienko *et al.* [6] present a visual analytics approach to find places of interest in movement data by clustering related movement events, such as low speed events to find locations with traffic jams. The temporal patterns of the movement data in these places of interest can then be studied using spatio-temporal aggregation. In our case we do not use automated methods to find places of interest as

our users already know where to look and instead offer a visual method to select traffic flows of interest.

The exploration and visualization of Origin-Destination (OD) data is a growing research topic [220]. OD data is movement data for which only the origin and destination points are stored with additional attributes of the trips. Ferreira *et al.* [78] present a tool in which queries on OD data can be defined visually using graphical widgets. The query results can be visually explored and further refined. Guo *et al.* [84] automatically select and extract flow patterns from OD data, and visualize these patterns using arrow glyphs on a map. Though improvements have been proposed [30], resolving clutter in flow mapping remains a challenging problem. In OD exploration and visualization, however, the actual route and directions of the trajectories are not important, whereas in our approach we look at traffic flows where the origins and destinations of the individual trajectories are of less importance.

Krüger *et al.* [121] present *TrajectoryLenses*, a system that uses lenses to support visual, set-based filter expressions to select trajectories. It supports three types of lenses: origin, destination, and waypoint. These lenses can be grouped with set operations to create more complex queries. We follow a similar Focus + Context [44] technique where we visualize our selections and their dynamics within the context of all trajectories. Our selections can also be combined using set operations. Our selection widget, however, is more flexible. The selection area is user-defined and the selection can be filtered on multiple attributes.

Selection of lines by their direction or angle is also relevant in other domains. For example, Hauser *et al.* [90] introduce *angular brushing* in parallel coordinate plots to select records by their slope, or the relation between axes.

Van den Elzen and van Wijk [200] present a method to explore both the structure and the attributes of multivariate networks by creating selections of interest and combining a visualization of the selections in the network with a high level, infographic-style overview, showing the structure and multivariate attributes of the selections. We follow a similar approach, but we select traffic flows rather than locations. In our method, we can also generate high level, infographic-style visualizations based on user-made selections.

Muigg *et al.* [143] present a method to address clutter and overdraw in dense line plots. First a tensor field is generated with a distribution of line orientations for each pixel. Line orientations are then visualized by applying anisotropic diffusion to a noise texture. We get a similar effect using our method, however, animated particles perform better in areas where multiple directions overlap.

Willems *et al.* [217] visualize large numbers of moving object trajectories using density maps. Scheepens *et al.* [179] extend this by using multivariate filters. Density maps of subsets of the data can then be interactively generated. These density maps can be aggregated or visually composed using a variety of operators. An overview can be given of how traffic evolves over time, however, it requires much interaction, and traffic flows cannot be selected or investigated seperately. Neither can users see the direction of the traffic flows. In a further extension to this work, Scheepens *et al.* [176] use a scripting language to define more complex, composite density maps. This method can be used to extract and

visualize dominant traffic flows in the data, but still cannot show direction. We also use density maps, but only to show an overview of the spatial distribution of the traffic flows.

Blaas *et al.* [22] use animated textures to display directionality in connected graphs to explore time series of an extended state-graph. Our work differs in many ways, but uses a similar visual technique to display flow direction.

Flow visualization is a large field of research in which many methods have been developed to visualize properties of flow fields, such as streamlines, glyphs, and textures [134]. However, such flows differ fundamentally from our data in that they are fields with (time-varying) properties tied to location, whereas our time-varying properties are tied to moving objects, whose trajectories may overlap.

3.3 Problem description

In this section we describe our problem area in more detail. We start by describing our input data and the type of movement data we focus on. Following this, we discuss our target users and their requirements.

3.3.1 Data

Data that can be explored using our approach consists of a large set of trajectories of moving objects. We focus on movement data of objects that are free to move around in space, but are constrained by some rules that encourage dynamic collective behaviour, such as aircraft that follow flight routes, or vessels that follow shipping lanes. We do not focus on moving objects that move around freely without such rules, such as animal movements, or players in a football match. Our approach may also work for objects that are not free to move around in space, but follow a predefined track, such as trains on a railway track, but the highly constrained nature of these movements allow for simpler selection methods, such as brushing.

3.3.2 Requirements

Our approach is intended for analysts that want to investigate traffic flows. In our design we are mainly inspired by air traffic analysts. These analysts perform traffic flow analysis and are mostly air traffic controllers with extensive knowledge of existing flight rules and the structure of the airspace. Traffic flow analysis is required to optimize flight routes and sectors in the airspace structure, but can also provide important input for efficiently assigning the required number of air traffic controllers during a day. Gaining deep understanding of flow locations and their dynamics, enables the extraction of even more information. To facilitate this, the user needs to see an overview of the flows and their departure/arrival

locations; in addition, multidimensional filtering to focus on a specific flow in space and time, information on the number of aircraft per hour in an area, and traffic flow comparison should all be available .

Based on discussions with air traffic analysts, we have formulated the following requirements for our approach:

R1 Visualization: The user should be enabled to see the direction of traffic flows while looking at the overview of all trajectories.

R2 Selection: The user should be enabled to dynamically, and visually select any, and multiple, traffic flows of interest using simple interaction techniques.

R3 Analysis: The user should be enabled to investigate and compare the dynamics of the selected traffic flows, *i.e.*, how the flows evolve over time.

3.4 Task Analysis

Initially, we tried to find and visualize potentially interesting areas using automated methods inspired by [6]. We automatically found areas with a high dispersion of trajectory directions by computing the entropy [99] in a fine grid of cells. All points with a sufficiently high entropy are considered interesting and are clustered. The resulting clusters are used to define a Voronoi space division and for each Voronoi cell, relevant information is visualized, such as the distribution of directions, and the traffic density over time. Our users, however, generally know which areas they want to investigate, and preferred to define the areas of interest themselves.

Using the requirements stated in Section 3.3.2, we now identify several user tasks and describe these tasks according to the typology of visualization tasks by Brehmer and Munzner [27]. To investigate the usage of space, such as flight routes for aircraft, or shipping lanes for vessels, the user needs to be enabled to investigate the dynamics of traffic flows, *i.e.*, how the traffic flows behave over time. We have identified the following tasks–see Figure 3.1.

Task 1: Selecting traffic flows.
A user must be enabled to visually *discover* traffic flows. The user does this by *searching* for these traffic flows and *identifying* them. Traffic flows and their locations may or may not be known beforehand. To support the discovery process, we *encode* the traffic flows and their flow directions using a density map and a particle system, and allow the user to *navigate* the map. The user may then *select* the traffic flow and *filter* the selection to *identify* the desired traffic flow.

Task 2: Exploring the dynamics.
In this *discovery process*, the user must be enabled to *explore* the dynamics of a traffic flow selected by task 1, and *summarize* the dynamic attributes of all trajectories in the traffic flow. We do this by *aggregation* and we *annotate* the

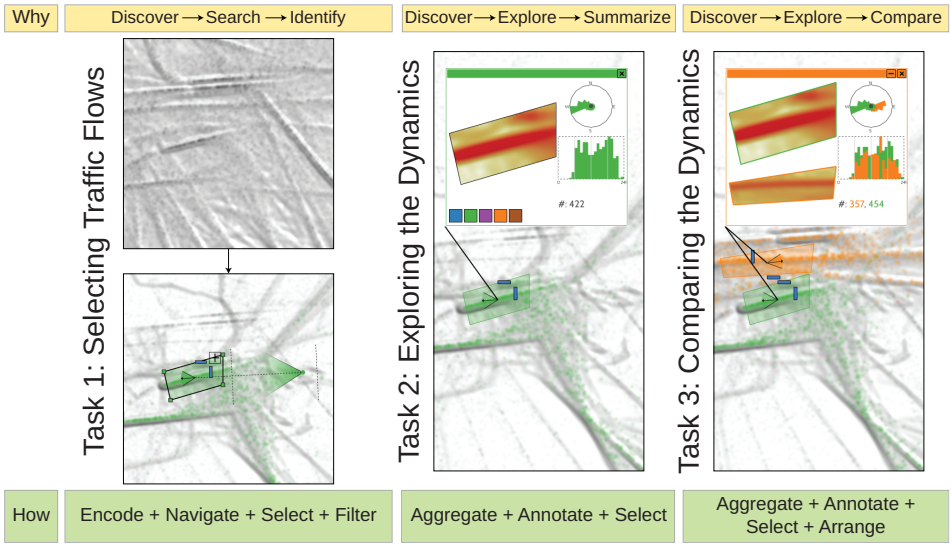


Figure 3.1: An overview of our tasks, based on the typology of visualization tasks of Brehmer and Munzner [27]. In task 1 the user can discover, search for, and identify traffic flows when navigating through a visualization of traffic flows encoded using a density map and a particle system. Traffic flows can be selected, and the selections can be filtered. In task 2 the user can discover, explore, and summarize the dynamics of a traffic flow selected by task 1. In task 3 the user can compare multiple traffic flows by arranging multiple windows produced by task 2 on top of each other.

dynamics of the selected traffic flow in a window that the user can arrange within the visualization space. Furthermore, the user can browse through the aggregated time bins in this window by *selecting* individual bins.

Task 3: Comparing the dynamics.

A user must also be enabled to *discover* the relationship between multiple traffic flows selected by multiple instances of task 2. The user must *explore* and *compare* the dynamics of these traffic flows. Similarly to task 2, we do this by *aggregating* and *annotating* the dynamics of the selected traffic flows in windows. To enable the *comparison* of selected traffic flows, the time bin selection in all windows is linked. Additionally, the user can *compare* multiple traffic flows by *arranging* their respective windows on top of each other, which automatically *aggregates* the visualization of both windows.

Task 4: Infographic-style visualizations.

In the final task, the user may want to create an infographic-style visualization, in which case a visualization is *produced* for a third party. For this task, any combination of the previous tasks may serve as input. While this task does not follow from the requirements, we find it interesting to explore, nonetheless.

3.4.1 Overview

The user can visualize, select, and analyze traffic flows using our approach as follows: we show the user an overview of the traffic flows in a user-defined time window by combining a density map [179] with animated particles. The density map shows the spatial overview of the trajectory data, while the particles show the direction of the traffic flows. The user can then interactively select traffic flows using a selection widget. We use the shading of the density map, and the color channel is used to visualize trajectory selections in the animated particles. To analyze the dynamics of the selected traffic flows and compare the dynamics of multiple selected traffic flows, the user can pop up windows to gain insight in the dynamics of the selected traffic flows.

3.5 Visualization: Particles

The user should be enabled to see the direction of traffic flows while looking at the overview of all trajectories (**R1**). We have chosen to show these traffic flows using an animated particle system. There are several different methods to visualize the direction in an aggregated visualization of a large number of trajectories, such as color maps [100], or glyphs such as arrows [3], which may be animated [33]. By encoding direction using color maps, blending issues arise and it is more difficult to use color to encode other attributes. By encoding direction using glyphs, a large amount of clutter is introduced, making the visualization harder to read, even if large traffic flows were bundled [100]. Instead, we have opted for animation, using a particle system, as we have found this does not interfere with other visual parameters, while it allows the user to clearly see and separate dominant traffic flows in the data–see the supplemental video. We have found that this does require a smooth animation at a high frame rate. We initially feared having animated particles would distract the user, but we have found this is not the case.

Having chosen for a particle system, we now derive the following new requirement from requirement **R1**: The particles should be visualized and animated such that the user can observe and distinguish traffic flows, even in areas where traffic flows intersect. Our particle system is produced as follows. First, we generate, for each trajectory $\boldsymbol{\alpha^0}(t)$, a set of particles geographically equally spaced along the trajectory. Their spacing needs to be such that the particle flows are dense enough to be able to observe traffic flows at all times, and sparse enough to be able to see particles moving along a traffic flow and to distinguish crossing traffic flows. The particles are then cycled over the trajectories with some speed defined in screen space. To avoid distracting regular patterns, some jitter in time and space may be required, however, we have found that in our data sets this was not needed. To render the particles themselves, we explored two possibilities–see Figure 3.2. First, we can render the particles as part of the density maps, *i.e.*, as part of the shading similar to [179]. Second, we can render the particles as

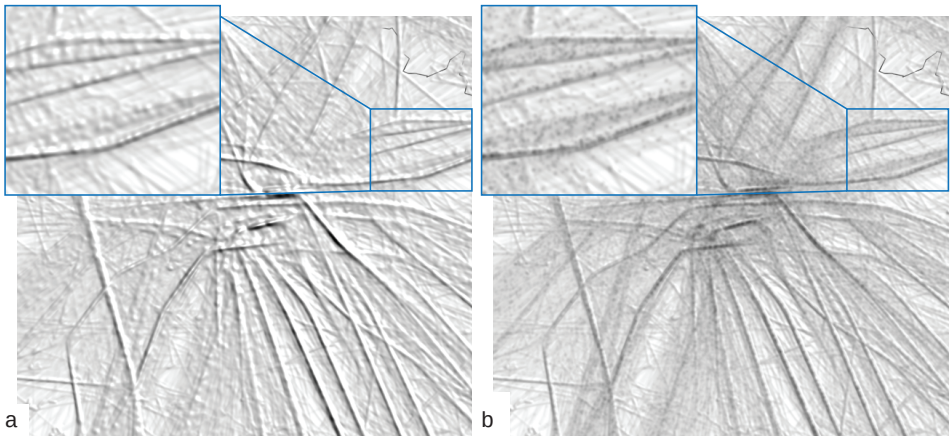


Figure 3.2: (a) The particles rendered as part of the shading of the density map, and (b) the particles rendered as Gaussian bells.

blended Gaussian bells with finite support. Rendering the particles in the shading makes it harder to use color to visualize additional attributes. Moreover, it appeared harder to interpret than the blended Guassian bells–see the supplemental video. Therefore, the particles are rendered as black or colored Gaussian bells, which are alpha blended using some weight p_α. Both the particle spacing and the weight are automatically estimated based on the number of trajectories per km^2 in the data set such that the particles reveal clear movement patterns without dominating the view. For example, in the aircraft data set shown in Figures 3.3 and 3.2, a particle spacing of approximately 40 km is used.

Because particles are generated for each trajectory, the particle system also naturally shows traffic density. We explicitly asked all domain experts we consulted what they thought the particles represent. Even without any further explanation, none of the domain experts confused a particle as representing a single aircraft or vessel. We believe this to be due to the density of the particles, *i.e.*, there are many more particles than the domain experts would expect there to be actual aircraft. For larger time windows, such as a year, this distinction may become less clear.

A slider can be used to change the particle weight p_α of non-selected particles, which also enables the user to hide non-selected particles by setting their weight to 0. To support zooming, the radius of the particles is defined in world space.

The velocity of the particles can be used to encode attributes of the moving objects they represent. We have experimented with attributes such as the actual velocity of the moving objects, and other attributes such as altitude. Encoding the actual velocity of the moving objects appears to distract the user and reduces the visual strength of traffic flows as too many particles move in different velocities in the same area. Mapping the altitude of aircraft to a limited number of discrete bins with particle velocities, using the principle that objects that are closer, *i.e.*, higher, to the camera, appear to move faster, seems to allow the user to separate

traffic flows by altitude. The effect, however, is not pre-attentive and makes separating traffic flows in general slightly harder–see the supplemental video. We therefore use a constant particle speed, which can be changed by the user if needed. Additionally, we tried applying the semantic depth of field technique [117], by allowing the user to *focus* on a specific attribute value or range of values, such as a range of aircraft altitudes. The particles with attribute values within the focus range were rendered as normal, while particles with attribute values outside the focus range were blurred. The effect, however, was too subtle and does not allow the user to properly focus on the traffic flows in range, while maintaining an overview of the general traffic flow–see the supplemental video.

Since moving objects can have multiple dynamic or static attributes, we have chosen to map these attributes to additional dimensions, such that the user can change axes (by rotating the view [104]), to explore these attributes in a similar way as ScatterDice [66]. Another way to map these additional attributes would be by coloring particles, however, we already use color to mark selections of traffic flows. By dragging the sliders shown in Figure 3.4a, the user can control the rotation over the attributes. Lastly, while in the rotated view, the user can also rotate over the spatial z-axis to view the particles from different angles.

3.6 Selection

The user should be enabled to dynamically and visually select any, and multiple, traffic flows of interest using simple interaction techniques (**R2**). We enable the user to create a selection S, a set of trajectories representing traffic flows of interest, using novel selection widgets. The user can draw a polygonal selection area by clicking at least three points on the screen. All trajectories that visit this area are now selected. We have chosen a polygonal area as opposed to a circular area or brushing a line to give the user the flexibility to define their own areas of interest, which may be later used to analyze traffic flows as described in Section 3.7. The selected area can be modified interactively, by dragging one of the vertices of the polygon or by moving the polygon in its entirety. The selection S can be further refined by filtering a range of directions, such that only trajectories within the filtered direction range in the selection area are selected. In Figure 3.3 such a selection S is shown, with our selection widget that allows the user to intuitively filter a range of directions. The user can change the direction filter by moving the anchor point, rotating it around the area to rotate the direction filter and dragging the anchor point closer to or further from the selection area to, respectively, widen or narrow the range of the filter–see the supplemental video. The selection is updated immediately after the user releases the mouse button. To reduce the amount of user interaction required, we divide the trajectories in the selected area into segments of equal length. We then apply Principal Component Analysis (PCA) to the set of displacement vectors of these segments to find the principle direction of the trajectories. Finally, we align the widget direction range to this principal direction. This allows us to define a sensible

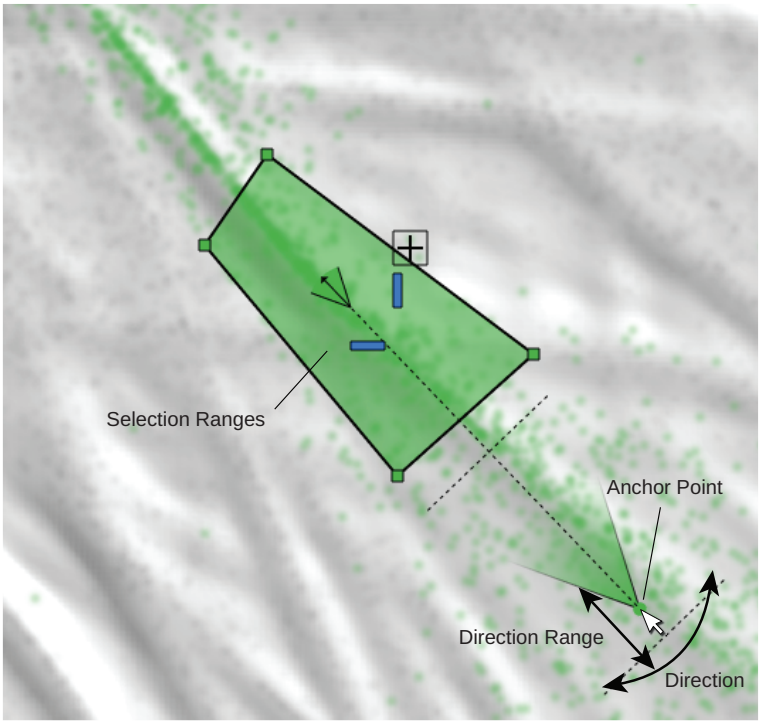


Figure 3.3: The user can create a selection area by clicking a polygonal area on the map. An initial direction range selection is estimated, which can be changed by the user in an intuitive way, by dragging the anchor point. The glyph in the center of the selection area shows the current selection ranges of the direction; and also the additional ranges using blue rectangles.

direction range filter immediately after the selection area has been created. This approach works well for normal distributions of directions, *i.e.*, in areas with one dominant traffic flow, but may give unsatisfactory results for multimodal distributions which can generally be found at intersections of two or more traffic flows of comparable density. We do not, however, consider this a problem as from the selection of an intersection area alone it is not obvious in which of the multiple intersecting traffic flows the user is interested. In this case the user can explore intersecting traffic flows and select the most interesting one by manipulating the selection widget.

Ranges from additional attributes can be filtered using range selectors to the left or at the bottom of the screen–see Figure 3.4b, or by rotating the view and changing the selection area there. These selection ranges are also visualized using small, blue rectangles in the selection glyph–see Figure 3.4. In Figure 3.10 we see an example of aircraft traffic rotated into the altitude view. In this view, the user can interact with the selection areas to change their altitude filters–see supplemental video.

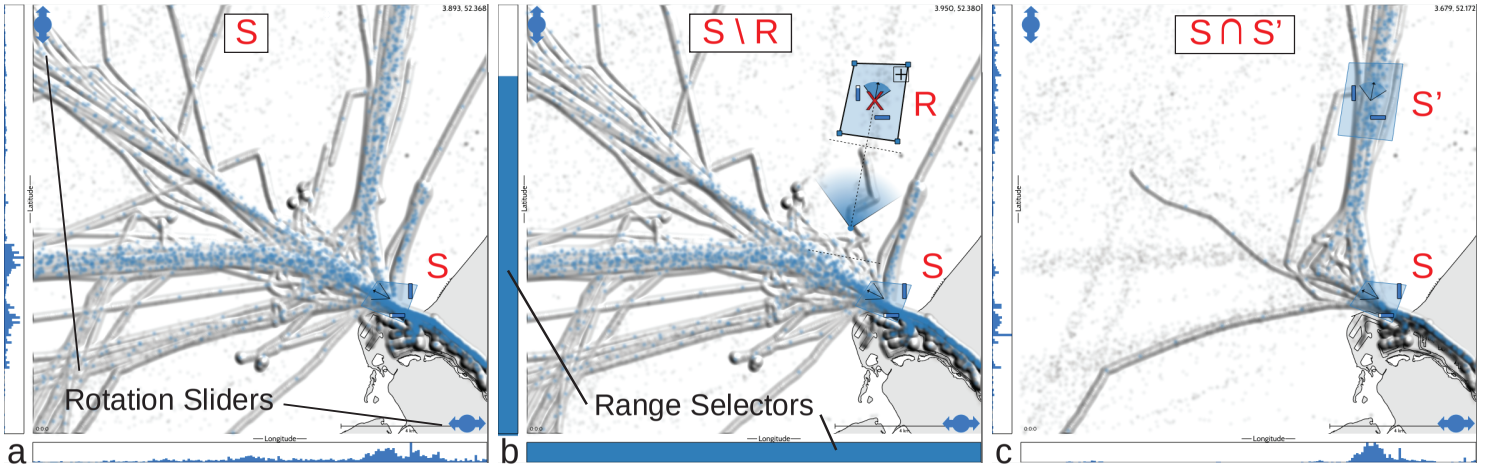


Figure 3.4: Selections of traffic flows of vessels near the harbor mouth of the port of *Rotterdam*. (a) Vessels leaving the harbor have been selected (S), (b) vessels leaving the harbor (S) except vessels heading north (R) have been selected, and (c) vessels leaving the harbor (S) that head north (S') have been selected.

To enable the user to create more complicated selections, we support a compound selection S, *i.e.*, a selection created using multiple selections S_i–see Figure 3.4 and the supplemental video. Additionally, we support *inverse* selections R_i to remove trajectories from the compound selection S. A compound selection S is then defined by

$$S = \bigoplus S_i \setminus \bigcup R_i \tag{3.1}$$

where $\oplus$ can be configured to be $\cap$ or $\cup$. This allows the user for instance to define selections where all trajectories going through a selected area S are selected except those that go through another selected area R (Figure 3.4b), or where all trajectories that go through selection area S *and* through selection area S' are selected (Figure 3.4c).

The user can also create multiple, independent compound selections. These selections are colored using a qualitative ColorBrewer color map [89]. Selected trajectories are visualized by coloring their particles using their associated selection color. To improve visibility, selected particles are rendered on top and with a higher particle weight p_α. The user can choose to only render the selected trajectories in the density map by selecting the "render only selected trajectories" option. Additionally, the user can choose to hide the selection widgets of all colors, or just of one color.

3.7 Traffic Flow Analysis

The user should be enabled to investigate the dynamics of the selected traffic flows, *i.e.*, how the traffic flows evolve over time (**R3**). Specifically, the usage of an area over time by objects moving through the selected areas, and the change of the directions of these objects over time should be presented. Also, the user should be enabled to compare the dynamics of multiple selections (**R3**).

We realized this by showing windows on demand for any selection area, that serve as annotations to the selected areas. These windows show the number of trajectories within the selection area over time using a histogram, a density map showing the trajectories within the selected area, and a polar area diagram showing the distribution of directions similar to a *wind rose* plot [184]–see Figure 3.5a. By hovering the mouse over bars of the histograms, the user can investigate specific time intervals. Both the density map and the direction diagrams show the density and direction, respectively, of the highlighted time interval–see Figure 3.5b and the supplemental video. To aid comparisons, all windows are linked, that is, the same time interval is highlighted in all windows.

The windows visualize the information of all tracks visible within the selection areas. The user can deselect individual selection colors by clicking the color buttons at the bottom of the window–see Figure 3.5c. Selection areas may also serve as selections of areas of interest, instead of selections of traffic flows, for instance, to investigate the dynamics of selected traffic flows in a specific area.

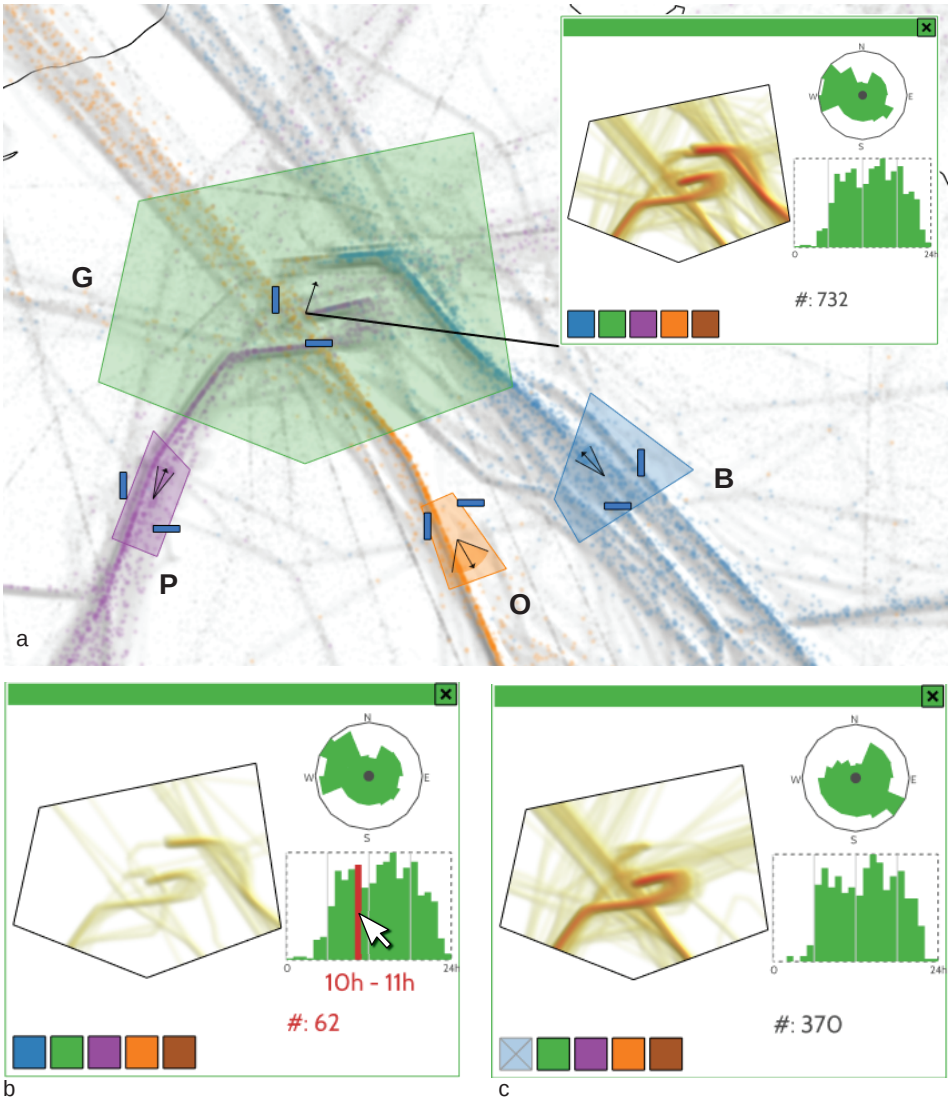


Figure 3.5: Three different sets of traffic flows have been selected: Purple (P), orange (O), and blue (B). Also, a green (G) area has been created that does not select any traffic flow, but represents an area of interest. (a) A window has been opened showing the dynamics of the traffic in the green area, *e.g.*, distribution of traffic over time, and the distribution of directions. (b) By hovering the mouse over the time histogram, different time instances can be explored. (c) The color buttons in the bottom of the window show which traffic flows are visualized in the window. Here, the blue traffic flow has been deselected.

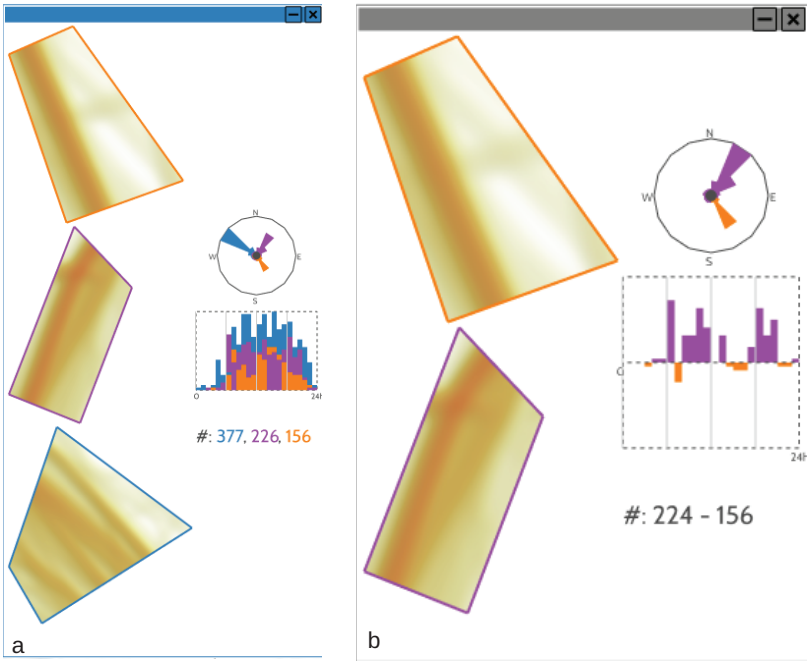


Figure 3.6: (a) The user can compare multiple (in this case, three) traffic flows by *stacking* multiple windows, which results in stacked visualizations. (b) Additionally, the user can investigate the difference between two traffic flows.

The green selection in Figure 3.5a is an example of such an area of interest selection, which does not select any traffic flows as can be seen by the direction range in its glyph, *i.e.*, the range is a line, which means the direction range is empty.

Traffic flows can be compared in multiple ways [113]. The user can juxtapose multiple windows, but the user can also superimpose the movable windows by dragging these on top of each other–see the supplemental video. The stack of floating windows is then displayed as a single window containing all density maps in the stack. Both the polar area diagrams and the histograms are stacked by, for each bin, depth sorting the values, from large (back) to small (front) similar to the braided graphs of Javed *et al.* [114]–see Figure 3.6a. We allow the user to choose between juxtaposition and superimposition because both approaches have their unique strengths [114]. While superimposition (space sharing) is better for comparing local maxima, juxtaposition is better for dispersed comparison.

Apart from stacking, the user can also investigate the difference between traffic flows by dragging a line between two selection areas while holding the shift key. A new floating window appears with density maps for the selected areas. In this window, the difference between the traffic flows is shown in both the direction diagrams and the histogram. In the direction diagrams we show, per direction bin, the absolute difference between the traffic flows by the color of the area using the color of the largest traffic flow. In the histograms we show,

per time bin, the difference between the two traffic flows where bars are drawn above the line if the bottom traffic flow is larger, and below the line if the top traffic flow is larger–see Figure 3.6b.

3.8 Evaluation & Use Cases

In this section we discuss expert feedback and a number of use cases using a set of aircraft trajectories over France of a single day, and a data set of vessel trajectories near the Dutch coast. The aircraft data set contains 17 841 flights with a total of 424 546 sample points. The vessel data set contains 16 421 vessels with a total of 420 335 sample points.

Since we used a user centered design process, we conducted several evaluation sessions with air traffic controllers. This section reports the latest evaluation in which two air traffic controllers, that have not been involved in the design process, used our system and gave feedback. The two air traffic controllers have 10 years of experience in the Paris area: *Roissy Charles de Gaulle* and *Orly*, the two biggest airports in France. During the evaluation, we first presented the goal of the tools with its available features. Then we asked them to freely use the software while thinking aloud to better understand their reasoning.

The two air traffic controllers followed the same sequence of investigations: they first tried to validate what they already knew and then, they tried to understand the rationale of a number of outliers. They both first investigated the traffic flows of the airport they were most familiar with. What they found out about these traffic flows was consistent with what they already knew apart from two outliers. They investigated these outliers: unexpected traffic at 5am at Orly and a peak at 8pm at Roissy–see Figure 3.7. At 5am, Orly airport is closed, but a number of aircraft appear to have landed anyway. After some investigation, these aircraft appear to have landed at *Villacoublay* airport, a military air field very close to Orly. These aircraft are most likely medical evacuation services. Regarding the 8pm peak of traffic at Roissy Charles de Gaulle, the air traffic controller figured out that this corresponds to numerous Air France aircraft which fly back to Paris to spend the night at the airport before leaving again on the next morning.

Both controllers saw the main strengths of our tool in an educational setting, more specifically for air traffic controllers and analysts in training. To the best of their knowledge, no previous tool is available to display the recorded traffic flow with their directions. They reported that this tool provides a perfect visualization system to better understand the structure of the air space and how traffic flows *tangle*. Furthermore, the flow evolution is of a great interest to better grasp the temporal traffic density. Additionally, they said this tool can also be used for communication purposes, to show to a general audience how traffic is distributed over France. The displayed density map is really helpful to correctly grasp the actual flow density. Combined with the particle density, it creates a background image which better emphasize the local density.

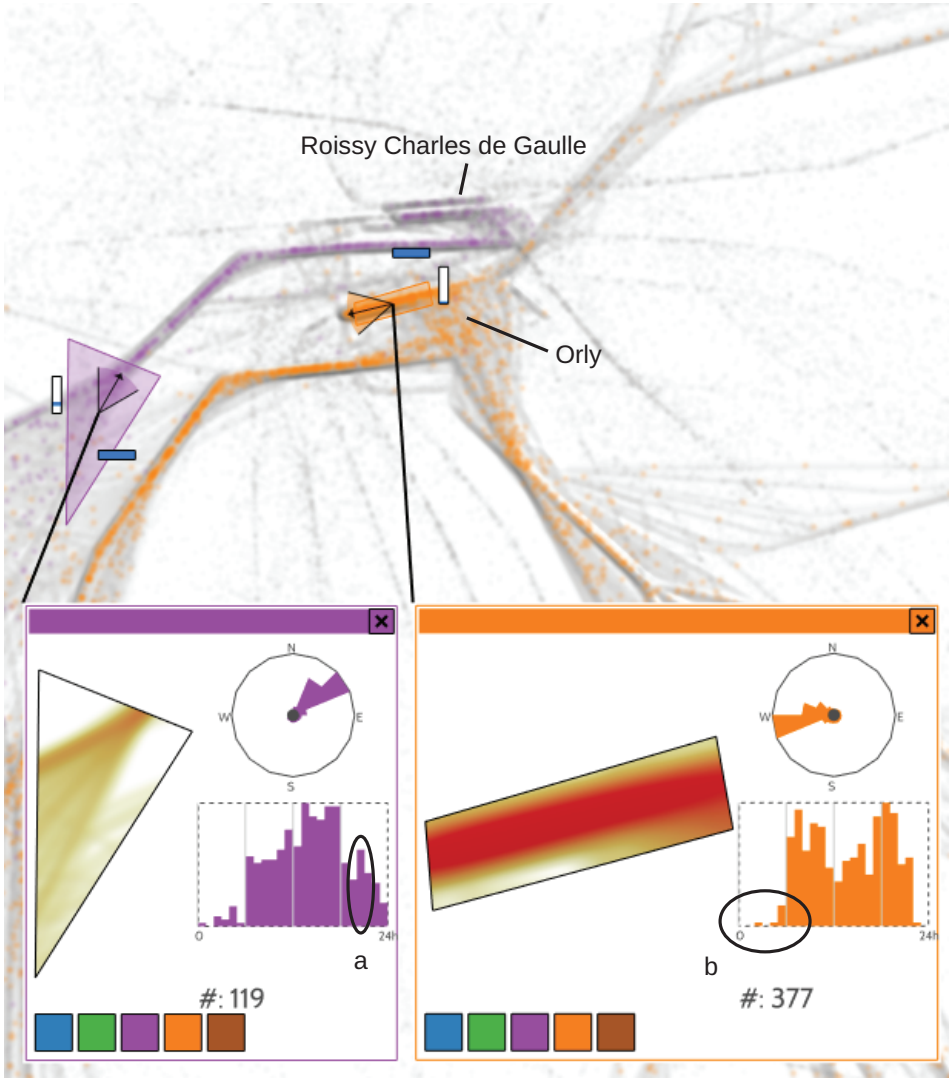


Figure 3.7: Selections of a traffic flow coming into *Roissy Charles de Gaulle* from the south-west (purple) and a traffic flow landing at *Orly* (orange) with the outliers identified by the air traffic controllers: A peak of traffic around 8pm in the purple flow (a), and landing aircraft at Orly while the airport is not yet open (b).

The experts also saw many more operational usages: This tool can be complementary to existing ones to perform statistics. This can be a great asset in order to modify, forecast and assess air space structure and evolution. The few available tools do not display visual information but only textual and graphical statistics.

Our air traffic experts were regularly consulted during the design of our tool and the effectiveness of our design choices were regularly validated with the

experts. They especially appreciated our selection widget to filter flows. The multidimensional filtering technique (rotation to a different data dimension) was also qualified as appealing to better understand how trajectories tangle in 3D.

We have chosen not to overlay our visualization of the aircraft data on top of a map because our domain experts know the region very well and consider a map unnecessary clutter. If our visualization is to be used for educational or communication purposes, however, a map would be advisable. In the vessel use case of Section 3.8.5, our visualization has been overlaid on top of a map.

In the following sections, we describe more use cases for our tool in both the air traffic domain and the Maritime domain.

3.8.1 Comparing Flows

Aircraft follow flight routes, ordered sequences of geographically referenced locations. Actual aircraft trajectories do not always follow the exact location of the flight route. Air traffic controllers can dynamically alter the route for reasons of safety and optimization. Investigation of such flight routes, or, traffic flows, is valuable to better understand airspace congestions and to improve flight regulation and safety. In this use case, we investigate two traffic flows crossing France in opposite directions. These flows correspond to flight transits between cities from the south-west to cities in the north-east, and vice versa: *Geneva*, *Lyon*, *Toulouse*, and *Madrid*. To avoid colliding aircraft, these parallel flows are geographically separated. Using our approach, we can select each flow with our selection widget, and define the altitude range corresponding to the desired flight routes–see Figure 3.8 where the altitude ranges are visualized using the small blue bars. The dynamics of these flows can be investigated separately, or can be directly compared by stacking their windows. As we can see, the traffic flow heading south-west is denser in the morning, while the traffic flow heading north-east is denser in the afternoon. According to our experts, this can be explained by passengers preferring to arrive in the morning in the south of Europe and wanting to return at the end of the day.

3.8.2 Landing and Take-off

In this scenario, we investigate landing and takeoff events at *Roissy Charles de Gaulle*, the main airport in France–see Figure 3.9. To do so, we create two different selection boxes located before and after the runways. Next, we rotate to the altitude view to display the altitude and change the altitude range of the selections to only select aircraft that are landing or taking off, respectively, *i.e.*, aircraft with low altitude. In Figure 3.9, blue traffic flows correspond to take-offs, while green traffic flows to landings. Both traffic flows are similar with around 850 aircraft each. The visualizations of the traffic flow dynamics show the specific departure and arrival sequences. Our experts explained the patterns as follows: In the time line of the departing aircraft (blue in Figure 3.9), we can see (1) postal aircraft taking off between 0am and 5am, (2) two big departure sequences, called

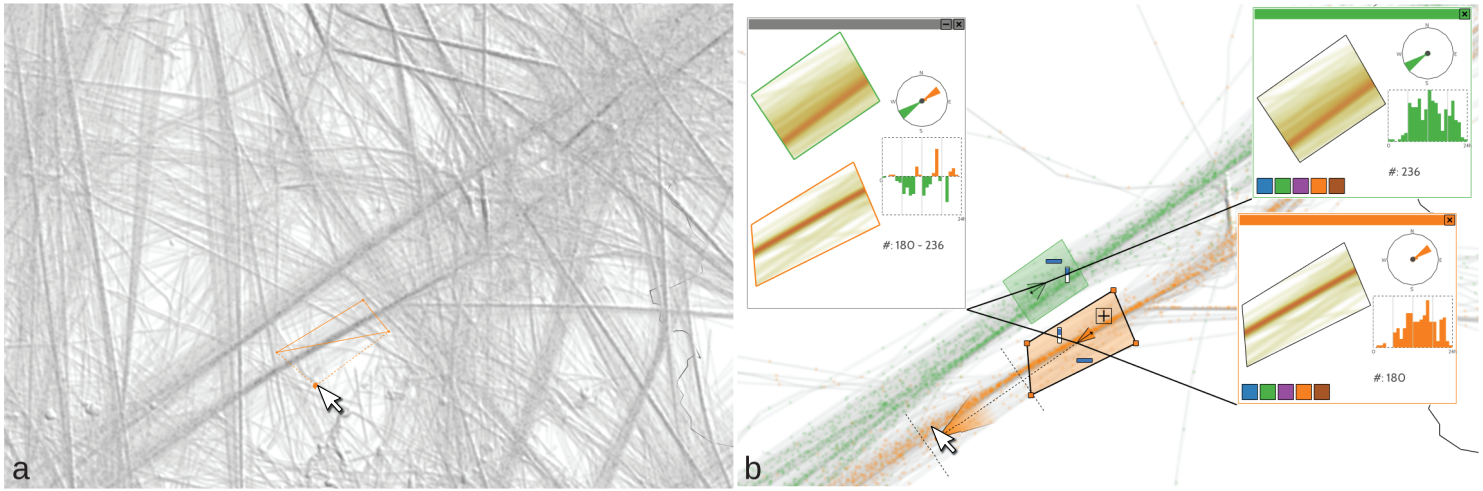


Figure 3.8: (a) Traffic flows are visualized using a density map and a particle system. They can be selected using our novel selection widget, and are visible through coloured particles. (b) Two traffic flows of aircraft moving in opposite directions have been selected. The green traffic flow is moving south-west, and the orange traffic flow is heading north-east. Their dynamics can be explored and compared through movable windows, which show distributions of direction, and density over time. The gray window shows the difference between the two traffic flows.

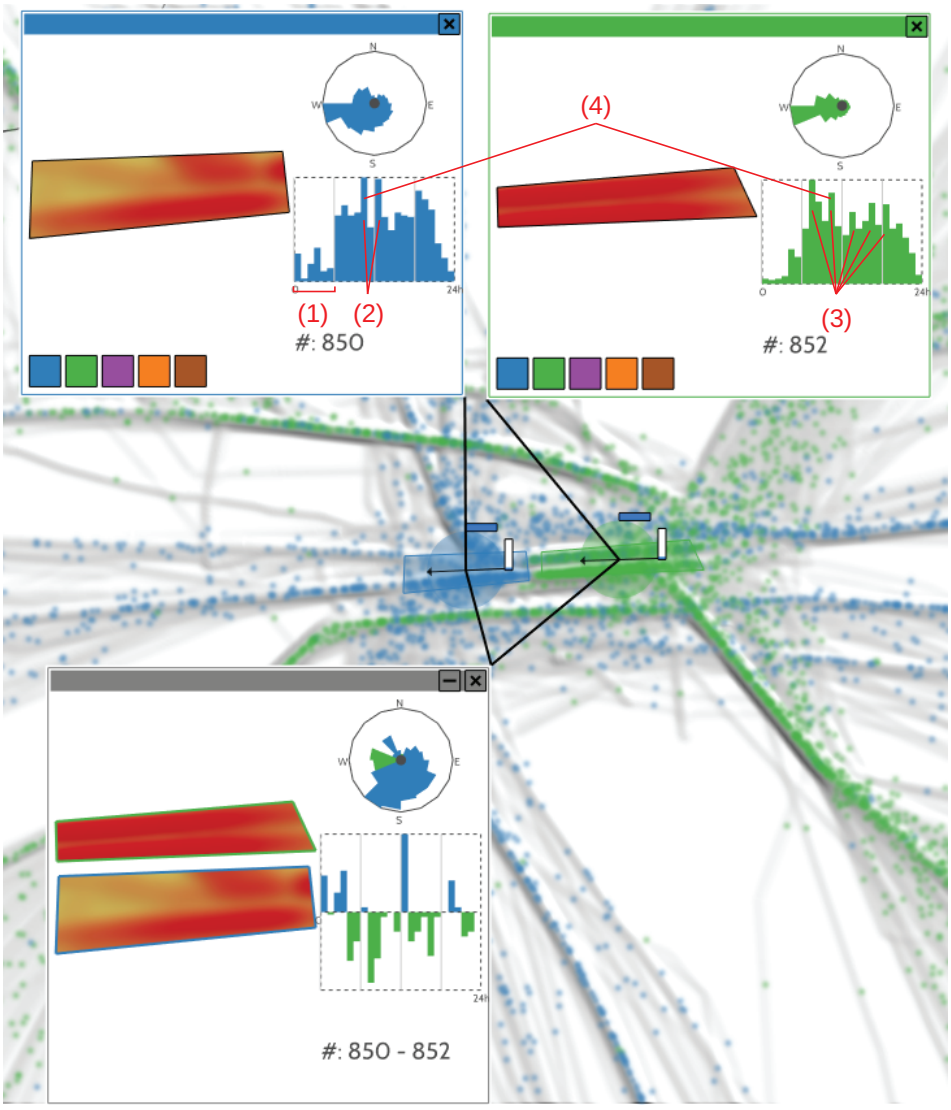


Figure 3.9: Aircraft traffic flows over the airport of *Roissy Charles de Gaulle*, France. Take-off (blue) and landing (green) sequences have been selected and their dynamics can be compared using the windows. According to our experts we can see postal aircraft taking off (1), several large departure (2) and landing (3) sequences, and a peak in both landing and take-off at the same time.

hubs, at 10am and 12am. In the time line of the landing aircraft (green in Figure 3.9) we can see (3) several peaks in landing aircraft starting after 5am. The aircraft landing in the evening stay at the airport during the night. The peaks in landing and take-off generally do not coincide except for a peak between 10am and 11am (4). This information can be used to optimize aircraft scheduling.

3.8.3 Altitude

In this scenario, we investigate the aircraft distribution over different altitudes. In order to optimize fuel consumption, aircraft remain at the same altitude, also called Flight Level (FL) as much as possible. FLs are expressed in feet; for instance an aircraft flying at 30 000 feet (approximately 10 km high), has a Flight Level of 300. Stabilized aircraft heading east (with a direction between 0 and 179 degrees) have an *odd* FL (*i.e.*, 310, 330, 350, 370, etc.). Aircraft heading west (with a direction between 180 and 359 degrees) have an *even* FL (*i.e.*, 300, 320, 340, 360, etc.). This mandatory rule helps to better separate aircraft traffic flows. When selecting flows by altitude, we can see the opposite aircraft direction between odd and even FL–see Figure 3.10. Since only powerful aircraft can reach high altitudes, the flow density decreases with higher FL.

3.8.4 Traffic over Paris

In this use case, we investigate the traffic flow over *Roissy Charles de Gaulle*. When zooming into the Paris area, the flows appear entangled and spread out (see Figure 3.11a). A common approach, however, to deal with such clutter is to apply a visual simplification technique such as edge bundling [102]. Since flows are oriented, we use an extended version of the edge bundling technique which bundles trajectories with compatible directions [103]. This technique reduces visual clutter by aggregating edges into bundled flows. Edge Bundling provides a trade-off between empty spaces and overdrawing [100, 103]. As a drawback, the trajectories are distorted and thus not geographically accurate compared to the original trails. As shown in Figure 3.11b, we can still easily select the desired flows using our approach. Figure 3.11b shows the double cross flow system (four incoming flows and four outgoing), which can be investigated and compared. We can see some flows operate more in the morning, such as flow (1) heading east, while other flows operate mainly after midday, such as flow (2) coming from the south-east. Peaks correspond to hubs, *i.e.*, when more aircraft arrive at the same time to maximize efficiency.

3.8.5 Harbor Infographic

In this use case we use vessel trajectory data of a single day near the Dutch coast based on AIS [111] tracks–see Figure 3.12a. We show how our approach can be used to quickly construct an infographic-like visualization to compare the outgoing traffic of three harbors along the Dutch and Belgian coast: *IJmuiden*,

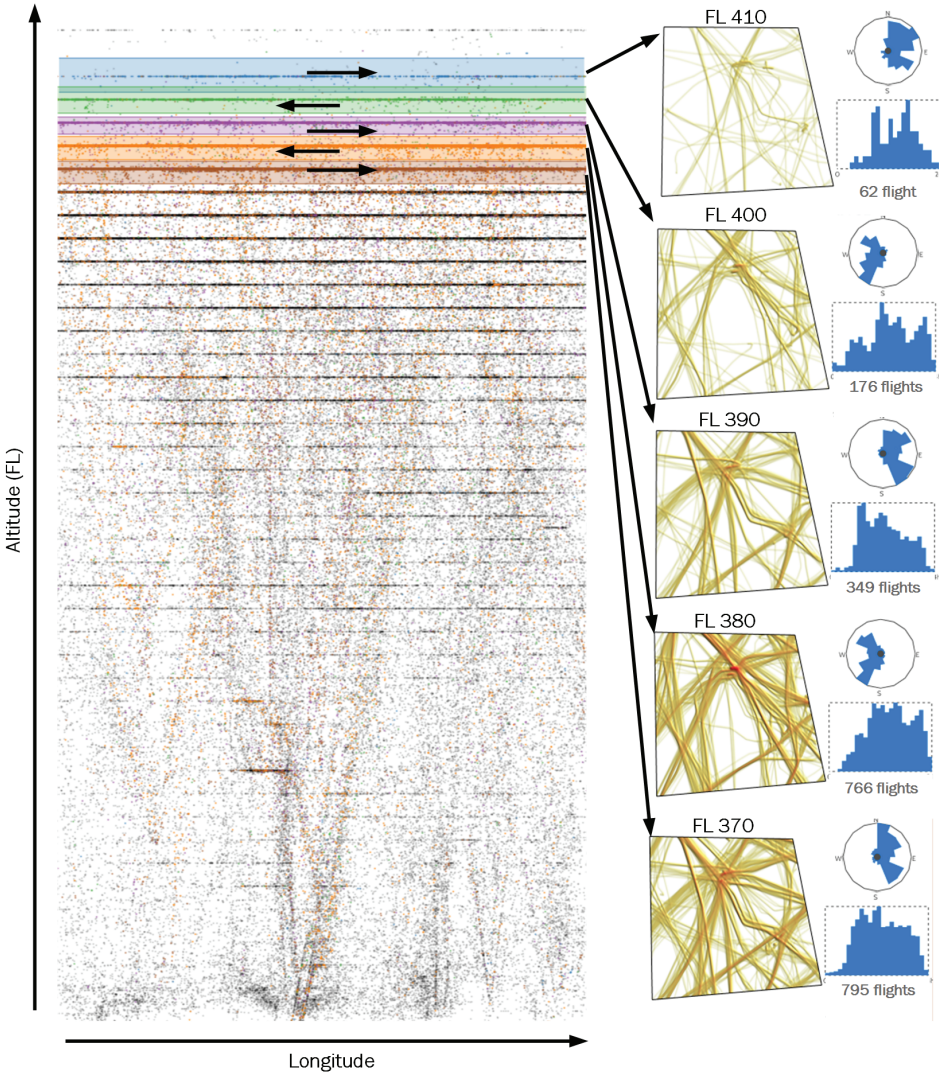


Figure 3.10: The view has been rotated to aircraft altitude where we can see aircraft flying east and aircraft flying west fly at alternating flight levels.

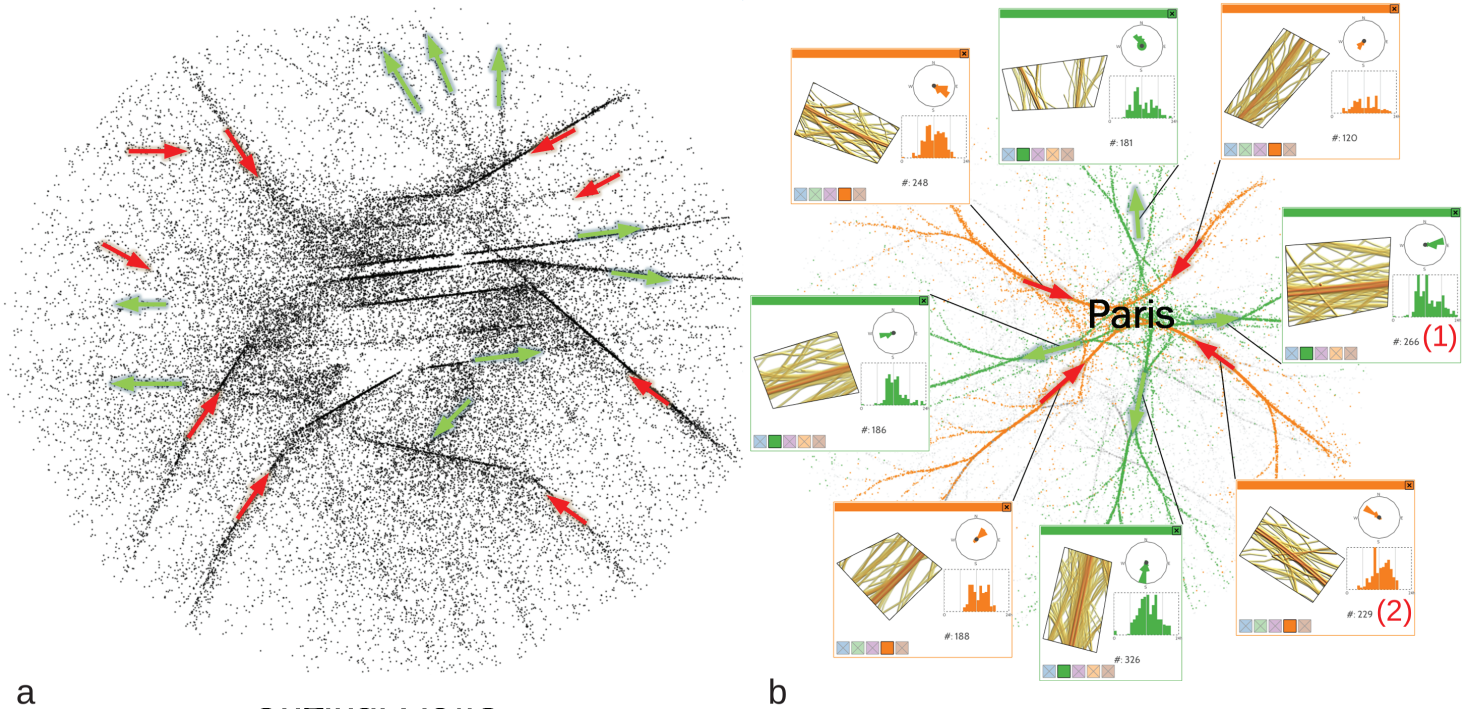


Figure 3.11: (a) An overview of traffic flows over the Paris area. Outgoing traffic flows have been marked with the green arrows, while incoming traffic flows have been marked with a red arrow. (b) The traffic flows have been bundled, selected, and the dynamics of these traffic flows are displayed using the movable windows.

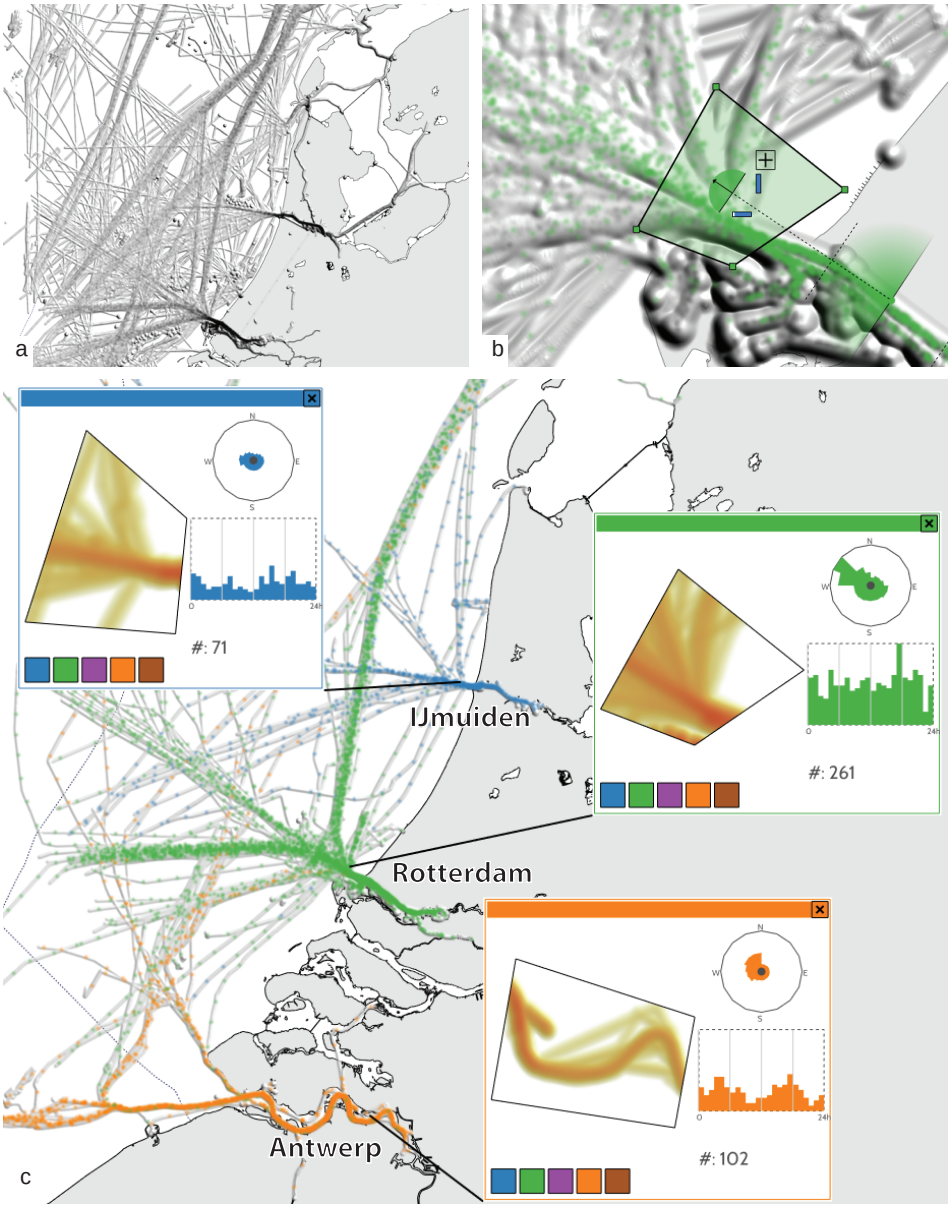


Figure 3.12: (a) An overview of vessel traffic flows near the Dutch coast. (b) Vessels leaving the port of Rotterdam have been selected. (c) A visualization of the outgoing traffic of three harbors along the Dutch and Belgian coast: IJmuiden (blue), Rotterdam (green), and Antwerp (orange). The wave pattern in the histogram of the port of Antwerp is caused by the access to the port, the river *Scheldt*, being subject to the tides.

Rotterdam, and *Antwerp*–see Figure 3.12c and the supplemental video. For each of the three harbors, we use a different selection color to select an area around the harbor mouth using a heading range of vessels moving away from the harbor–see Figure 3.12b. We increase the minimum of the velocity range so that we only select moving vessels, and open a window for each selection. By selecting the "render only selected trajectories" option, hiding unselected particles, and hiding the selection widgets, we get the visualization of Figure 3.12c.

Due to the colored particles, we can now see where vessels leaving the respective harbors typically go. The floating windows now serve as annotations for each harbor that show the dynamics of traffic leaving their respective harbors. We can see here that the port of Rotterdam is by far the largest of the three. Also, we can see a wave pattern in the histogram of the port of Antwerp. This is because the access to the port of Antwerp, the river *Scheldt*, is subject to the tides.

3.9 Conclusions & Future Work

In this chapter, we explored the design space to visualize and interactively explore traffic flows in moving object data sets. Traffic flows have intrinsic properties that can be displayed and analyzed even in dense visualizations: location, direction, and intensity. In addition, traffic flows can have other dimensions, such as altitude, speed, type of moving object, etc., which can evolve over time. The user can discover and identify traffic flows using a visualization which is a combination of a density map and a particle system. Then, using our novel selection widget, these traffic flows can be selected. When selected, the user can explore traffic flow dynamics using annotation windows, which can be dragged on top of each other to compare multiple traffic flows. We demonstrated our work using a number of use cases, which have been validated by air traffic analysts.

As future development, we plan to extend this software together with air traffic control practitioners to provide "a what if" system that enables the user to remove, change, or simulate trajectories. This will help the user to understand the impact of traffic flow modifications: What if this traffic flow is redirected in that direction? What is the impact on the other traffic flows? This can, however, also be used for other data sets. For example, to study what the effect on shipping lane usage is if more vessels leave the harbor in the morning.

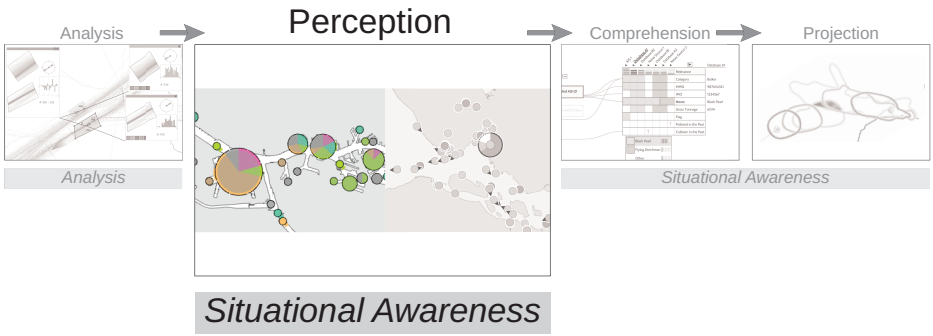
We would also like to investigate the scalability of our approach by using larger data sets. Currently, we estimate parameters for the particle system, such as particle spacing and particle weight P_α, automatically for the data sets discussed in the use cases. These parameters are based on the average density of the data set. We would like to generalize this estimation so that we can automatically generate suitable parameters for any input data set. Additionally, we would like to improve this parameter estimation to take into account other relevant data characteristics such as the spread of the trajectories.

And finally, we would like to investigate generating infographic-style visualizations further using our approach. For example, the arrows shown in Fig-

ure 3.11b and similar annotations can be generated automatically, but currently are not.

4

Non-Overlapping Aggregated Multivariate Glyphs for Moving Objects



How can we help an operator in perceiving a situation?

The operator needs to be enabled to perceive the status, attributes, and dynamics of relevant elements in the environment. In moving object visualization, such as in the maritime operational picture, objects and their attributes are commonly represented by glyphs on a geographic map. In areas on the map densely populated by these objects, visual clutter and occlusion of glyphs occur, which is very harmful to obtaining situational awareness. We propose a method to solve this problem by partitioning the set of all objects into subsets that are each visualized using an aggregated multivariate glyph which shows the distribution of several attributes of its objects, such as heading, type and velocity. We choose the combination of subsets and glyph design such that the glyphs do not overlap and the number of subsets is approximately maximal. The partition is maintained and updated while the objects move.

The contents of this chapter have in part previously appeared in [175].

A video on the contents of this chapter can be found at: https://youtu.be/kUxqFd3F7xM.

4.1 Introduction

Plotting glyphs on a 2-D area is a common operation in many domains. Visual clutter is a classic visualization problem, especially for mobile devices with a limited display size and resolution or very large data sets. When a large number of glyphs are plotted, they may overlap, possibly resulting in complete occlusion of some items. So, in general, only a subset of all data items is visible on the screen. Based only on what is visible, erroneous assumptions may be made [165] or important information may be missed [64]. This type of visual clutter can occur in any type of visualization of a large number of data items, ranging from scatter plots to cartography. Rosenholtz *et al.* [165] define clutter as "the state in which excess items, or their representation or organization, lead to a degradation of performance at some task". The clutter problem typically occurs in geographic applications where many glyphs or labels [16] are used to mark the location of items of interest on a map.

The clutter problem can be tackled in a variety of ways, often depending on the type of data and the application domain. In the case of glyphs, a common solution is to displace them to eliminate overlap [17, 61, 196] or to simplify the data set by leaving out a subset of the data elements [20, 61], or a combination of both [21]. Reducing the number of displayed data items or changing their position may be undesirable in some domains. This is especially so in the maritime domain where the situational awareness of an operator is based on the presence and actual position of vessels. If the positions of individual vessels are changed or some vessels are not visible due to sampling, the operator may get a false impression of the situation. Another common solution to cluttering is to aggregate data items and visualizing aggregate information in a density map [179] or using multivariate glyphs [210]. Additionally, user interaction can be applied for zooming or for using a lens that simplifies [65], deforms or clusters [101] the underlying data.

Clutter is also a problem when plotting dynamic data such as trajectories of moving objects on a 2D surface. Two general solutions to reduce this clutter problem are clustering the trajectories [7, 12], or aggregating the trajectories in a density map [124, 217].

In this chapter we focus on time instances of time-series data. This data can range from moving objects in geographic space to time-series data displayed in a scatter plot. We focus, in particular, on the maritime domain and design our visualization with that in mind. Our visualization, however, can be generalized to other domains with moving objects or time-series data in general, such as in traffic management and congestion control, crowd monitoring for public safety and animal movement [55].

We start with multivariate time-series data, which is a set O that contains a number of objects o. Each object has a set of attributes $\boldsymbol{\alpha}_o(t)$ that contains both constant and dynamic attributes. We assume that at any time t we can only retrieve the attributes of the objects at a time smaller than or equal to t.

A visualization is required that allows a user to recognize the density distri-

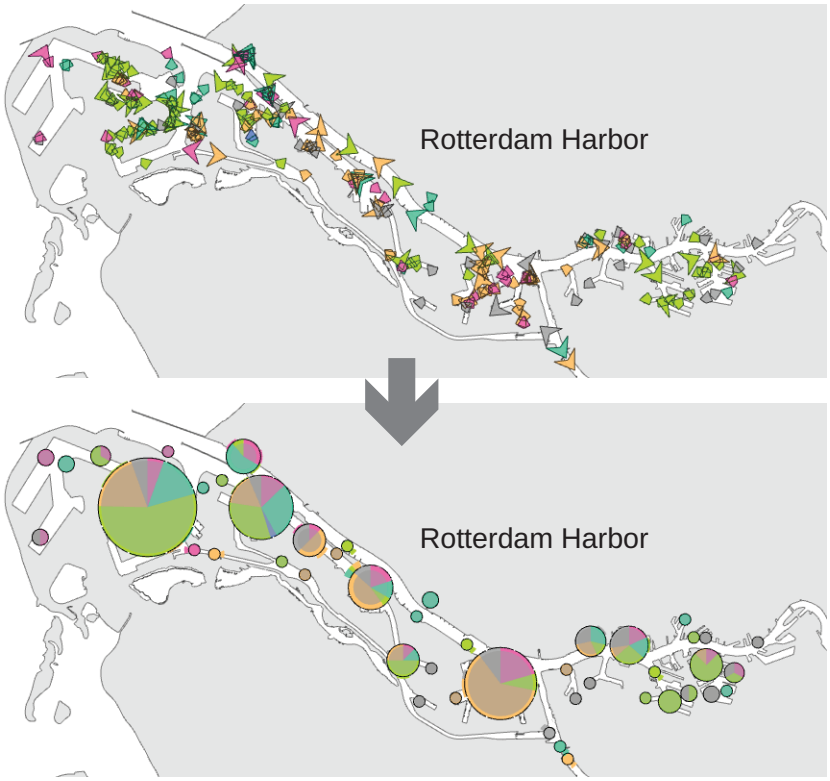


Figure 4.1: Single moving objects are grouped into larger visual representations to reduce visual clutter and overlap. On the left we show individual vessels in a harbor visualized using vessel glyphs, while on the right we show them grouped and visualized using aggregated multivariate glyphs.

bution and global patterns of the data set despite the clutter problem, for static as well as dynamic cases. We propose an aggregation method that maintains a dynamic partition of the objects as they move over time. Each subset of the partition is visualized using a circular, multivariate glyph designed to show the distribution of a select number of attributes–see Figure 4.1. These glyphs are scaled based on the number of moving objects in the subset. The moving objects are partitioned such that the following requirements hold:

- **R1** There should be no overlap or occlusion between visual representations of the subsets;
- **R2** The subsets should be as small as possible;
- **R3** The user should be enabled to estimate the point density of areas;
- **R4** The user should be enabled to recognize patterns in the attributes of the data;
- **R5** The user should be enabled to see more detail by zooming in.

In a user study we compare our method to two alternatives. We investigate tasks such as comparing groups of objects and maintaining situational awareness. We find that our method does not perform significantly worse than competitive visualizations with respect to correctness and performs significantly better for density comparison tasks in high density data sets. From the results of a questionnaire, we find that the participants of our user study find our method significantly better than its competitors for all tasks combined.

4.2 Related Work

We first discuss a number of techniques related to clustering and clutter reduction in both an algorithmic and visualization context in Section 4.2.1 Following this, we discuss some related work on multivariate glyphs in Section 4.2.2.

4.2.1 Clutter Reduction

A common approach to reducing or eliminating overlap is to simplify or sample [52] the data. De Berg *et al.* [47] present an efficient algorithm that simplifies a set of points on a map. A subset of the original point set is chosen such that the distribution of the subset approximates that of the original set. Ellis *et al.* [63] propose a sampling lens that allows a user to interactively reduce the plotting density in scatter plots and PCPs under the lens. This enables a user to investigate saturated areas in the visualization. Similarly, Hurter *et al.* [101] use a lens to interactively cluster or push aside edges in a dense graph. Bertini and Santucci [20] use non-uniform sampling to simplify the representation of a scatterplot. In a later paper [21] they combine sampling with displacement to get a simplified representation of a scatterplot that preserves the observable density differences of the original. Trutschl *et al.* [196] use an algorithm based on Self Organizing Maps (SOM) to displace records in a k-dimensional visualization. The resulting spatial reordering of overlapping records eliminates their overlap and emphasizes the relationships between neighboring records. Since our main focus is the maritime domain, no individual object can be displaced (*i.e.*, a false presentation of position) or have its presence hidden. Hence, we do not use simplification or displacement.

Our solution for the overlap problem is related to clustering. Methods already exist to cluster moving objects over time using a static clustering. Har-Peled [88] proposes a clustering method that provides a static clustering for a set of moving points in $\mathbb{R}^d$. This assumes that all movement data is known beforehand and gives a single clustering. We, however, consider an operational real-time view in which only the current and possibly historical positions of the moving objects are known, but nothing is known about their future positions. Gao *et al.* [80] propose a randomized algorithm that maintains a set of clusters of points moving over the plane. Each moving point has a fixed region defined by an axis aligned square. The clustering is achieved by finding a minimal subset of points called centers,

such that each point in the data set is within range of at least one center. Whereas Gao *et al.* cluster for more efficient communication between collaborating mobile devices, we group points to overcome the problem of visual overlap.

Aggregation is another common approach to visualize large data sets. Scheepens *et al.* [179] use convolution to visualize large sets of vessel trajectories in a density map. The authors split the data set on attributes into multiple density maps, which can be visualized using multivariate visualization techniques. The main difference is that their method is intended for finding and analyzing trends of movement, while here our method is intended for a real-time view of the data. Andrienko and Andrienko [3] propose methods for visualizing spatio-temporal trajectories for the interactive, visual analysis of large trajectory data, such that it can be investigated from the point of view of trajectories or from the point of view of the traffic. In the traffic-oriented view, attributes of traffic, such as velocity or direction, are shown using aggregated glyphs. In the trajectory-oriented view, entire trajectories are aggregated based on origins and destinations and visualized using arrow-like glyphs. In another paper, Andrienko *et al.* [10] use spatial generalization and aggregation of large movement data. Based on significant points extracted from the data a partition of the underlying space is created. The trajectories are then visualized as aggregated flows between these areas. We use an aggregation represented by a multivariate glyph.

4.2.2 Multivariate Glyphs

A common approach in cartography for the visualization of point sets is to replace a set of overlapping or crowded glyphs by a larger glyph of the same type [61] or a composite glyph. In their traffic-oriented view, Andrienko and Andrienko [3] draw glyphs with directional bar diagrams that show movement data aggregated by compass directions and circles that show the proportion of slow-moving or stationary objects.

In a geo-referenced information visualization system intended for mobile devices, Carmo *et al.* [36] propose to divide the map space into a regular grid. The number of elements in each cell is counted and if it exceeds a pre-defined amount, an aggregated glyph is shown with up to four symbols depending on the symbols associated with the elements in the aggregation. This aggregated glyph is visualized as a stack of multiple glyphs, giving a user the intuitive idea that the glyph represents more than one element.

Chlan and Rheingans [41] propose an aggregated glyph to visualize the distribution of large quantities using a 2D or 3D glyph. In the 2D case the glyph is rendered using three shells. The outer shell is an ellipse showing the maximum extents of the data range in both the X and the Y attribute, the middle shell is an ellipse showing the variability of the data in the X and Y range, and the inner shell shows both the average and the standard deviation in the X and Y range. The colors of the shells are used to visualize a third attribute.

Scheepens *et al.* [179] use a pie chart glyph to enhance their visualization of multiple density maps. These glyphs are placed at peaks of aggregate density and

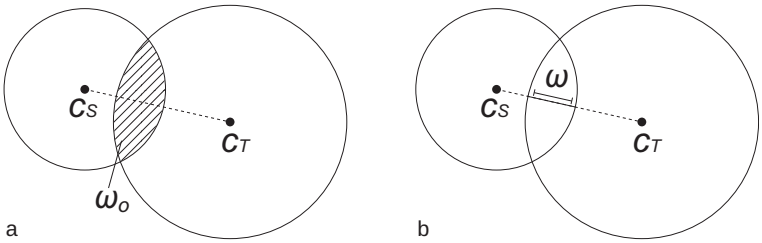


Figure 4.2: The areas of influence of subsets S and T overlap with the area of overlap ω_o (a) or the penetration depth ω (b).

show the distribution of the underlying density maps. These glyphs are spread out as evenly as possible over the density map to give an overview of the density distributions over the entire map without overlap.

4.3 Approach

We divide our object set O into a partition $\{S_1, ..., S_m\}$ of non-empty disjoint subsets S_i that together span O. Each subset S has a circular area of influence defined by its centroid $\mathbf{c}_S$ and a radius $r_S = r(|S|)$, where r is a function of the number of elements of subset S. We now define two additional requirements:

R6 The areas of influence of different subsets do not overlap;

R7 Small changes in object positions have small effects on the partition.

Since we are interested in solving overlap of glyphs on the screen, we project our positions onto screen space and give the radii r_S in pixels. This also allows the partition to adapt to changes in zoom level, satisfying requirement **R5**.

For reasons of symmetry we chose a circular area of influence. Figure 4.2 illustrates two measures of overlap of two areas of influence: Figure 4.2a uses the area of overlap ω_o, Figure 4.2b the penetration depth ω. The latter is easier to compute and used here:

$$\omega(S,T) = r_S + r_T - ||\mathbf{c}_S - \mathbf{c}_T||. \tag{4.1}$$

If $\omega(S,T) > 0$, we say that the visual representations of subsets S and T overlap–see Figure 4.2b.

4.3.1 Constructing the Partition

We use a greedy algorithm to construct the partition. We construct an initial partition K containing a singleton $\{o\}$ for each $o \in O$ (see function Init in Algorithm 1) and continue with function Merge: While there are pairs of subsets S and T in K for which the overlap $\omega(S,T)$ is larger than zero, we merge a pair with

maximal overlap into a single subset. This loop ends, because the size of the finite set K reduces each step. Furthermore, since only overlapping subsets are merged and since the loop terminates when no more subsets overlap, requirements **R2** and **R6** are satisfied.

The positions of the objects are periodically updated. To satisfy requirement **R7**, instead of constructing a new partition from scratch, we update the existing partition to again satisfy the requirements **R2** and **R6**. In a greedy approach, a new local optimum based on the current partition with updated positions (see function Update) is found by performing a splitting step on all subsets, followed by a merging step on them. In the splitting step, each subset S is partitioned by applying an Init to the objects it contains. In the subsequent merging step, the resulting partition K is cleared of overlapping subsets by applying Merge.

For performance improvement of Algorithm 1, a binary Search Tree (BST) sorted on the overlap measure ω of pairs $(S, N(S))$ is maintained, where $N(S)$ is a nearest neighbour of S in the sense that for no other subset T $\omega(S,T)$ is larger than $\omega(T, N(S))$. Whenever such a pair or its overlap measure changes, its position within the tree is updated. The pair with the largest overlap is now simply the right-most node in the tree. To find the nearest neighbor of a subset efficiently, the subsets are stored in a k-d tree [48] based on the position of their centroids. Whenever two subsets are merged, the k-d tree is incrementally updated.

4.3.2 Analysis

We show the running time of the initialization and update parts of our algorithm by taking the average running time of 100 runs. There are two critical factors in the object data that determine the running time of our algorithm: the total number of objects and the degree of overlap of the representations of the objects. We represent the degree of overlap by the number of objects per pixel. Figure 4.3a and Figure 4.3b show the running time of Init and Update, respectively, for various combinations of object numbers and density. For reference, Figure 4.4 has

Algorithm 1 Partition Construction

```
function Init(O)
    K ← {{o} | o ∈ O}; return Merge(K)
function Merge(K)
    while ∃_{(S,T)∈K×K} ω(S,T) > 0 do
        (S,T) ← pair with maximal overlap ω(S,T)
        K ← (K \ {S,T}) ∪ {S ∪ T}
    return K
function Update(K)
    for all S ∈ K do                                  // split subsets
        K ← (K \ {S}) ∪ Init(S)
    return Merge(K)                                   // merge subsets
```

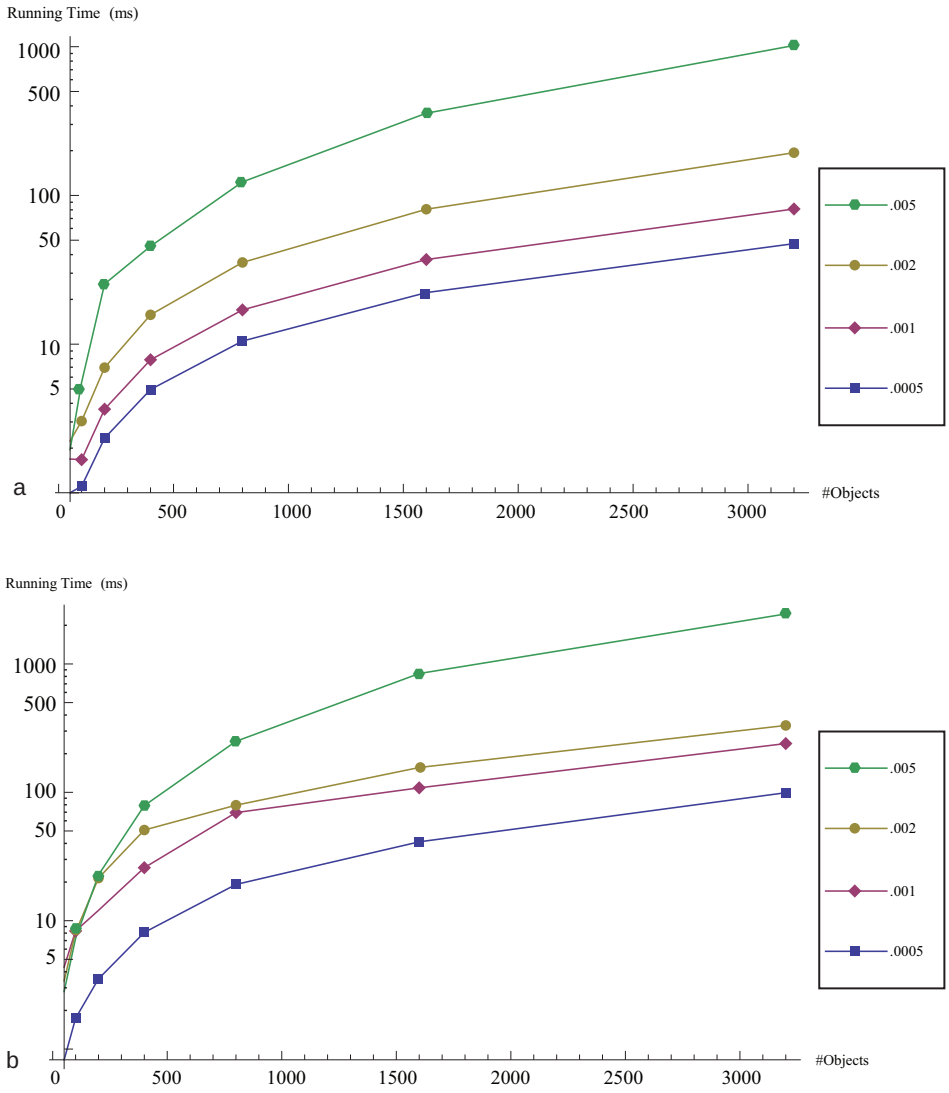


Figure 4.3: The running time in milliseconds for a number of different densities for the partition construction algorithm Init (a) and partition update algorithm Update given in Algorithm 1 (b).

an object density of .002 with 400 objects.

From the graphs we can see that both functions have similar running times, which is not surprising as the update function depends heavily on the initialization function. During updating, however, we need to record and compute additional information required for interpolating and animating between states as explained in section 4.4.2. This would explain the slightly higher running time

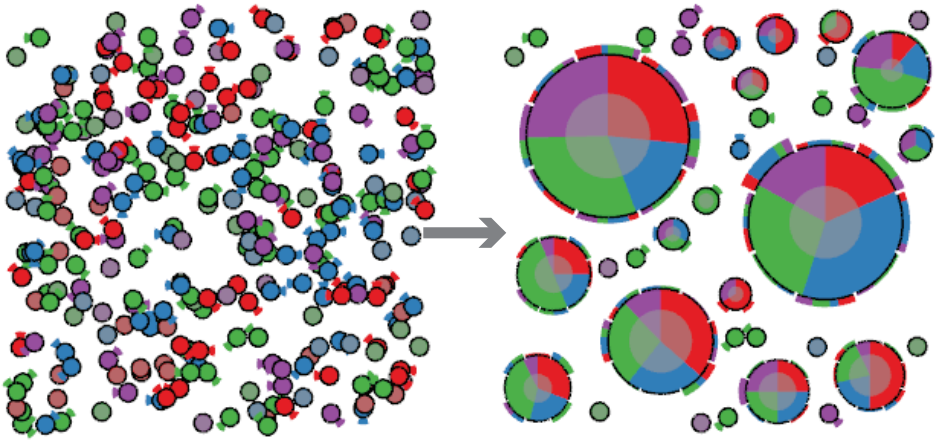


Figure 4.4: Left: A set of overlapping objects. Right: The same set of objects partitioned.

for the update algorithm; even so, it appears to be a constant factor.

We ran our tests on a PC with an Intel Core i7 2.8Ghz processor, 6GB of RAM and a NVidia GeForce 285 videocard. The algorithm is implemented on the CPU. To speed up our algorithm, we could construct the KD-tree on the GPU [223] and parallelize our algorithm as much as possible.

4.4 Visualization

We visualize each subset using a multivariate glyph showing a number of attributes of the moving objects. As a basis we use the position time series of the objects given in $\mathbb{R}^2$, as $(x(t), y(t))$ pairs. The heading $h(t)$, an angle, may then be derived from the displacement vector in this pair over time. While designing our glyphs, we focus on the maritime domain and have the following additional attributes: Vessel type τ and velocity v.

We want to visualize the heading and an abstraction of the velocity of the objects as well as the distribution of object types over these attributes. According to Tufte [197], the surface area of a graphical element should be directly proportional to the numerical quantity it represents. Therefore, to enable visual comparison between subsets, we consistently keep the area of visual elements proportional to the number of objects it represents.

In the following sections, we discuss how we visualize the above attributes in a single multivariate glyph and how we visualize the transition between partitions over time.

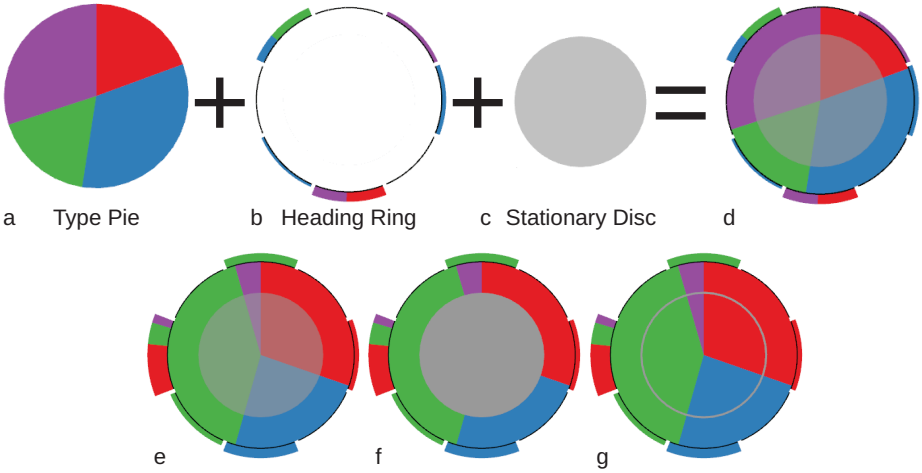


Figure 4.5: Our subset glyph (d) is composed of a Type Pie (a) showing the distribution of object types, a heading ring (b) showing the types and number of objects moving in a limited number of direction ranges, and a Stationary Disc (c) showing the proportion of stationary objects. A number of design alternatives for the stationary ring: A translucent disc (e), a solid disc (f), and a ring (g).

4.4.1 Subset Glyph

We require a simple, scalable and compact visual element to visualize the distribution of the object types in a subset. Candidates are a divided bar chart and a pie chart. Other candidates such as a star glyph can give the unwanted impression of directionality or orientation when the distribution is heavily skewed. We have chosen for a simple shape, instead of, *e.g.*, a shape that follows a river, to keep the type distribution visualization clear and consistent and the glyphs simple. We also prefer the visual element to be radially symmetric in order to enable encoding the heading around it. So, the natural choice is the pie chart. Furthermore, a pie chart performs better than a divided bar chart for comparisons, if they come in various sizes [96]. In Figure 4.5a, we show the distribution of the vessel types that occur in a subset with a standard pie-chart using a domain prescribed color coding. Due to the limited amount of categories that can be displayed using colors or using a pie chart, we suggest aggregating types using some hierarchy, or a selection of types considered interesting by the user, when necessary.

Replacing a number of visually conflicting glyphs by a single larger glyph is a common operation in cartography [61]. We use the size of the type pie chart to visually encode the number of objects in the corresponding subset S. We get the radius $r(|S|)$ of the area of influence from the equation

$$\pi r^2 = |S| c_0, \tag{4.2}$$

where c_0 is a scaling constant. An alternative could be to use a set of discrete sizes to represent ranges of $|S|$ using some optimal scale [127]. A disadvantage of using

size to represent the number of objects $|S|$ is that more geographic context may be obscured by the glyph, however, we solve this using the interaction discussed in Section 4.4.4.

The velocity attribute is reduced to a binary variable: A vessel is either moving or it is stationary according to some velocity threshold. This gives us two attribute sets: $S_{h,moving}$ and $S_{h,stationary}$. The heading and type distribution of $S_{h,moving}$ is shown in a ring outside the pie chart, similar to the directional rings of Guo *et al.* [85]. A configurable number N of segments is created, where each segment s_i represents headings in the interval $[b(i-1/2), b(i+1/2)]$ with $b = 2\pi/N$. The intervals are shifted by $b/2$ such that the segments center around the standard compass directions if $N = 4$ or $N = 8$. Each segment is then visualized as a part of the heading ring showing the distribution of types, visualized using the angle, and the number of objects in the segment both in total and per type using the surface area–see Figure 4.5b. Since we use the angle to visualize the distribution of the types over a segment, we determine the height h_i of a segment s_i from

$$\pi(r + h_i)^2 - \pi r^2 = 4|s_i|c_1/N, \tag{4.3}$$

where r is the radius of the type pie, $|s_i|$ is the number of objects in s_i, and c_1 is a weight representing the intended number of pixels per object. To visually separate the segments, a small white space is rendered between each segment–See Figure 4.5b. An alternative would be to use interaction, *e.g.*, by showing a set of weighted arrows on mouse over. Our maritime domain experts, however, have indicated that the heading always needs to be visible for global movement patterns to be recognized without the need for interaction.

The number of stationary objects $n_s = |S_{h,stationary}|$ is visualized using a semi-transparent disc on top of the type distribution–see Figure 4.5c– with the radius of the disc set to $r(n_s)$. There are two alternatives for visualizing the stationary disc: A ring and a *solid* disc–see Figure 4.5. For a ring it is not intuitive whether n_s is represented by the radius or the area contained within the ring. For a solid disc, it seems intuitive that its area represents n_s, but the disc itself may occlude large parts of the type distribution. An additional alternative would be to visualize the number of stationary objects per type by changing the radius of the stationary disc per slice accordingly. This makes the glyph, however, harder to read and does not give a clear overview of the proportion of total stationary objects.

4.4.2 Animation

The changing of the partition due to the movement of the objects may confuse the user. Therefore, we propose to use a short animation to show that subsets have split or merged such that the user can maintain a mental map. We identify two basic cases: A subset S is either merging into a new subset S^+ with one or more other subsets, or a subset S has split off from a larger subset S^+. In both cases we interpolate between the glyph of the original state and the glyph of the new state. Since objects continue to move, it may, for instance, occur that a

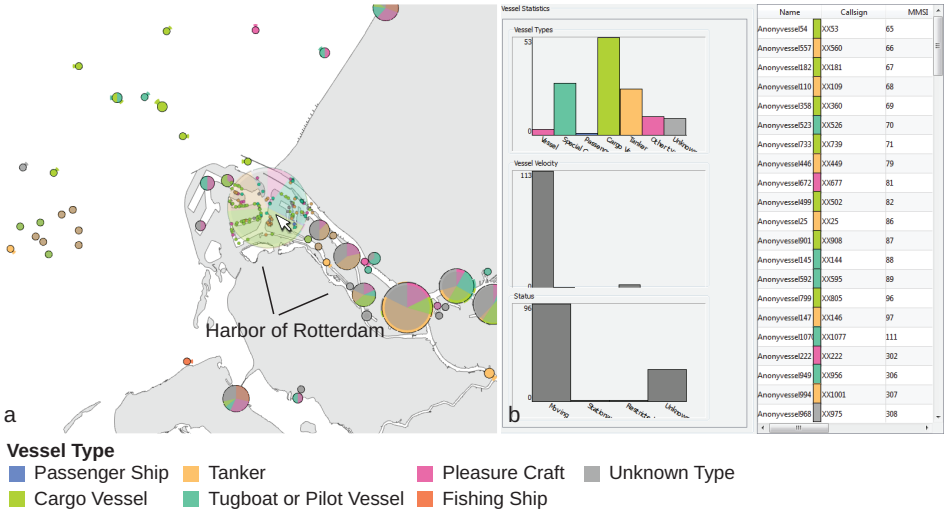


Figure 4.6: An application of our partition method in the maritime domain (a). Through interaction the user can investigate point groups: By holding the mouse over a glyph, the user can see the spatial distribution of the points in the subset, and by clicking on a glyph the user can see statistics of the vessels in the subset (b), such as the distribution of vessel types, vessel velocities or vessel states (*e.g.*, the vessel is stationary).

subset that is merging has to split again. Therefore, each animating subset S has a collection A_S of sets O_i of objects; each O_i is either merging into S or splitting off from it. Each set O_i has a weight w_i that decreases from one to zero for splitting sets and increases from zero to one for merging sets. The attributes used during visualization of S are computed as a weighted sum of the attributes of S and the attributes of each O_i. For example, the number of elements n is given by $\sum w_i|O_i|$. By changing the weights at every update step depending on a configurable, in our case linear, function, the glyph of subset S morphs into its new state.

Even though moving data elements may overlap visually, the human visual system is still able to track the individual moving objects to some extent [148]. This means that an alternative to partitioning all points may be to partition only stationary and slow-moving objects, while visualizing the fast movers as single points. A user can then see global patterns within the slow-moving cluttered objects and is still able to visually track the fast moving objects. In the maritime domain, however, there are generally no vessels moving fast enough on screen to warrant such an approach. We have therefore decided to partition over all objects.

4.4.3 Real-world application

Our method is designed for the visualization of moving vessels. In harbor areas, busy shipping lanes or anchorage areas, a large degree of overlap can occur.

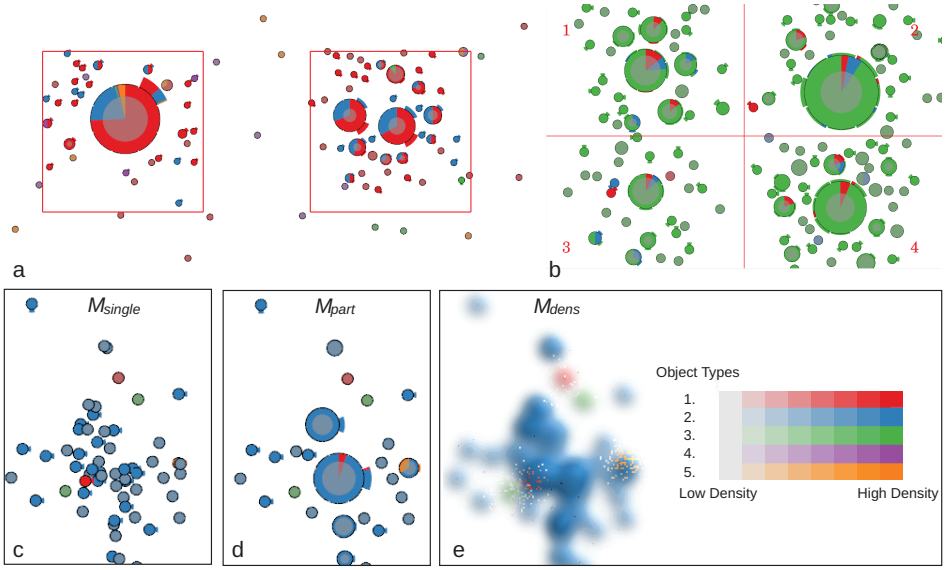


Figure 4.7: In our user study we show two red squares containing a distribution of points each for the static tests (a), and for the dynamic test we split the screen into four numbered quadrants (b). The visualizations used are M_{single} (c), M_{part} (d), and M_{dens} (e).

An operator, whether working for the coast guard, a navy or a port authority, may miss crucial information if glyphs representing vessels occlude each other. We show the application of our method for an especially busy harbor area in Figure 4.6a. We use a color coding for vessel types that is standard in the maritime domain. In the video accompanying this chapter we show an application of our method to the maritime domain.

4.4.4 Interaction

To enable a user to further investigate the data we offer several forms of interaction. A user can perform zooming and panning operations, resulting in automatically updated grouping, due to the partitioning in screen space. When the mouse hovers over a subset, the individual points in the group are visualized to show their spatial distribution–see Figure 4.6b. Clicking on a subset results in an enlarged view with details on its objects in a separate window–see Figure 4.6c. The user can select parts of the subset to investigate further. The accompanying video shows this interaction in practice.

4.5 Evaluation

We evaluate our visualization by comparing our partition method M_{part} to two alternatives that are commonly used in the maritime domain: a KDE [182] based method M_{dens} and a single point visualization M_{single}. For each method we use the five element color map as shown in Figure 4.7e. For M_{dens} we use the visualization techniques proposed by Scheepens *et al.* [179]. For each object type, the positions of the objects are convolved onto a density field. Each density field is assigned its own color map where the saturation increases with the density. The density fields are then combined using the block composition operation of Scheepens *et al.* [179] and shaded–see Figure 4.7e. To avoid bias towards a specific visualization and to allow testing of the aggregation method's validity, the single point visualization M_{single} as shown in Figure 4.7c is chosen similar to the partition visualization for subsets of size 1.

We tested the ability of subjects to recognize density and patterns in a static visualization and their ability to maintain situational awareness in a dynamic one. We believe these tasks represent the process of basic understanding of the situation in the maritime domain. In the user study we first explained the three visualizations and the setup of the study to the subject (10 min). Then the subject performed three static tests (each 5 min) and one dynamic test (10 min) and finally, filled out a questionnaire (5 min). The actual study took 30 to 45 minutes per subject. We tested with 13 males and 2 females with a computer science background, on a single machine with a 1690×1050 screen, and with the visualizations in 1200×800. To show our method can be generalized and does not require specific maritime domain knowledge, we have chosen not to use maritime domain experts.

4.5.1 Static tests

In our static tests we asked the subjects to perform three different tasks T_i with varying visualizations and data sets, such that each data set is tested for each visualization in random order. For each task two red squares are displayed, one on the left side and one on the right side–see Figure 4.7a. The subject is posed a question about the objects in the squares and can answer the question by either entering *left* or *right* using arrow keys on the keyboard. The questions differ per task and are as follows:

T1 Which square contains more points? Left or Right?

T2 Which square contains more blue points? Left or Right?

T3 Which square contains more blue points heading approximately North-East? Left or Right?

With these three tasks we test the abilities of the subjects to compare densities, to compare the type distributions of Section 4.4.1, and to interpret the heading ring of Section 4.4.1, respectively.

Each task has a set of *special* points which the subject has to compare, *e.g.*, for **T1** all points, and for **T2** all blue points. A dataset is parameterized by two parameters (n, p) for **T1** and (p_s, p) for **T2** and **T3**. The number of points n is set to low density (50 points), medium density (500 points) and high density (1000 points) for task **T1** and a random number in the range $[50, 1000]$ for **T2** and **T3**. The total number of special points n_s is determined by $n_s = p_s \cdot n$, where p_s is set to 100% for **T1** and to small (5%), medium (10%) and large (15%) for **T2** and **T3**. And finally, the percentage difference p between the number of special points in the left square and those in the right square is set to small (5%), medium (10%) and large (15%). We generated a data set for all combinations of each of the parameters n, p and p_s, leading to 9 unique configurations of parameters per task. For each of these configurations we generated three variants leading to a total of 27 unique data sets. For each subject all unique configurations of parameters are used, but only one of the three variants is randomly selected.

Data Generation

We generated our test data by first placing n objects within each square using a normal distribution with a random variance. To prevent the objects in each square from forming a circular shape, we randomly displaced their positions. We next added 10% extra objects randomly placed on the 1200 by 800 area, outside of the squares to represent random noise. Following this, we randomly assigned types, headings and velocities to the objects. We made sure we had exactly the right amount of special points as determined by parameters p_s and p by changing the properties of the objects as needed.

For this test we took a random number of total points n between 50 and 2000 (low and high density) and varied the total number of blue points n_{blue} as a percentage of the total number of points by 5%, 10% and 15%. Additionally we varied the percentage of difference between blue points in the left and in the right square by 5%, 10% and 15%.

We varied the data sets for this task in the same way as the previous task instead that for blue points, we took blue points heading approximately North-East. Since the density visualization does not show direction we only compared M_{part} and M_{single} in the last test. For each of the above test sets we added 10% extra points as random noise. For each variation we generated three test sets leading to a total of $3 \cdot 3 \cdot 3 = 27$ unique data sets for each task which are tested for each visualization: Three times for **T1** and **T2**, and twice for **T3**.

4.5.2 Dynamic test

In our dynamic test we show the subject a dynamic data set running over 2 minutes. The screen is divided into numbered quadrants–see Figure 4.7b–containing moving objects. Each quadrant contains a large number of green objects and a few red and blue objects. We tell the subject that red objects represent ambulances and blue objects represent police vehicles. For adequate emergency response, each quadrant should contain at least one ambulance and one police vehicle. We present the user with the following task:

T4 When a quadrant no longer contains both blue and red objects, press its number.

The response time is measured as the time difference between the key press and the time the requirement was broken.

We generated three types of data sets, with the parameters $(n_{green}, n_{emergency}) \in \{(100, 1), (200, 2), (300, 4)\}$, to represent a calm situation, a medium situation and a busy situation, respectively, where n_{green} is the number of green objects and $n_{emergency}$ is the number of ambulances and police vehicles per quadrant. For each variation we generated three test sets leading to a total of $3 \cdot 3 = 9$ unique data sets which are tested for each visualization.

Data Generation

We generated our dynamic data sets by first simulating the green objects for 2 minutes. We simulated objects by changing their velocity and heading each second using a normally distributed delta. The parameters are chosen such that the simulated movements approximate realistic vessel movements, *e.g.*, objects will not suddenly change speed or direction. The initial positions, headings and velocities of the objects were chosen at random. Whenever the green objects encounter the edge of the screen they are given a new random direction heading away from the edge. We then picked a time instance τ between 10 and 110 seconds and a quadrant in which our requirement must be broken. We next simulated our blue and red objects such that the requirement is broken at τ, remains broken until the end of the simulation and is only broken in the chosen quadrant, ensuring there is only one correct answer. The starting positions of these objects were chosen at random such that each quadrant contains $n_{emergeny}$ blue and red objects at the start.

4.5.3 Questionnaire

A questionnaire is presented to our subjects. Directly after the corresponding task, they answer a question, using a five-point Likert scale, on the suitability of each visualization type for each task Ti. At the end of the user study, we ask the subjects to name a few pros and cons for each visualization, and finally, we ask which visualization was their favorite and why.

4.5.4 Hypotheses

While density maps M_{dens} have been shown to perform well for the visual extraction of vessel movement features in trajectories [218], we expect them to perform worse than both M_{single} and M_{part} on correctness for all four tasks. Also, M_{dens} cannot be used for **T3**, as even though there are methods to incorporate direction in density [120, 179], it cannot be visualized clearly. Furthermore, we expect M_{part} to perform better than M_{single} on correctness, due to the problem of overlap in M_{single}, especially in large data sets. For low density data sets we expect M_{part} to behave as M_{single}, because the latter then also exhibits a low degree of

overlap. Lastly, we do not expect any significant differences in response times between the visualization methods. This leads to the following hypotheses:
*H1: For **T1**, **T2** and **T3**, M_{part} performs (a) better than the other methods both in correctness and response time for high density data sets, and (b) no worse than the others in both correctness and response time for low and medium density data sets.*
We expect that M_{part} performs better than the other visualization because M_{single} suffers from overlap and M_{dens} may be hard to read due to the convolution. This will be especially apparent for data sets with a higher density and thus more overlap.
*H2: For **T4**, M_{part} performs no worse than M_{single} and M_{dens}.*
While we expect M_{part} to perform better for the static data sets, we expect it to perform no worse than the other visualizations for dynamic data sets. Since humans are good at tracking moving objects despite overlap [148], we believe M_{part} will not be significantly better than the other visualization methods.

4.5.5 Results and Discussion

In Figure 4.8 we show four scatter plots in which each dot represents a subject and one of the three visualizations. The axes are mean correctness and mean response time, computed over all relevant trials. We cannot immediately see a trend between visualization methods, but we can see that tests **T1** to **T3** are varying in difficulty, while for **T4** there were few incorrect answers.

In order to determine whether differences in mean correctness or response time between visualization methods are significant, we analyze our results using the one-way ANOVA test and Tukey's HSD post hoc test at a 5% significance level. The latter test results in homogeneous subsets (HS) of visualization methods and is only performed for the tasks where three visualization methods are compared. Between methods in different subsets a significant difference exists, while within a subset, no significant difference exists.

In Figure 4.10 we show the mean correctness with a confidence interval of 95% for the three static tests, separated by the data set size (n) or relative data set size (p_s) at the top, and the percentage difference (p) between the left and the right hand side at the bottom. For the three static tests we can see that the mean correctness often goes up as the percentage difference between the left and right side p increases, as is to be expected. We do not observe such a clear relation for n and **T1** in the top left of Figure 4.10 or p_s and **T2** in the top right of Figure 4.10. For **T3** we see that the mean correctness goes up if the percentage points p_s goes up. So, apparently this task is easier if the proportion of special points increases. An explanation could be that a smaller proportion suffers relatively more from overlap for M_{single}, and is less visible due to smaller glyphs for M_{part}.

The top row tables of Figure 4.9 show the results of the analysis for all combinations (p, n) for **T1**. For the response time we only see a significant difference for a large p where M_{part} is significantly faster than M_{dens}. For correctness we do not see any significant difference between the three visualiza-

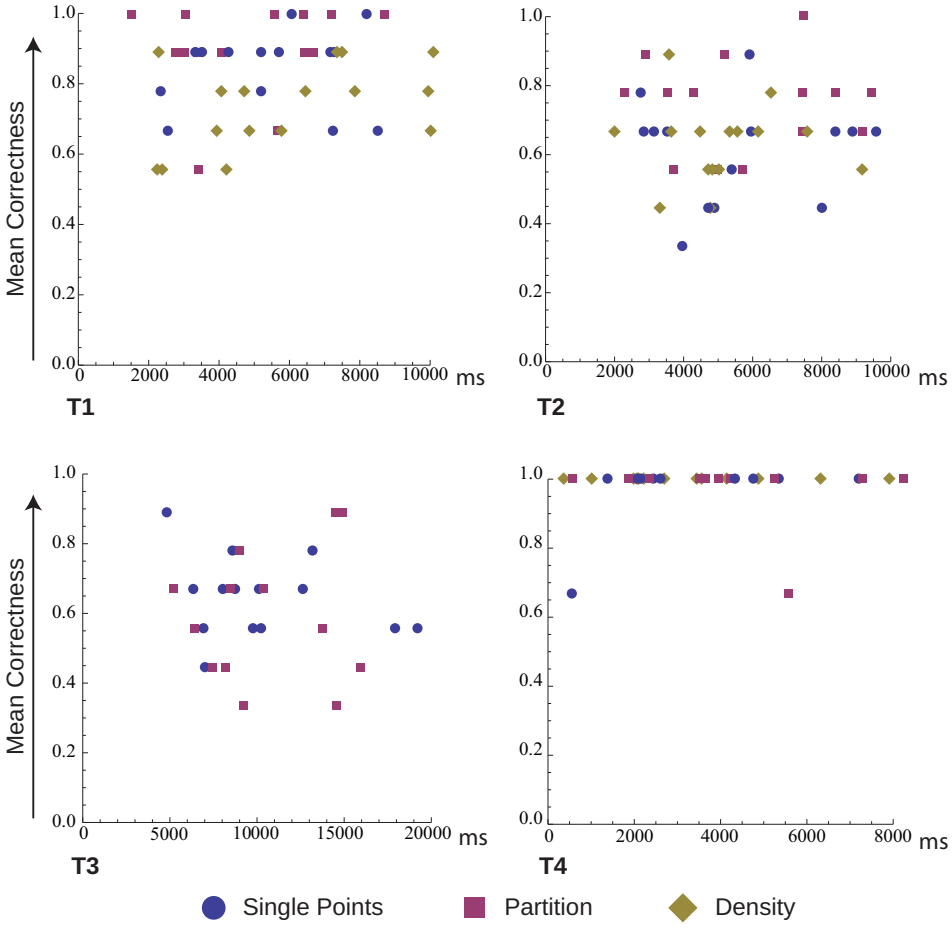


Figure 4.8: A scatter plot of the mean response time versus the mean correctness for each visualization for tasks **Ti** where each point represents a single subject. Each test Ti has 9 unique configurations, resulting in 9 possible values for mean correctness.

tion methods for data sets with a small or medium n. However, for datasets with $(p,n) \in \{(small,large),(medium,large)\}$ M_{part} is significantly better than M_{single} and M_{dens}, respectively. For $(p,n) = (all,large)$ we disregard the percentage difference p and M_{part} is significantly better than both other visualizations.We can therefore not reject hypothesis H1 for correctness, but we must reject H1a for response time.

For $(p,n) = (all,all)$, we find that M_{part} performs better than M_{dens}. From the column $n = large$ in the top-left table in Figure 4.9 we can conclude that M_{part} is overall the best choice. We also observe that M_{part} performs significantly better on correctness than M_{dens} for $(p,n) \in \{(medium,all),(large,all)\}$. Finally, there does not appear to be any clear relationship between the response

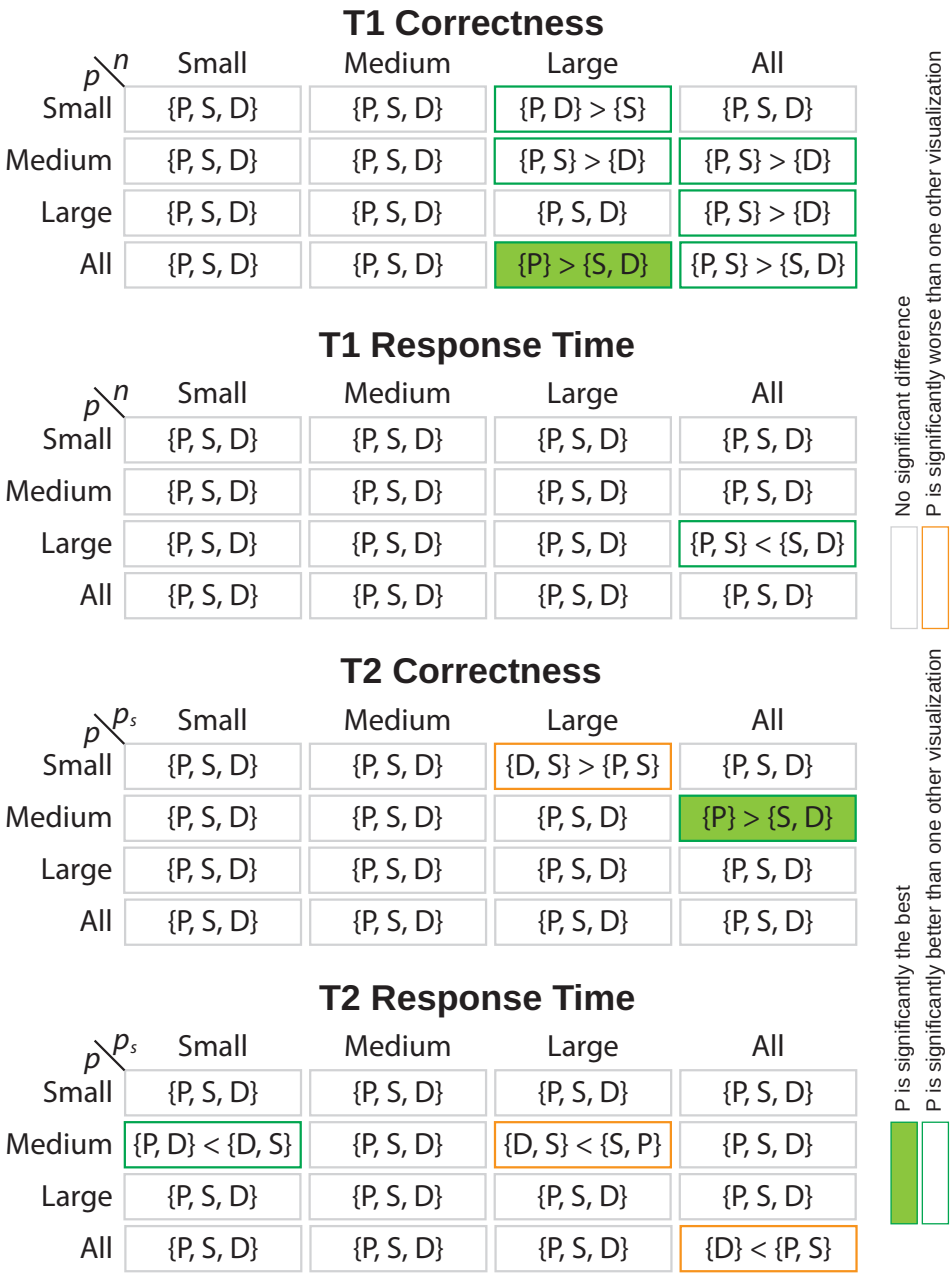


Figure 4.9: The homogeneous subsets for mean correctness and response time for **T1** and **T2** for all combinations of n and p, and p_s and p, respectively. S is M_{single}, P is M_{part}, and D is M_{dens}. Significant results are highlighted. Our method is never significantly the worst.

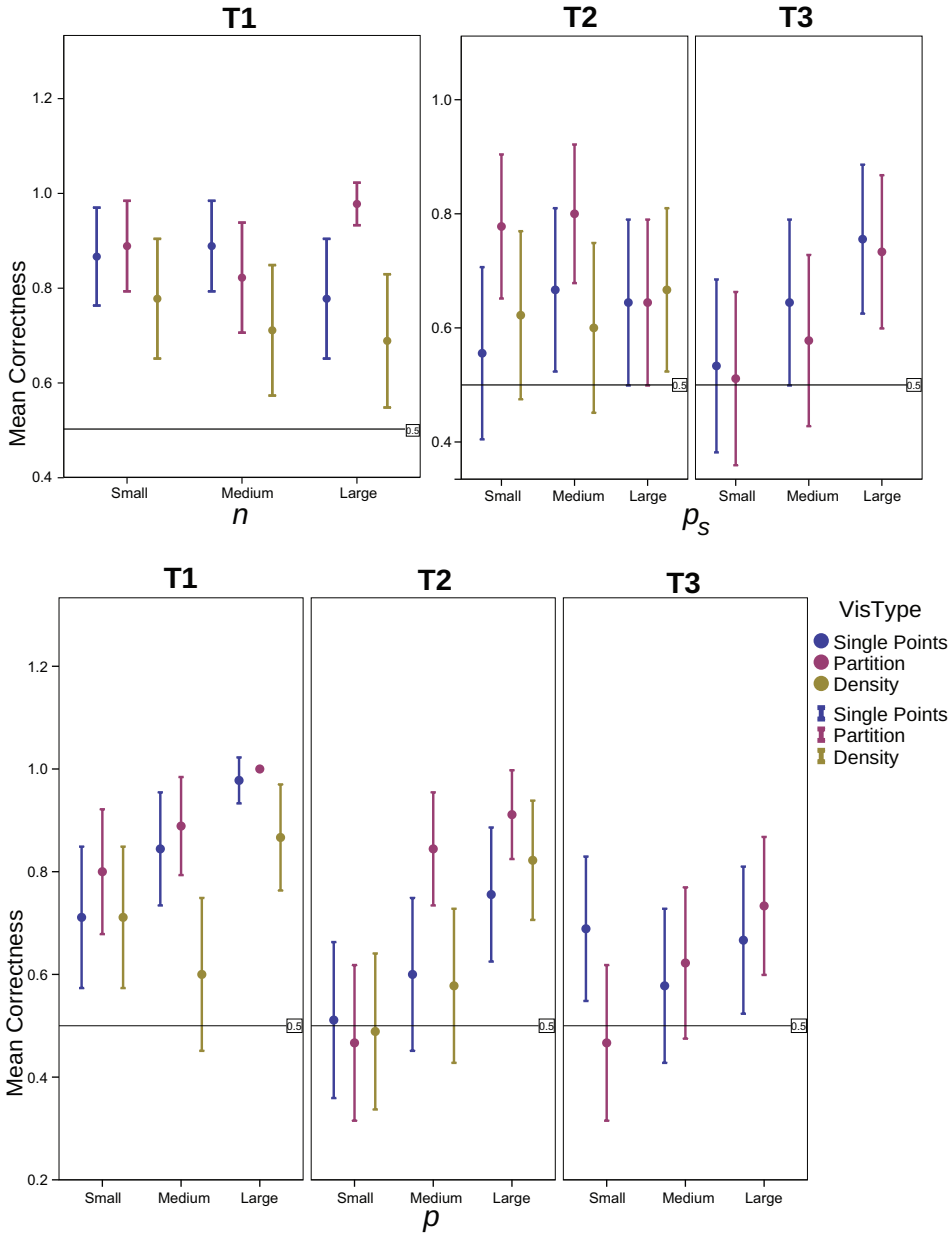


Figure 4.10: Plots of the mean correctness with a confidence interval of 95% are shown for each visualization method for each static test separated by data set size (top) and percentage difference between the left and right hand size (bottom).

time and parameters p or n–see the top right table in Figure 4.9.

For **T2** we find that M_{part} performs significantly better in correctness than M_{single} and M_{dens} for a medium p as shown in Figure 4.9. In contrast, we can also see that M_{dens} performs significantly better in correctness than M_{part} for large p_s with small p. We believe this is in part due to the fact that the mean correctness of M_{dens} for large p_s decreases from large to medium p, but increases again for small p, which is similar to what we observed in **T1**. As can been seen in Figure 4.10, for small p all visualization methods score around 50% correctness for **T2**, which is the expected result for a random guess. We observe that M_{dens} performs significantly better in response time overall than the other two visualization methods.

For **T3** we find no significant difference between the visualization methods for correctness or response time. So, we can conclude that our heading ring does not have the intended effect and requires further research. We can reject H1a and accept H1b for **T3**.

Since almost all trials for **T4** were answered correctly, we can find no significant difference in correctness between the visualization methods. Task **T4** appears to have been easy. Also, we do not detect any significant difference in response time between the methods. This means that M_{part} does not perform significantly worse than the other methods and we can therefore accept H2.

While M_{part} does not perform significantly better across the board on the static tests, we do find that it does perform significantly better under certain circumstances. For comparing densities, M_{part} seems to work especially well for large data set sizes, *i.e.*, in the presence of a lot of overlap. We believe this is because M_{part} has less visual items to compare due to the partition. Additionally, M_{dens} does not work well for estimating the difference in number of vessels, and M_{single} suffers heavily from overlap and may require more visual processing. This does not, however, surface in the response time. For comparing the number of objects of a certain type we can see that M_{part} only performs significantly better than the other visualization methods for medium p.

Questionnaire

In Table 4.1 we show the results of the questionnaire of Section 4.5.3. As before, we use one-way ANOVA to determine whether our results are significant and use Tukey's HSD post hoc test to find HS of the visualization methods. The numerical and statistical results for the questionnaire show that M_{part} is a clear favorite among the participants for all three static tasks, whereas for the dynamic task there is no significant difference between M_{single} and M_{part}. Furthermore, the favorite overall visualization method is M_{part} for 9 participants, M_{single} for 4 participants, and M_{dens} for 1 participant. The remaining participant stated that it depends heavily on the task.

Participants noted that M_{part} is intuitive, makes it easier to estimate the number of points, and offers less clutter and overlap. While participants said that the heading of a group of objects can be more easily seen, some users also found the heading ring complex to interpret. As largest disadvantage of M_{part}, some participants wrote that the animation is too distracting and a small change in

movement can lead to a large change in configuration. We believe this is largely due to the speed of the moving objects during the study. The speed is high on purpose to limit the amount of time our participants have to spend on the tests. In, *e.g.*, the maritime domain the objects generally move over the screen slowly as is shown in the accompanying video. While multiple participants liked the aggregation of all properties, others found it a disadvantage that there are no exact positions, as these are also aggregated. We did not allow any of the interaction discussed in Section 4.4.4 during the user study. The mouse-over view that shows the point distribution may help to see the exact positions of the objects in the group. Furthermore, in some tasks it is more important to see the presence of the object than its exact position, which we believe our method supports.

The most named advantage of M_{single} is the simple and intuitive representation for low densities with objects being countable. Furthermore, a number of participants found moving points easier to follow using this representation. A disadvantage named by almost all participants (14 out of 15) is occlusion. Also, some participants find it hard to compare different groups of objects using M_{single}.

As an advantage of M_{dens}, participants say that it gives an estimate of distribution quickly, especially the distribution of different types. While we can see that the response time of M_{dens} for **T2** is generally lower, it is not significantly so. Also, the reduction of clutter and overlap compared to M_{single} is mentioned. On the other hand, participants find individual moving objects hard to follow and they find it hard to estimate and compare densities. The lack of directional information is also mentioned as a disadvantage.

Task	Vis	n	-2	-1	0	+1	+2	Mean	HS
T1	S	15	2	5	4	4	0	-0.33	$\{P\} > \{S, D\}$
	D	15	3	3	2	7	0	-0.13	
	P	15	0	0	4	10	1	0.8	
T2	S	15	4	5	4	2	0	-0.73	$\{P, D\} > \{S, D\}$
	D	15	2	4	5	4	0	-0.26	
	P	15	0	1	6	8	0	0.47	
T3	S	15	2	4	5	4	0	-0.27	$\{P\} > \{S\}$
	P	15	0	1	2	11	1	0.8	
T4	S	15	0	2	3	8	2	0.67	$\{P, S\} > \{D\}$
	D	15	0	10	2	3	0	-0.47	
	P	15	1	0	4	9	1	0.60	
Total	S	60	8	16	16	18	2	-0.02	$\{P\} > \{S, D\}$
	D	45	4	19	11	11	0	-0.36	
	P	60	1	2	16	38	3	0.67	

Table 4.1: The answers to the questionnaire with 'Strongly Disagree' is -2, 'Disagree' is -1, 'Neutral' is 0, 'Agree' is 1, and 'Strongly Agree' is 2 with their frequency and mean for each test T_i and for each visualization method. S is M_{single}, P is M_{part}, and D is M_{dens}.

4.6 Conclusion & Future Work

We have presented a method to remove visual overlap of the glyph representations of moving objects. The set of all objects is partitioned into subsets, which are visualized with visually non-overlapping aggregated multivariate glyphs that show the distribution of attributes such as heading, type and velocity. The size of the glyph represents the size of the subset. The partition is updated as the objects move such that the change in subset configuration is approximately minimized. Changes in subsets in between frames are smoothly animated. We use examples from the maritime domain, but our method is applicable to any dynamic data with high and non-uniform density. As glyph position, we use a projection of geographic coordinates on the screen, but this can be generalized to any pair of attributes for any time-series data.

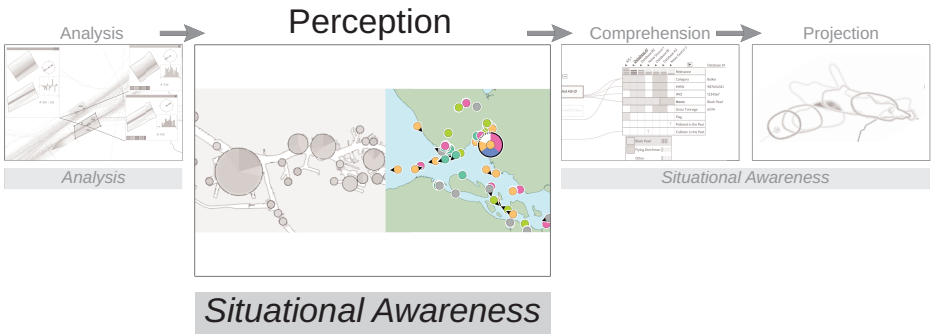
We can conclude from our user study that our method is competitive overall, and better than its competitors for density comparison tasks in the presence of a lot of overlap. Additionally, the participants of the user study have a preference overall for our method.

Based on the results of the user study involving direction and the comments in the questionnaire, we feel that the heading ring needs further research. Additionally, in real-world maritime data sets, especially in harbors, the heading ring becomes hard to read due to the proportionally large number of stationary vessels. Possible solutions could be to increase the overall size of the heading ring, or to only show the heading information after user interaction using a clearer and bigger visualization. Another limitation of our method is that comparing individual items is harder due to the aggregation. This can, however, be solved by introducing additional interaction and multiple selection. The animation also requires additional research. We believe it works for slow moving objects such as in the maritime domain, but users do not seem to like it for fast moving objects. One approach is to investigate a different, less distracting form of animation. Another approach is to introduce conditional overlap: for example, only points moving below a certain velocity threshold are considered overlapping, or, objects moving in opposite direction at a relative high speed are not considered overlapping. Additionally, we would like to test our methods on maritime domain experts using real-world maritime scenarios, especially dynamic scenarios. We would also like to apply our method to dynamic data sets from different domains, especially those without a geographic context. And finally, to be able to interactively handle larger sets of moving objects we intend to implement and parallelize our algorithm on the GPU.

4

5

Design of Glyphs for Uncertain Maritime Data



How can we help an operator in perceiving a situation?

The operator needs to be enabled to perceive the status, attributes, and dynamics of relevant elements in the environment. Through a case study we show how to design a specialized multi-variate glyph in cooperation with domain experts. This glyph shows the attributes and dynamics of elements in the environment and attracts the attention of the operator for elements that are relevant. From a problem definition and requirements derived with domain experts, a design process is started in which a number of design parameters and feasible choices for them are determined. We then go through these options with domain experts to determine the best choices and, finally, we evaluate the chosen glyph with other domain experts.

The contents of this chapter have in part previously appeared in [174].

5.1 Introduction

In the maritime safety and security domain it is important to maintain situational awareness, which means an operator must comprehend the current situation and the projection of the current situation into the future [72]. In current operational systems, situational awareness is established mainly through kinematic properties of vessels and the behavioral models distilled from these properties. Additional information about vessels is obtained through sensors and human observation in increasing quantity [204]. On the, mainly kinematic, information retrieved about vessels, reasoning can be applied that aids in comprehending the current situation [23]. The goal now is to improve the operator's understanding of the situation and confidence in the identity and intents of vessels through visualizing the quality of information using a glyph that represents both the most important attributes including their uncertainty, and the interest of the vessel to an operator. Information is now not only gathered by trusted sensors on board of a vessel, but is also retrieved from external sources with varying trustworthiness. Much of the information required to understand a situation can be gathered from such sources [93]. These sources can vary from trusted databases such as IHS Fairplay [76], to less trusted online databases containing observations by enthusiasts, to news posts or even Twitter messages. The information from these sources is fused and reasoned with [139], resulting in uncertain distributions of possible attribute values, a number of intent hypotheses, and an aggregated *attention* value [146].

In this chapter we present the design process, in cooperation with domain experts, for a new glyph to represent individual vessels. To determine which attributes are relevant for an operator, and especially, which are relevant for visualization in a glyph, we consulted three maritime domain experts with operational experience. Some of these attributes, however, are currently not available to operators and it is therefore hard to determine, even for domain experts, which attributes should be directly visualized in a glyph. After discussing with our domain experts, we have determined that the following attributes should be conveyed by a glyph representing a vessel for any mission:

- Kinematic attributes: position $\mathbf{P}$, heading h, velocity v.
- Vessel type τ for which our domain experts would like to see both the vessel type τ_a , what the vessel claims to be according to the Automatic Identification System (AIS) [111] and the vessel type according to reasoning τ_s, what the system thinks the vessel is, in the form of a distribution of values with probabilities. This allows an operator to quickly see whether a vessel is trying to hide its identity.
- Operator attention value A according to reasoning, which states how much attention an operator should pay to a vessel on a scale of 0 (not important) to 1 (very important).

Currently, vessels are visualized using ship-shaped icons in a geographic information system based on information gathered from AIS. The icons are oriented

along the heading of the represented vessel and are colored according to its AIS type using some standardized color map. We propose to extend this by replacing the standard icons by a parameterizable glyph that can visualize more attributes, including uncertain attributes. First, we explore the visualization design space, and based on visualization and maritime domain knowledge from domain experts, we can dismiss a number of design choices, leaving us with a number of feasible options. Next, we discuss these with several groups of domain experts to arrive at a single design. In this chapter we focus on the design process and the lessons learned during this process that are applicable to a wider range of problems.

5.2 Related Work

Matthews *et al.* [133] designed and evaluated a glyph, to accompany the standard symbology, to represent uncertainty of sensed attributes. These attributes are, however, limited to the uncertainty of identity and spatial location, and to the timeliness of the information.

Maguire *et al.* [131] propose a systematic method to design a glyph by drawing a parallel between categorical hierarchies in the data and the ordering of visual channels. In a similar fashion, we systematically design our glyph based on the structure and semantics of the data.

5.3 Design & Evaluation

With the problem defined, we now have the following requirements for the glyph design after discussion with our domain experts. The glyph should be:

- **Expressive**: Convey the right degree of information;
- **Legible**: Easy to understand and unambiguous;
- **Separable**: Visually separable from both the background and other glyphs;
- **Pop-Out**: Glyphs with a higher operator attention value should be easier to notice for an operator [25];
- **Compact**: Not take up more space than necessary;
- **Consistent**: Have a consistent appearance.

Based on these requirements, we have identified the following visual parameters in the glyph design related to the requirements and the attributes:

- **Glyph Shape**;

Figure 5.1: The final glyph design in a sandbox environment.

- **Background Separation**: How to make glyphs separable to the background and each other;
- **Direction Indicator**: How to visualize direction;
- **Operator Attention**: How to attract operator attention to glyphs that require extra attention;
- **Uncertainty Visualization**: How to visualize uncertain attributes.

We evaluated our glyph design parameters with three pairs of domain experts, with backgrounds varying from development and research in the naval domain to navy operator service. The evaluations had the following structure. At the start of the session the subjects filled out a short questionnaire about their expectations of the glyph before having seen the actual glyph design. Afterwards, the subjects were shown a small presentation to explain the structure of the evaluation sessions, the glyph choices and the glyph parameters that we would be evaluating. The subjects were allowed to ask questions if anything was unclear to them. Following this we showed the glyphs in action in a sandbox environment, see Figure 5.1, where the parameters of the glyph design can be interactively changed. Using this, we systematically went through the glyph parameters. For each parameter we asked the subjects what they believed was the best choice and to thoroughly motivate their answer. The subjects were encouraged to discuss their answers with each other, which we considered might lead to extra insight. When all glyph parameters had been reviewed and the subjects were confident about their answers, the subjects were given another questionnaire with questions about their opinion on the glyph and were asked for additional comments.

In the following, we discuss the design parameters, how we selected the most suitable options, and how our domain experts made their final choices.

5.3.1 Glyph Shape and Direction

We place the vessel type τ_s centrally in the glyph and came up with the designs shown in Figure 5.2a. We eliminate three of them: Sunburst chart for its implicit directional bias; bar chart for its lack of compactness; and autoglyph because it cannot be easily oriented. As shown in Figure 5.2b, we visualize τ_a separately so the two attributes can be visually compared. On the pie-chart based glyph, τ_a is visualized as a smaller, but sufficiently large [186], circle inside the pie-chart, while in the bar chart τ_a can be visualized as an external element or inside the bar chart. Some of our domain experts initially had a strong preference for the stacked bar chart, as it can be oriented and is familiar as an abstract representation of a vessel. After seeing a circular implementation in action, however, all domain experts agreed that the pie chart is preferable due to its rotational invariance which makes it easier to interpret, even though direction is more clearly visible using the oriented bar chart. Another argument given against the ship-analogy is that the colors inside the vessel may be interpreted as an indication of cargo type, whereas an abstract shape does not have this implication.

Displaying the direction on the two chosen representations can be done as shown in Figure 5.2b. The velocity v of a vessel is divided into four classes, which are encoded as shown in Figure 5.2c. With the pie-chart-based glyph already chosen as base shape, all domain experts preferred using the arrow on the outside of the circle to visualize direction and velocity as it gives the most clear and familiar indication of direction.

5.3.2 Uncertain Distribution

The uncertain distribution of τ_s can be visualized in the four ways as shown in Figure 5.2d; (1) by showing all possibilities, which is considered an information overload as low probabilities are considered not relevant; (2) by showing all possibilities above a given probability threshold, which is the preferred method; by showing only the most likely value, in which case the probability can either be visualized as (3) a proportion, or, (4) by reducing the saturation of the visualization, which is considered too subtle, especially considering possible heavy use of color in the background.

The experts preferred to use a to be determined threshold to display the uncertain distribution of τ_s. Only showing the most likely value is considered undesirable as other possibilities may exist that have a probability close to that of the most likely possibility. The option using saturation was considered too subtle, especially considering possible heavy use of color in the background.

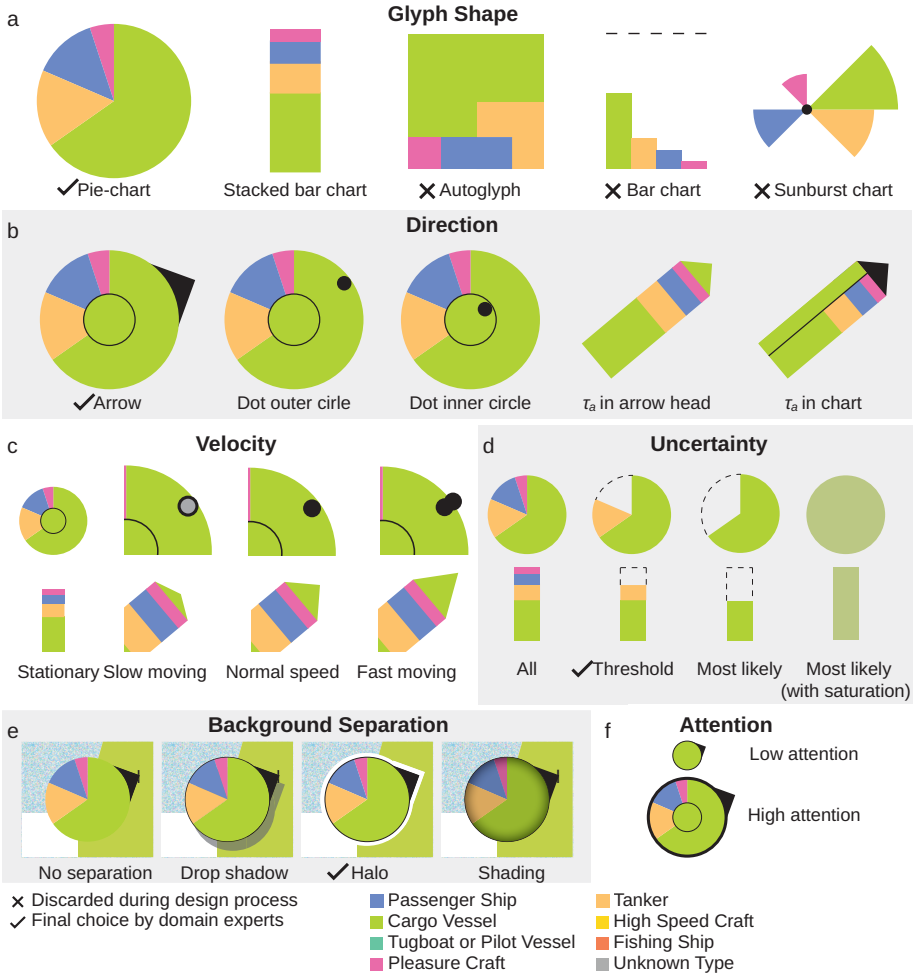


Figure 5.2: A number of ways to visualize direction: On the pie-chart based glyph, with τ_a visualized as a smaller circle inside the pie-chart, as an arrow on the outside of the glyph (a), as a dot on the edge of the inner (b) or outer circle (c), or as orientation for the stacked bar chart where the direction is indicated either using a triangle colored with τ_a (d), or with a black triangle where part of the stacked bar chart is used to indicate τ_A (e).

5.3.3 Background Separation

Since the background may consist of a detailed sea chart and additional colorful and detailed layers, it is important that we separate the glyphs from the background such that they remain distinguishable and legible. As shown in Figure 5.2e, we can introduce a drop shadow combined with a black ring, which according to the domain experts makes determining the position of the vessel ambiguous; a white halo [13] combined with a black ring to separate from both light and dark background colors; or by using shading, which is considered undesirable by our domain experts. For the background separation most domain experts had a strong preference for the halo, while one expert preferred not using any background separation to keep the screen area of the glyph minimal. This expert, however, did agree after seeing the glyph in action that some background separation is necessary. The color of the halo can be any color that contrasts with the black ring, but the experts indicated they preferred a white halo.

5.3.4 Operator Attention

Glyphs of vessels that require more attention from operators, defined by the operator attention value A, should pop out more and should be easy to spot for an operator. The geographic context may already contain color, we use color to visualize vessel type, and we cannot use orientation as a visual channel. Hence, the most suitable visual channels to attract attention to a glyph are size and shape [131, 25]. Since changing the shape of the glyph will distort the information presented within the glyph, we have chosen to use size. Attribute A can be mapped to glyph size linearly, mapped to a small number of different sizes, or a set with the highest A can be shown as large glyphs based on a number or a percentage. Another option is to show a simpler glyph for vessels that do not require attention and a more detailed glyph for vessels that do require attention using size and saliency [77] (Figure 5.2f) to attract attention. The latter is preferred. Arguments against using just size are that the size of the glyphs may be understood to be related to the physical size of the vessel, or to positional uncertainty. When using saliency as well as size, however, these are no longer seen as issues. Additionally, using a single threshold to separate vessels that require attention from vessels that do not require attention is preferred.

5.4 Evaluation

In a later session with a group of other domain experts and operators with active operational experience, we presented the same evaluation. These experts arrived to the same conclusions as the previous domain experts with some additional remarks. Operator trust in the validity of automated reasoning is not considered an issue. While operators of older generations may not readily accept automated methods, younger generations will. Since automated reasoning technologies are

only just emerging in the naval domain, current operators do not know what they want to see, making it hard to define what information needs to be visualized. Also, overlap is considered a problem, especially if larger glyphs are used to visualize vessels that require attention. Finally, the experts miss the information on whether an object is a sea-, land- or aircraft. Currently, we only target seacraft, however, in the future we may solve this issue by using the shape of the glyph.

5.5 Discussion

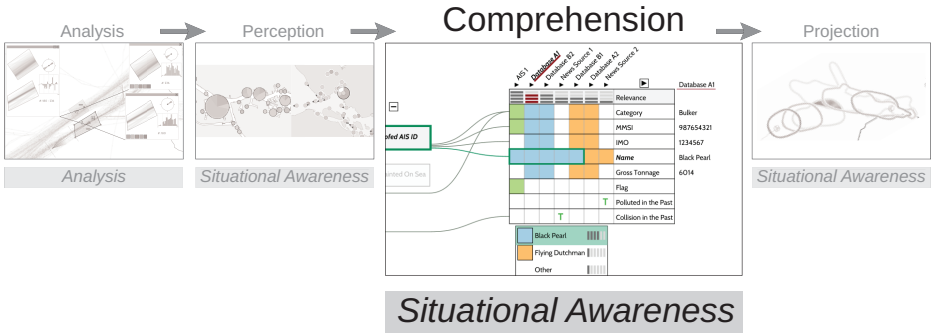
Real-world data often has various attributes of different types with different meanings. These attributes need to be treated independently and differently, depending on their type, semantics, and relative importance. To prevent information overload, continuous attributes, such as in our example velocity or operator attention value, need to be descretized. We took inspiration for the design of the visualizations of individual attributes from existing information visualization solutions. The resulting design choices are, however, not independent, and the effectiveness of each individual design choice is influenced by the other design choices. It is therefore important to be able to visualize and change the design parameters in a real-world environment such that expert users can see the effect of the design choices in all combinations. It is also very important to involve actual end-users, not only in evaluating the design choices, but in making the design choices themselves. We have found that showing our design in a real-world context, allowed our expert users to look beyond traditions and conventions, enabling them to look at the merits of the design. Familiarity is important for domain experts, and it requires visual evidence for these experts to accept new or different ideas. Showing static images of our design did not have this effect.

5.6 Conclusions & Future Work

We have demonstrated the design and evaluation process of a multivariate glyph with uncertainty using a case from the maritime domain. The pitfalls and lessons learned from such a process are discussed. In future work we would like to expand the glyphs to be able to represent additional, mission-specific information. Additionally, we would like to apply our design techniques to a problem in another domain.

6

Rationale Visualization for Safety and Security



How can we help an operator in comprehending a situation?

The maritime environment can contain a large number of vessels. Due to the number of vessels and the amount of information that can be found on these vessels, it is not feasible for the operator to reason on each of these vessels to determine whether they are relevant and require attention. This is where automated methods, such as automated reasoning, can help the operator. These automated reasoners, however, typically only output a conclusion and this conclusion alone does not help the operator in comprehending the situation. In this chapter we present a method to visually explain the rationale of a reasoning engine that raises an alarm if a certain situation is reached. Based both on evidence from heterogeneous and possibly unreliable sources, and on a domain specific reasoning structure, this engine concludes with a certain probability that, e.g., the vessel is suspected of smuggling. To support decision making, we visualize the rationale, an abstraction of the complicated reasoning structure. The evidence is displayed in a color-coded matrix that easily reveals if and where observations contradict.

The contents of this chapter have in part previously appeared in [171, 172].

A video on the contents of this chapter can be found at: https://youtu.be/8LlJ_er-1LU.

6.1 Introduction

Safety and security tasks, such as law enforcement or emergency management, require continuous monitoring of objects of interest (OOI). Examples of such objects are people, vessels, and transactions. In a decision support system, a reasoning engine may reason about these OOIs and their context, using information gathered from multiple, heterogeneous, and possibly unreliable sources [98]. These sources may vary from trusted databases to unreliable news articles and twitter messages, and from automated sensors to human observations. Based on a specific domain model, the information is fused [137] and reasoned with. The system will raise an alarm whenever the reasoning engine decides certain conditions have been met, *i.e.*, the probability that some task-related hypothesis is true is above a set threshold. As data becomes increasingly available and complex, the need for such automated methods, and especially, automated reasoning rises. Automated reasoning methods, however, are often monolithic black boxes, where data goes in and a hypothesis with a probability for its validity comes out, leaving the user to draw conclusions while guessing the rationale behind the reasoning. Especially when these conclusions are required for high-cost decision-making, this poses a problem. Trust in the system, an understanding of the situation, and also acceptance of the results are essential to decision making [29, 130, 164]. Explaining the reasoning process has been shown to be beneficial for decision-making [57, 188, 164]. Therefore, we propose to visualize the rationale of a reasoning engine. Our visualization shows why an alarm was raised or not, and based on what evidence the reasoning engine came to its conclusion. This can then be used to confidently take appropriate action–see Figure 6.1.

Our reasoning engine is based on a *first-order probabilistic logic* model, which allows to define joint probability distributions, modeling the uncertain relations between the objects in the domain of interest. The structure of the model expresses the domain knowledge in a natural way [98]. A model can be used to compute probabilities of arbitrary statements, given information known in the current situation, called *evidence*. Such a reasoning approach is similar to widely used *Bayesian Networks* [150], but more general due to the first-order nature of the used language. The increased expressive power is necessary for safety and security tasks, as they involve a dynamic number of objects and dynamic amounts of information. A more detailed description of this model as used in the Maritime Safety and Security Domain (MSSD) is given by Michels *et al.* [138].

Our contribution is as follows: we introduce a novel way to visualize agreement and contradiction between sources in heterogeneous, multivariate data. By combining this with a graph representation of the reasoning structure, we visualize the rationale of a reasoning engine based on first-order logic for the (maritime) safety and security domain. To explain our visualization method, we use a hypothetical problem scenario that is easy to understand. Further on in this chapter we discuss case studies from the MSSD and evaluate with experts from that domain. During the design process, these domain experts have been continuously involved and consulted. However, we expect our method also to be applicable to

other domains that use decision-support systems, such as law enforcement, medical diagnosis, machine diagnostics, and financial and identity fraud investigation.

6.1.1 Background

In the MSSD, whether on board of a vessel or in land-based stations, operators are monitoring vessels. The operators are monitoring for suspicious behavior, which is defined by the mission they are on in an area of interest that may contain hundreds of vessels. Threats the operator is looking for may include illegal activities such as smuggling, environmental degradation, illegal fishing, human trafficking, piracy, and terrorism [180, 45]. As a response to such threats, operators may need to deploy resources that are sparse or costly with potential risk of life, such as sending a vessel to intercept or search the vessel [45]. Vessels are monitored using the Automated Identification System (AIS) [111] augmented by radar.

Traditionally, automated reasoning or analysis is performed mostly on kinematic attributes of the vessels [173]. Decisions and reasoning on the intents of the vessels are based on intuition and operational experience [45]. The trend, however, is that more and more additional information from additional sources is being used [93]. Our system works as a part of a broader safety and security system [93, 98] intended to support law enforcement agencies on their missions. For each vessel, large amounts of information can be gathered from distributed and homogeneous sources, from trusted sources such as official databases, analyzers that detect suspicious behavior in kinematic attributes such as rendezvous [180] or loitering in a forbidden area, or less trusted sources such as unstructured news sources [93, 206], or even Twitter messages. In general, knowledgeable users can do the reasoning themselves, but for large amounts of information and large numbers of OOIs this becomes infeasible and impractical and it is best to automate the process. This, however, requires presentation of an automated reasoning process to enable the user to understand the process [142, 164]. Using resource management and reconfiguration [146], the reasoning engine reasons on the intents of all vessels present and raises an alarm to the operator whenever a particular vessel requires attention in the context of the current mission. To spare resources and minimize risk, operators need to be confident in their decision before taking action [164]. The operator can use our visualization to gain understanding of the situation and the conclusions of the automated reasoning, and take better informed decisions with confidence–see Figure 6.1.

After discussions with potential users and domain experts, mainly from the MSSD, and reasoning experts, we formulated the following requirements:

R1 The user should be enabled to see anomalies;

R2 Hypotheses requiring attention should pop out visually;

R3 Only hypotheses relevant for the current situation should be shown;

R4 The visualization layout should be stable;

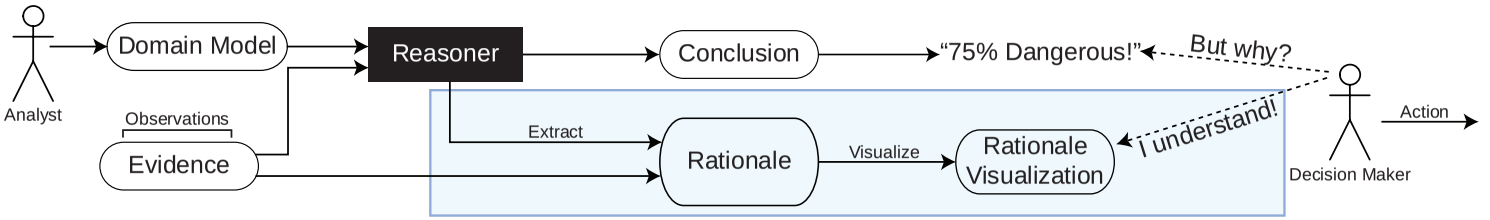


Figure 6.1: Evidence is reasoned upon by an automated reasoner, a monolithic black box based on a domain model, which supplies a conclusion with a certain probability. This is, however, not sufficient for a decision-maker to take decisions confidently. We therefore propose to extract an abstraction of the reasoning engine, and visualize this such that the decision-maker can understand the reasoning process. Our contribution is highlighted in blue.

R5 The user should be enabled to see and identify agreeing or conflicting observations quickly;

R6 If this information is available, the user should be able to see the timing of the observations.

A user will generally see the rationale visualization of many different OOIs. The user has to be enabled to quickly see what is important in the current situation (**R1**, **R2**), and not be distracted by irrelevant information (**R3**). Furthermore, to preserve the mental map [59] of the user, we require the layout of the visualization to remain stable (**R4**), enabling faster visual processing. For example, in the MSSD, each mission has a single main hypothesis, therefore the reasoning structure will remain highly similar for all OOIs. Lastly, to understand the source of possible conflicts or uncertainty, the user has to be enabled to find and identify these conflicts (**R5**, **R6**). For example, an observation, 10 years ago, relating a vessel of interest to criminal activity may be of less value than one of only several months old.

The reasoning engine works based on a domain model created by domain experts [98]. This model can be validated using our visualization. Doing so has proven valuable, *e.g.*, for the models used in this chapter. Note that the decision-maker, *i.e.*, the operator, is a user of the reasoning engine, but is not expected to modify the domain model.

6.2 Related Work

Visual representation of both evidence and reasoning used by an expert system may be used to enhance trust in its rationale and as such can improve decision making of the decision support system it is in.

Evidence consists of a set of agreeing or conflicting observations, *i.e.*, multivariate tuples of attributes with heterogeneous, including nominal, data and possibly missing values. Multivariate data visualization is a heavily researched subject in Information Visualization [185], especially for large numbers of tuples, which we do not have in our situation. Here we relate our work to some key visualizations. Scatterplot matrices [43] may be used to do cluster analysis to show the degree of conflict between tuples; the plots, however, suffer from data overlap, and are, even for small numbers of attributes, not compact. Using parallel coordinates [106] instead is also not compact and missing values are problematic. Aggregated views of data, like in (stacked) histograms and heat maps, are also unwanted, since both individual tuples and attributes are of interest. Diversity maps [151] do visualize the diversity in a set of tuples, but not between tuples since they also use aggregation. Multidimensional icons like Chernoff faces [40] and Star glyphs [213] explicitly map attribute values, which results in an unnecessary cognitive and perceptual load when checking only for conflict and agreement. The same problem holds for matrix visualizations [221], however, less so if they are properly sorted. Furthermore, they can handle missing values, are

realtively compact and do not aggregate values. Our visualization for evidence extends them by more explicitly showing conflict and agreement.

A *reasoning* typically forms a directed acyclic graph (DAG) which may be directly visualized. Argument mapping, for instance, is a visual aid to show how the premises and the structure of an argument are used to reach a conclusion in which both opposing and supporting premises are shown; see, for instance, [202]. In contrast to our explanation graphs, uncertainty plays no role in argument mapping. In the reasoning diagrams of Pike *et al.* [152] uncertainty plays a role. In these diagrams the emerging knowledge of users during interactive sessions is stored as nodes with data, visualizations, and confidence levels; the nodes are connected with directed edges containing evidence and a support level. Confidence levels and support levels are editable by the user, and after propagation give confidence levels for alternative hypotheses. Given a mission our system uses a static graph with a single main hypothesis, confidence levels in the nodes, and support levels on the edges; these levels are all computed based on the available evidence, which in our case is attached to nodes.

Visual explanation of the reasoning of expert systems is not unique to our system. Madigan *et al.* [130] already present such explanations modeled by, so-called, belief networks. Similarly to our explanation graphs, these networks are also DAGs. Unlike our networks these networks are complete (large) models and as such not useful for a clear and compact explanation. Šutovský and Cooper [208] present a hierarchical explanation method for a Bayesian network with a large number of agents, sources of observations. Their approach is similar to ours in the sense that they have a hierarchy with hypotheses, which are connected to the relevant observations. In their case all observations have the same single attribute; hence, they have no need to show the level of conflict between observations and simply group agents with the same attribute value.

6.3 Rationale

The actual reasoning structure is a complicated structure of interdependent hypotheses and too complex for a normal user to understand. For example, the reasoning structure behind Figure 6.7 consists of over 300 nodes, fully connected, and may grow depending, in part, on the amount of input information used. Therefore, to make it comprehensible we create an abstraction (see Figure 6.1) by extracting the most important hypotheses and connections. Selecting important nodes and grouping nodes into hypotheses is done manually by domain experts as the abstraction depends on the domain's needs and expectations. The connections and their weights can then be computed automatically. According to our experts, we can assume that working with an abstraction is fine, since an operator of a decision-support system is considered knowledgeable in the domain. This abstraction contributes to the *rationale*, which consists of the following:

- A directed acyclic graph (DAG), the *explanation graph*, in which each node

represents a hypothesis, and each edge the dynamic influence one hypothesis has on another;

- The evidence, a set of observations, where each observation is a tuple of attribute values that are, by definition, about the same object;
- Relations between the hypotheses and the observation attributes they depend on.

To explain our method, we use a simple example scenario. In Section 6.5 we demonstrate our visualizations using a more sophisticated, real world scenario in the MSSD. In our example scenario we look at a fictitious customs office where the OOIs are individuals crossing a border. The customs officer is using a reasoning system that, based on a set of observations, raises an alarm if the individual may be involved in criminal activity. The officer can then use our visualization to understand why and take further action if needed. The observations in this scenario come from the following sources:

- Direct observation by the officer;
- A passport check;
- An eye witness report (tip) that someone is involved in a criminal activity;
- A database containing criminal convictions.

The observations may contain one or more of the following attributes:

- Height;
- Sex;
- Individual is a child;
- Passport ID;
- Name;
- Individual has been seen involved in criminal activity by an eye witness;
- Individual has a criminal record.

The explanation graph for this scenario is shown in Figure 6.2. The main hypothesis, *criminal activity* (h_0), depends on whether the individual is trying to hide their identity (h_1), and whether there is past criminal activity (h_2). If the individual is an adult (h_3), it is considered more likely they are involved in criminal activity than if the individual is a child.

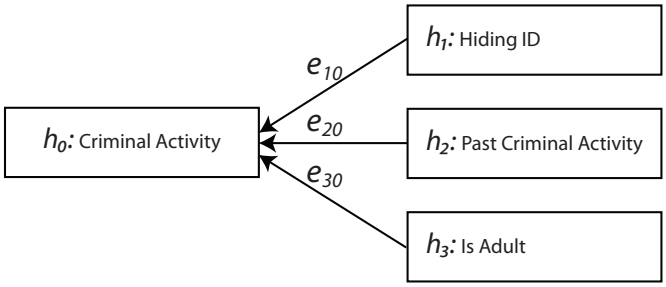


Figure 6.2: A schematic representation of the reasoning abstraction of our toy problem. The main hypothesis, h_0: *Criminal Activity*, states that the person crossing a border at a customs office is involved in criminal activity. This is based on whether the individual is trying to hide their identity (h_1), has past criminal activity (h_2), and is an adult (h_3).

6.3.1 Explanation Graph

Each node h_i in the graph represents a hypothesis which can be supported by observations and/or other hypotheses, for example *Criminal Activity* h_0 is supported by *Hiding ID* h_1: If individuals are trying to hide their identity, it is more likely they may be involved in criminal activity. Each hypothesis is formulated to either represent a situation that requires attention or to support another hypothesis that does. In the example scenario in Figure 6.2, reasoning about, *e.g.*, the hypothesis *hiding ID* involves a complex reasoning process, comparing the ID the person pretends to have with the other known information about the ID, of which a dynamic amount with varying reliability is available. This complex process is abstracted to a hypothesis that can be understood by the user. The nodes are connected by directed edges e_{ij} from node h_i (child) to node h_j (parent), where the hypothesis of node h_i is understood to support that of node h_j. The explanation graph has a single root h_0, which is the main hypothesis; in our example scenario, the individual is involved in criminal activity, or as in the example in Figure 6.7, the vessel is behaving recklessly.

For each node h_i, the probability p_i^e that the hypothesis is valid given the evidence currently available is computed. The prior probability p_i^{prior} is also given, which is the probability of the statement without any evidence. For example, the prior probability of hypothesis h_3 in Figure 6.2 is the chance that any given person crossing the border is an adult. Furthermore, each node has a short descriptive label. Each edge e_{ij} has a dynamic weight $w_{ij} \in [-1, 1]$ that signifies the influence, based on the dependencies, a hypothesis has on the validity of its parent hypothesis. The influence of one hypothesis on a parent does not only depend on how likely the hypothesis is, but also depends on the probabilities of other hypotheses. Therefore, the weights are dynamically computed to reflect that. For example in Figure 6.2, in general a high probability that an individual is trying to hide their id (h_1) increases the probability that the individual is involved in criminal activity (h_0). However, in case the individual is not an adult (h_2), it

is less likely they are involved in criminal activity, which means the influence of h_1 becomes lower.

6.3.2 Evidence

The evidence consists of a set of observations. An observation is a tuple consisting of a number of attributes, which are used to support hypotheses. These attributes can be of varying types, *e.g.*, a name, a number, a classification, etc. Multiple observations can contain the same attribute, but their actual values may differ or may be absent. Attributes may have completely different domains, *e.g.*, from strings to categorical values to continuous numerical values, and these domains are considered independent. Each attribute value has a probability that it is the actual value. Each observation o has a probability p_o^{OOI} which is the confidence of the reasoning engine that the observation is related to the OOI. Also, each observation o may have a date and time when it was recorded. Typically, the size of the evidence is in the order of 10 observations by 10 attributes.

6.4 Visualization

We aim to visualize the rationale such that the operator can understand and follow the reasoning that leads to the conclusion drawn by the reasoning engine. We show the explanation graph in a structured layout, which allows the user to quickly identify which hypotheses are relevant and which paths in the explanation graph lead to the main hypothesis. The hypotheses are connected to attributes that are visualized in an evidence visualization that allows the user to quickly find and identify agreement or contradiction between different observations.

We enable the user to explore the structure of the graph and evidence interactively (see the supplemental video) through highlighting as shown in Figure 6.7. The user can also change the order of the observations by clicking on one of the arrows above the matrix–see Figure 6.4. To return to the initial ordering, based on p^{OOI}, the user can press the arrow button above the attribute labels.

We have chosen a horizontal reading direction as it is more natural for our users and more space efficient. Our visualization can be seen as an explanation of the reasoning behind a conclusion, and as such would be read from the root hypothesis to the evidence visualization. It can also, however, be seen as a reasoning aid to help the user reach the conclusion, and as such would be read from the evidence matrix to the root hypothesis. We have chosen to use a horizontal left to right direction, from hypotheses to evidence.

6.4.1 Explanation Graph

We visualize the explanation graph by drawing boxes for nodes and curved lines with arrow heads for the directed edges. To layout the graph, we use a variation

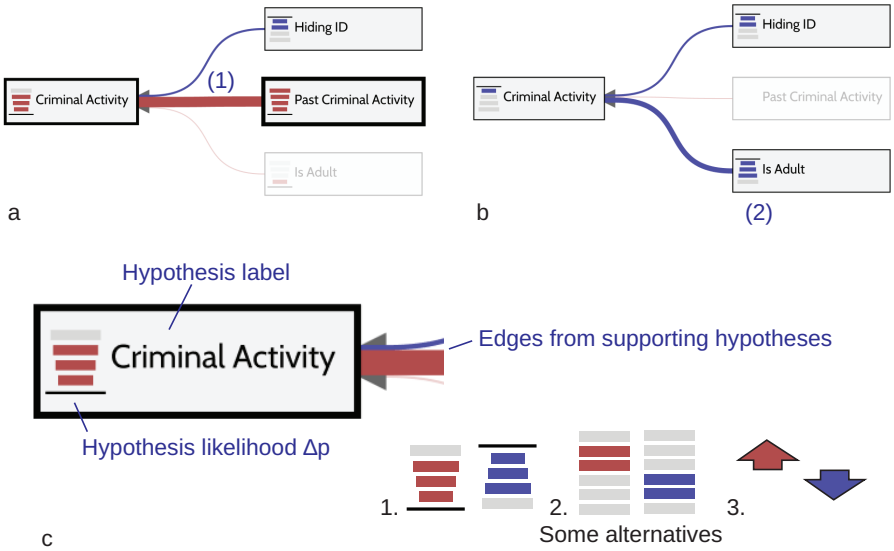


Figure 6.3: (a) A rationale visualization of a scenario in which there is a rise in probability that the OOI is involved in criminal activity. Because edge (1) is thick and red, we can see this is mainly due to a history of criminal activity. (b) There is a lowered probability of criminal activity. Because edge (2) is thick and blue, we can see this is mainly because the OOI is not an adult. In our toy model, we have made the assumption that a child is less likely to be involved in criminal activity. (c) A close-up of a hypothesis node, with some of the alternatives we considered for the probability indicator.

6

of the Sugiyama layout [189], where nodes are layered based on the shortest path to the root node. Within a layer, the nodes are ordered vertically to minimize edge crossings and to group nodes with the same parent–see Figure 6.7.

In an explanation of the main hypothesis for a given OOI, not all hypotheses are relevant. To make the explanation graph easier to understand (**R3**), we show only the sub hypotheses that are relevant for explaining the main hypothesis. For example, in Figure 6.3b the *Past Criminal Activity* hypothesis is not relevant since there is no evidence to support or refute it. We say a hypothesis h_k is not required to explain the main hypothesis h_0 if all paths from h_k to h_0 contain at least one edge e_{ij} with $|w_{ij}| < \epsilon$, where ϵ is some threshold. To preserve the mental map of the user and to satisfy **R4**, we use a fixed layout for the whole graph and for readability (**R3**) suppress irrelevant nodes to the background–see Figure 6.3a.

We assume that a user has domain knowledge and therefore restrict ourselves to showing the deviation from normal situations by using the difference of probabilities $\Delta p_i = p_i^e - p_i^{prior}$. This allows us to more easily show anomalies, *i.e.*, situations that are different from normal and may require attention (**R1**). For example, a probability of 10% that someone is a criminal might require attention, since for the average person the probability may be less than 1%. The hypotheses are formulated such that $\Delta p_i > 0$ means that more attention is required, while

$\Delta p_i < 0$ means no special attention is required in the context of the main hypothesis. Each node i is visualized as a box containing its label. If $|\Delta p_i|$ is above a threshold ϕ, its probability Δp_i is visualized in the left side of the box using an indicator with colored rectangular marks as shown in Figure 6.3c. The marks are filled to show the magnitude of p_i with unused marks shown in gray. For $\Delta p_i > 0$, we color the marks from bottom to top with red, and draw a line below the marks to indicate the zero line. The marks are stretched towards the top such that the higher Δp_i, the more visual impact the indicator has (**R2**)–see Figure 6.3c. For $\Delta p_i < 0$, we reverse the direction and color the rectangles blue. Initially, we tried using an arrow to point towards the direction of change, but our users found it too hard to read (Figure 6.3c). Therefore, we tried to use boxes similar to a bidirectional volume indicator (Figure 6.3c). The unused boxes, however, take up too much space, therefore, we decided to visualize a single direction with the zero line to emphasize directionality. We use a color scheme red (hot) and blue (cold) to, respectively, signify elements requiring attention (a dangerous or anomalous situation) and elements that do not. As a double encoding, hypotheses that require more attention, *i.e.*, $\Delta p_i > \phi$, are visualized with a thicker border proportional to Δp_i (**R2**).

The edges between hypotheses are visualized as colored, curved lines. Here we use the same color scheme as before, where red signifies a strengthening influence, and blue signifies a suppressing influence. The thickness of the edge is determined by $|w_{ij}|$ (see Figure 6.3a) (**R2**). An edge from a suppressed hypothesis is also suppressed–see Figure 6.3a (**R3**). We draw an arrow head at the end of the edges to emphasize the direction of the edges. Faded edges are rendered below the arrow heads, while active edges are rendered above.

Once the user is familiar with the structure of the explanation graph, the labels of the hypotheses can be deactivated by clicking on the contract/expand button above each layer, as shown in Figure 6.5. To support scalability, nodes may need to be contracted by default, or may need an even more compact representation for larger explanation graphs.

6.4.2 Evidence

We visualize the evidence in a matrix, where each column represents an observation o, and each row a unique attribute a–see Figure 6.4a. The confidence that an observation o is about the OOI, p_o^{OOI}, is visualized as a header at the top of the evidence matrix and labeled "Relevance". We use a gray indicator with rectangles to visualize the confidence p_o^{OOI} for each observation o. The observations are initially ordered based on this confidence from high (left) to low (right).

When hovering over a cell (a, o) in the evidence matrix, all attribute values of observation o are shown to the right of the matrix. At the same time, all unique values for attribute a are shown below the matrix, as shown in Figure 6.4b. Next to the attribute values, the confidence in the attribute value is visualized using an indicator with filled rectangular marks.

Since the evidence may contain many attribute values, from possibly con-

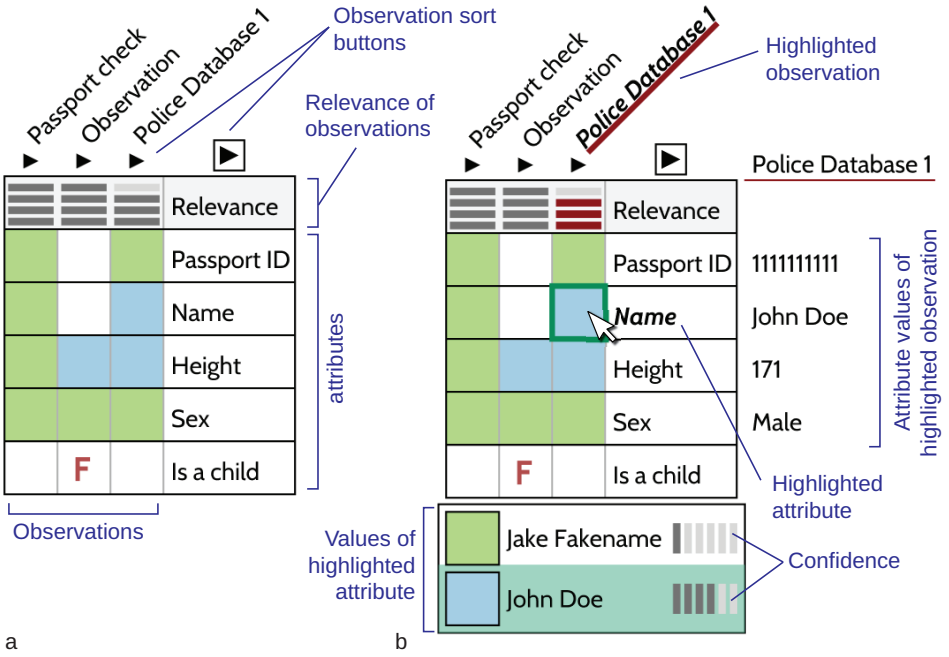


Figure 6.4: (a) The evidence matrix where attribute values are colored-coded to show contradiction or agreement between observations. (b) A user can investigate the evidence by hovering the mouse over the matrix.

tradicting observations, simply displaying the values using text in a table of attributes versus observations will not enable the user to quickly see the degree of agreement or contradiction between observations [136]. Also, since each attribute represents an independent and possibly completely different domain, we cannot map a pre-defined color map to attribute domains. Therefore, we have chosen to use color to emphasize agreement and disagreement in the evidence.

The cells (a, o) in each row are colored such that each attribute value receives a unique coloring. This allows the user to quickly see on which attributes observations agree and for which attributes there is a conflict (**R5**). Boolean attributes are visualized with a green T or a red F. Continuous attributes may be clustered or mapped to a discrete set of values by the reasoning engine based on some predefined domain-based rules. In Figure 6.4a we show an evidence matrix in which all observations agree on the passport ID and sex of the OOI, but disagree on the name and height. Cells are colored according to the following requirements:

- The total number of different colors K is minimal;
- For each row, two cells have the same color if and only if they have the same value;
- Cells with no value have no color;

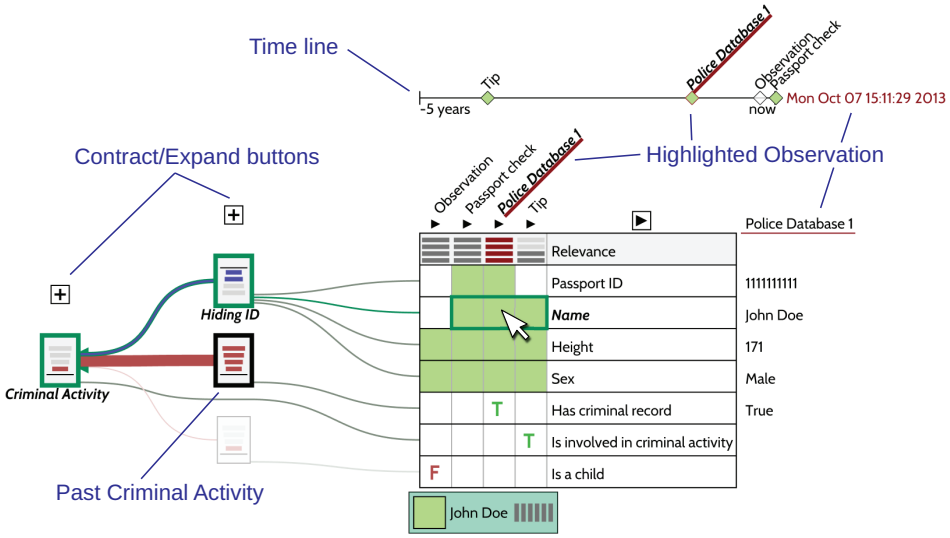


Figure 6.5: We have a slightly risen probability that the OOI is involved in criminal activity. This appears to be mainly due to past criminal activity, and an eye witness report with the OOI's descriptions. In the time line, however, the tip appears to be old, and even predates the criminal record. This is reflected by the lowered relevance of the eye witness report, and the only slight rise in probability of *Criminal Activity*.

- For each column, the number of colors should be minimal.

We minimize the number of different colors in a column to reduce visual clutter and to make it easier for the user to see contradictions between observations as color differences in the horizontal direction. From the requirements it follows that K is exactly equal to the maximum number of different attribute values on any attribute. We use a greedy heuristic to color the matrix. For each row a, we call cells with identical attribute values a group. For each column o, we maintain a set U_o with colors already used in the column. Given a coloring where group g is not yet assigned a color, we define the total cost $C(g,k)$ for all cells (a,o) in group g with color k as

$$C(g,k) = \sum_{(a,o)\in g} [k \notin U_o], \tag{6.1}$$

where $[X]$ is $\{1,0\}$ if X is $\{\text{True}, \text{False}\}$ respectively.

Initially, we can already color one of the rows with the highest number of groups. Since U_o, for each column o, is still empty, the cost of coloring the initial row is 0. Until all groups are colored, we keep on finding a group g and a color

k for which the cost $C(g,k)$ is minimal and coloring g with k. The sets U_o are updated where necessary. See Algorithm 2.

We use a ColorBrewer [89] color map to color the cells. We have removed the red color from the color map, since our users interpreted the combination of red and green as bad and good, which attracted undue attention to red cells. Without red, no additional meaning appears to be given to the colors themselves. We abstained from using blue as default color, as our users confused a fully blue evidence matrix with the blue hypothesis indicators. To make it clear that the colors are related in the horizontal direction and not in the vertical direction, we separate the rows using black lines and separate the columns using gray lines.

6.4.3 Time line

The relevance of information may depend on its age, *e.g.*, an eyewitness report of criminal activity of 20 years ago is less relevant than an eyewitness report of one week ago. Therefore, we also visualize a time line of the observations if such information is available (**R6**). In Figure 6.5, we see that the eye witness report is almost 5 years old and predates the criminal record from the police database. Hence, the relevance of the eye witness report is lower.

We visualize a time line simply as a line, with diamond shaped markers for each observation. If an attribute is highlighted, each observation indicator in the time line is colored according to the attribute value of the respective observation (see Figure 6.5).

6.4.4 Attribute connections

Attributes (rows in the matrix) are connected to the hypotheses they influence with curved, gray lines. We have chosen to layout the attributes on the rows instead of on the columns such that they can be more easily linked visually to the hypotheses using lines. These lines are faded if they connect to a suppressed hypothesis. To avoid visual clutter, they are bundled and routed around nodes where needed–see Figure 6.7. The lines are sorted to minimize overlap and a gap is left between bundled lines such that individual lines can still be distinguished

Algorithm 2 Matrix Coloring

```
Let r be a row with K groups
Color row r using K colors
while Exists uncolored group do
    Find uncolored group g and color k,
        such that C(g,k) is minimal
    Color g with k
    for all (a,o) ∈ g do
        U_o ← U_o ∪{k}
```

and followed. To separate attribute lines from edges between hypotheses, the attribute lines are connected to the bottom right of the hypothesis boxes as shown in Figure 6.7.

6.5 Results

In this section we demonstrate our method using some realistic use cases from the MSSD based on real data. The data is replayed in a functioning monitoring and decision support system and interesting cases have been selected for evaluation with domain experts.

6.5.1 Evaluation

We demonstrated our method at a convention, where the main attendees were potential users such as law enforcement agencies, harbor agencies, cities, and banking analysts, using example use cases from the MSSD, including the use cases discussed below. We were happily surprised that we did not have to explain the visualization before the attendees could tell us what they (correctly) thought was going on by looking at our visualizations.

Additionally, we consulted two independent groups of four domain experts and operational experts to evaluate how well our visualizations are understood by potential users. These groups did not include the domain experts we consulted while developing our visualization method. In these two sessions we gave a small introduction on our methods and visualization. Following this, we presented four use cases using real data that would require operator attention and asked the experts why an alarm was raised given our visualization. We describe these scenarios and the results and comments from the experts in the following subsections, and we further discuss our results in Section 6.6. All sources and all vessel information have been anonymized.

6.5.2 Use Case 1: Environmental Hazard

In this use case, the operator is looking out for vessels that may pose an environmental hazard. A vessel in a particular area has triggered an alarm due to a rise in the probability that it poses an environmental hazard. The operator is investigating why the alarm has been triggered, and whether action is required–see Figure 6.6. A group of hypotheses at the bottom requiring attention is what jumps out immediately for the experts. The root hypothesis, *Environmental Hazard*, has a rise in probability. We can see this is mainly due to *Restricted Area Violation*, as shown by edge (1). The rise in *Restricted Area Violation* is supported by the hypotheses that the vessel is a tanker, is above $10,000$ tons, and is inside a restricted area (2). Due to the red coloring and thick borders we can immediately see that this rule is being violated. This is related to an environmental protection rule

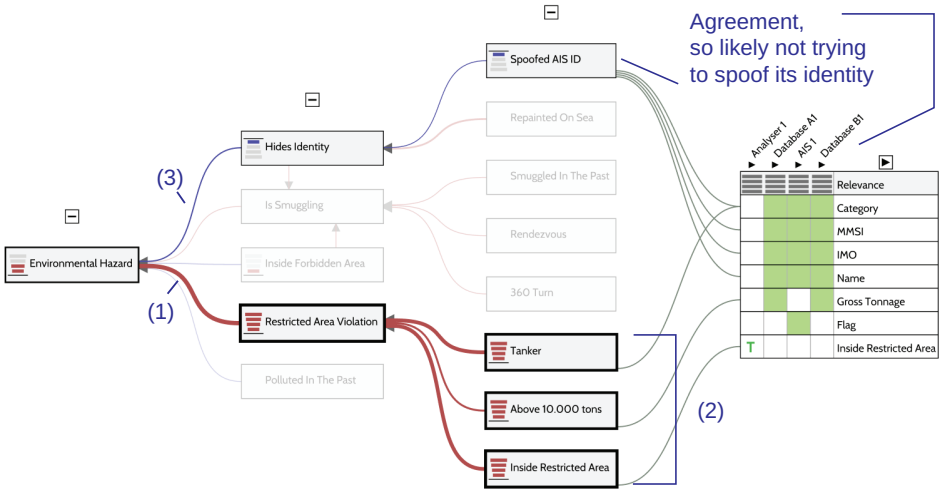


Figure 6.6: We see a vessel with a raised probability that it poses an environmental hazard. We can see this is mainly due to *Restricted Area Violation*, as shown by the thick red line in between these nodes. This node's probability is in turn raised due to a set of three nodes that together constitute a domain rule that tankers above 10,000 tons are not allowed to be in certain restricted areas. Because of the red coloring and the thick borders we can immediately see that this rule is being violated. We can also see that there is no contradiction in the evidence matrix because all cells have the same color. This leads to a lower probability for *Spoofed AIS ID*, which in turn leads to lower probability of *Hides Identity*. We also see a blue line from *Hides Identity* to *Environmental Hazard*, which is due to the assumption that vessels that do not hide their identity, are less likely to pollute.

that forbids tankers above a certain size to be in environmentally sensitive areas. This is assumed to be part of the operator's domain knowledge, and was indeed immediately recognized by the experts as such. In the evidence matrix we can see all observations agree on the identity of the vessel; it is therefore less likely to be spoofing its AIS ID and therefore also less likely to be hiding its identity.

6.5.3 Use Case 2: Reckless Behavior

In this use case, the operator is looking at vessels that are behaving recklessly–see Figure 6.7. Experts could immediately see that the root hypothesis, *Reckless*, has a rise in probability mainly due to the vessel's possible history of collisions, and that there is a small chance it is a smuggler due to its identity being unclear. The experts could quickly see through the evidence matrix that the latter is because what the vessel claims to be through AIS, cannot be supported by any other database records. As we can see in the evidence matrix, the green AIS attributes

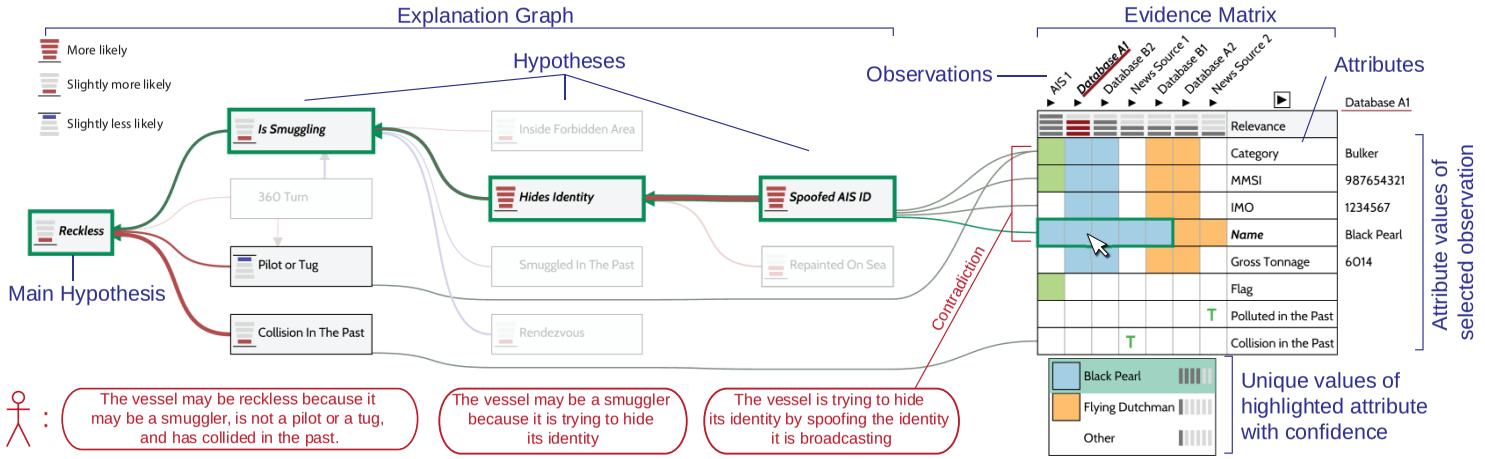


Figure 6.7: A rationale visualization of a maritime use case in which an alarm has risen due to possible *Reckless* behavior. By hovering the mouse over the name attribute, the operator is investigating the possible names of the vessel and the effect on the reasoning. All hypotheses that depend on the name attribute are highlighted in green.

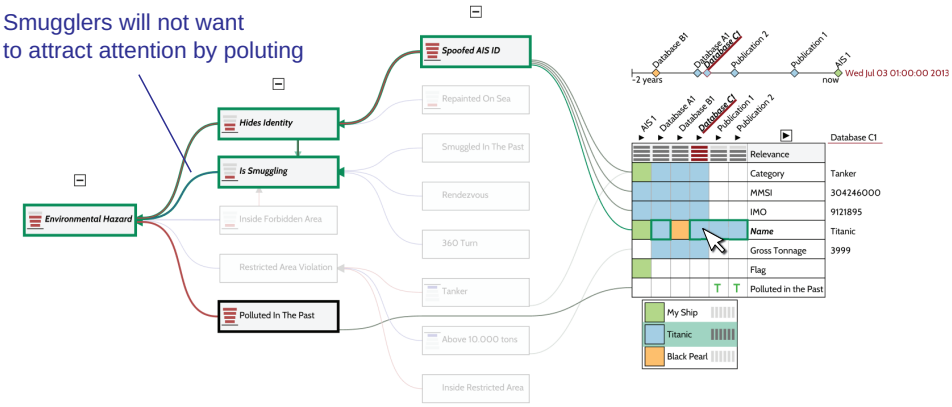


Figure 6.8: Here we see a vessel with a risen probability that it poses an environment hazard. We can see this is because the vessel is trying to hide its identity by spoofing its AIS ID, and because it already has polluted in the past. It can be immediately seen through the evidence matrix that the vessel is likely trying to hide its identity, because the information broadcasted by its AIS cannot be confirmed by other sources.

cannot be seen in any other observation. Also, it was clear to the experts that the additional records were found only by name, which is reflected by the lowered *relevance* p^{OOI} of the observations, which in turn is reflected in the low probability for *Collision In The Past*. According to the evidence, the vessel is less likely to be a pilot vessel or tug, both of which are expected to make more seemingly reckless maneuvers. This makes reckless behavior more likely.

6

6.5.4 Use Case 3: Environmental Hazard

In this use case, the operator is again looking out for vessels that may pose an environmental hazard–see Figure 6.8. A vessel in a particular area has triggered an alarm due to a rise in the probability that it poses an environmental hazard. Our experts could immediately see that the root hypothesis, *Environmental Hazard*, has a rise in probability because the vessel is trying to hide its identity by spoofing its AIS ID and because it has a possible history of polluting. It can be immediately seen through the evidence matrix that the vessel is likely trying to hide its identity, because the information broadcasted by its AIS cannot be confirmed by other sources. We can see this because the green AIS attributes cannot be seen in any other observation. Also, in the time line, we can see the publications on pollution events are separate events. Both groups of domain experts independently gave the same likely explanation based on the patterns in the evidence matrix: The vessel has changed some of its AIS attributes to hide its identity while it is illegally dumping waste, which is a common occurrence according to our experts. The MMSI and IMO identifiers, unique vessel identifiers, are harder

to change and using that, its real identity can be found through other sources. According to our experts, the differing name in the oldest observation *Database B1* means the ship has been renamed, but the database has not been updated yet.

In our explanation graph we see there is a rise in probability that the vessel is a smuggler, but this has a suppressing effect on *Environmental Hazard*. This is due to the assumption, as defined in the reasoning model, that a smuggler avoids attracting attention and therefore will less likely be polluting.

6.5.5 Use Case 4: Smuggling

In this use case, the operator is looking out for vessels that may be smuggling–see Figure 6.9. Our experts could see that the root hypothesis, *Is Smuggling*, has a rise in probability because the vessel is trying to hide its identity and, mainly, is inside a forbidden area. As before it can be seen through the evidence matrix, due to the variation in colors, that the vessel appears to be deliberately trying to hide its identity. The main hypothesis, however, seems to be mainly supported by the fact that the vessel is inside a forbidden area. Our experts agreed that this specific scenario is not cause for much alarm, which is supported by the low rise in probability of the root hypothesis.

We also see a slight decrease in the probability that the vessel is posing environmental hazard, because it is not violating the restriction area violation rule. We can see, due to the red edge (1), that this has a slightly raising effect on the main hypothesis. This is again due to the assumption that smugglers avoid attracting attention and therefore a vessel that does not pollute is slightly more likely to be smuggling.

6.6 Discussion

Our domain experts could easily follow the reasoning steps in the explanation graph, and found that by emphasizing relevant hypotheses and supressing irrelevant hypotheses it is easy to get a quick understanding of why and how the reasoning engine came to its conclusions. We also found that they, both in the evaluation sessions discussed above and in earlier discussions, could very quickly recognize and understand patterns relating to real world situations in the colors of the evidence matrix. An example of such patterns can be found in use cases two and three, where the vessel has deliberately tried to hide its identity by changing its AIS attributes. Another example is a pattern that occurs when the vessel has been recently sold, but not all databases have been updated yet. According to our experts, operators should have no problem understanding the rationale visualization.

While automated reasoning is required to filter the *uninteresting* cases from cases that warrant an alarm, the experts in both sessions found that they could easily reproduce some of the reasoning visualized by the explanation graph by

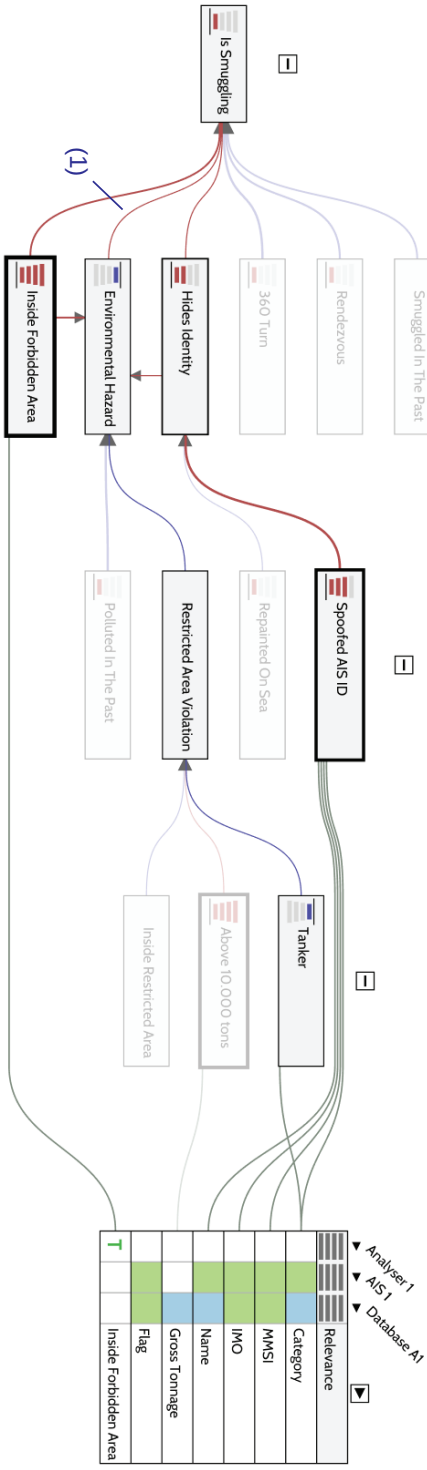


Figure 6.9: Here we see a slight rise in probability that the vessel is smuggling. This appears to be because the vessel is trying to hide its identity by spoofing its AIS ID, which in turn is supported by the varying colors in the evidence matrix, and because the vessel is inside a forbidden area.

looking only at the evidence matrix. The explanation graph was used only to, successfully, confirm their own reasoning.

In that light, the domain experts would prefer a reading direction from the evidence matrix to the root hypothesis. Additionally, the experts suggested using the evidence matrix as a compact glyph to show an overview of the degree of agreement for each vessel, while showing the explanation graph when the user drills down to gain a better understanding of why an alarm is raised and what the degree of agreement for a particular vessel means.

6.7 Conclusions & Future Work

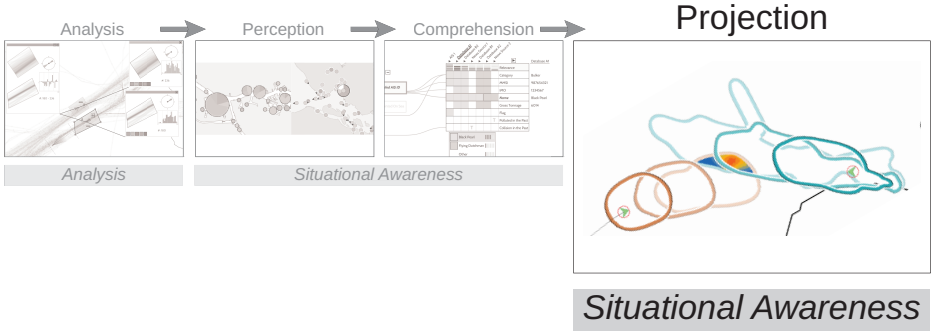
We have presented a method that uses an abstraction of an automated reasoning engine model to visualize the rationale behind its reasoning, enabling the user to understand the reasoning process and confidently take appropriate action. The rationale is visualized by showing a directed graph of connected hypotheses, which are in turn connected to observed attributes visualized in an evidence matrix. Despite their widely varying heterogeneous attribute values, we show the observations in the evidence in a compact and quick to read matrix. The visualization has been designed in close cooperation with both automated reasoning and MSSD experts.

While the visualization of the rationale is intended for end-users, it has also proven its value in developing the reasoning engine. Several flaws in the reasoning engine's model have been detected, which before remained undetected, using the visualization.

In future work we would like to visualize how the rationale changes over time when more observations become available. Additionally, we would like to allow the user to interact with the reasoning engine [187]. This can be done on the observation level by allowing the user to make their own judgments on the reliability of the observations, and to allow the user to add their own observations, *e.g.*, what operators on board of a vessel can see with their own eyes. This can also be done at the attribute level where the user can decide, for example, that two seemingly different names of the vessel refer in fact to the same name. Additionally, we would like to allow the user to turn observations off and on, so that the user can gain an even better understanding of what the effect of these observations are on the reasoning result. We would also like to add functionality to directly manipulate the domain model, such that domain experts can use our visualization as a toolset to create or adapt domain models [68]. While our domain experts have assured us that the abstractions of operational reasoning models will not be much larger than the ones used in the use cases, we would like to explore other domains with potentially larger reasoning models to investigate the scalability of our visualizations. Finally, our visualization of the evidence enables users to quickly see similarities and differences between objects with heterogeneous attributes, and we would like to study if this approach is useful for multivariate data visualization in general.

7

Contour Based Visualization of Vessel Movement Predictions



How can we help an operator in projecting a current situation into the future?

Once relevant vessels have been identified, the operator must be enabled to project the situation into the future, i.e., what will or can the relevant vessels do next. We present a visualization method for the interactive exploration of predicted positions of moving objects, in particular, ocean-faring vessels. The prediction models generate temporal probability density fields starting from a known situation. We use contours to visualize spatiotemporal zones of these density fields. Predictions are split into a configurable number of segments for which we render one or more contours. Users, investigating and exploring the possible development of a situation, can see where a vessel will be in the near future according to a given prediction model.

The contents of this chapter have in part previously appeared in [173].

7.1 Introduction

In the maritime domain, moving object analysis is concerned with large volumes of (fused) movement data. These data are gathered from multiple sensors with different properties, such as radar systems and the Automatic Identification System (AIS) [111], and can be enriched using additional sources or reasoning [219]. Sensors have uncertainty by nature due to measurement errors, and different sensors may give a different readout. This introduces a degree of uncertainty. Additional uncertainty is introduced when attributes such as position are predicted.

In this chapter we focus on uncertainty and prediction of kinematic properties of vessels, especially position, but also course and velocity. In particular, we focus on interactive, visual investigation of position predictions and the interaction between predictions for multiple vessels. Two types of predictions are relevant: one that extrapolates a given position report into the future, and one that interpolates between two given position reports that are geographically and temporally relatively far apart. When the time period of a prediction is relatively long, we are interested in uncertain areas where a vessel may be within a time period of interest. Our aim is to visualize temporal probability distributions produced by prediction models. We define a temporal probability distribution as a positional probability distribution that changes over time. In this chapter we focus on predictions based on extrapolation.

We apply our method to support *operators* in the maritime domain that use a cooperative system for public safety [93], and, we also show that our method is applicable to other domains such as urban law enforcement. Operators can have a large number of responsibilities and objectives such as managing emergency situations, enforcing the law, protecting economic interests (*e.g.*, oil and gas platforms), and controlling the border (*e.g.*, against smuggling). All these objectives are related to situational awareness [72], which is defined as the perception and comprehension of the current situation and the projection of the current situation into the future. Our goal is to provide operators with a visualization that allows them to have a better understanding of the evolution of a situation. We define prediction models to demonstrate our visualization and show its effectiveness in some maritime use cases and, to show generality, in a pedestrian use case. We argue that our method supports the operator in objectives that warrant prediction of future vessel positions such as collision avoidance, catching smugglers and protecting commercial vessels against piracy. In all these cases the operator is an external observer and may have limited to no information on the vessels involved, which in turn introduces uncertainty. To support an operator in such tasks, we need both a model to predict future positions and a visualization to communicate this prediction to the operator in a visual analytics setting.

Visualization is a powerful tool for increasing situational awareness and reducing information overload [163, 126]. An operator may want to know where a vessel of interest is likely to be during some time interval, whether two vessels are at risk of colliding, or whether two vessels, suspected of being involved in smuggling operations, will have the chance to transfer illegal goods. An operator

may also want to see where a vessel with engine problems may drift to. A drifting vessel is at risk of colliding with other vessels or stationary objects, such as oil platforms. Our visualizations can also be used for predictions of environmental calamities, *i.e.*, the predicted spread of an oil spill.

We make a distinction between two types of behavioral classes for maritime data. First is *normal behavior*, which is defined by the average behavior of vessels in historical movement data sets, or what is *expected* of a vessel based on its type. Second is *complete potential movement behavior*, or *unusual* behavior, often using the maximal capabilities of a vessel to circumvent law enforcement. We also distinguish between two user types. Apart from the *operator*, introduced above, we have the *modeling expert*, who is interested in the average behavior of the prediction models for all vessels, for example, to visually compare different models when developing or improving prediction models.

7.2 Related Work

Existing navigation systems, such as automatic radar plotting aids (ARPA) [209], provide some functionality for collision-avoidance such as plotting the closest point of approach (CPA) between two vessels, and a predicted area of danger (PAD). These systems, however, assume both vessels maintain course and velocity, whereas we assume vessels follow non-linear prediction models. Additionally, we display likelihood of interaction.

Bomberger *et al.* [23] and Rhodes *et al.* [157] present vessel position prediction based on the current position and velocity of a vessel by using algorithms that learn vessel motion patterns from movement events. Their work, however, focuses purely on the learning algorithms while paying no attention on how to visualize the results to be able to present them to an operator.

A popular way to model object movement, especially with uncertain positions, is the space-time prism [122]. It can be used to estimate movement between two known points or a future prediction from a known starting point. Kuijpers *et al.* [123] extend this by constraining the space-time prism by kinematic properties.

Animal movements are studied by estimating movement paths of animals using the Brownian Bridge Movement Model (BBMM). The animal's mobility is modeled using Brownian variance [32]. Horne *et al.* [97] show how to estimate Brownian variance using a maximum likelihood approach. Kranstauber *et al.* [119] extend this further by changing the characteristics of the BBMM based on likely behavior. We also use random movement for one of our prediction models. Our movement is, however, constrained in acceleration and turn rate to better model the characteristics of vessels. Also, the BBMM is used for estimating area usage of animals between sparsely sampled points, while we predict future positions. Furthermore, animal movement tends to be more unpredictable, while vessels are more inclined to follow maritime rules.

Clustering trajectories is a technique that is often used to analyze large quan-

tities of movement data [8, 6]. Etienne *et al.* [74] cluster trajectories by itinerary. Through statistical analysis, spatiotemporal patterns are extracted that are used to find unusual behavior such as veering off course or being late. We use techniques used to cluster trajectories to find similar trajectories for our prediction models.

An overview of a large amount of trajectory data can be visualized by convolving them onto a density map [217]. Scheepens *et al.* [179] extend this by allowing interactive, visual exploration of attributes, such as time, in the aggregated data, allowing a user to investigate temporal patterns. In a further extension [176], they propose a method to interactively define computational networks that generate specific task-oriented density maps. A use case shows how this method can be used to find vessel interaction. Since the density maps are aggregated over time, it can be seen where these interactions occur, but how many vessels are involved or when interactions occur is not shown. Demšar and Virrantaus [51] use the space-time cube to visualize space-time density. We agree that the use of the third dimension for time can be effective, but also, space-time cubes can be difficult to interpret for non-expert users. We use a 2.5D approach here [212]: we use depth cues, like occlusion and fading, to show order in time.

Kosara *et al.* [117] use a technique called semantic depth of field, in which objects of interest are displayed sharp, while objects of less interest are blurred. We use a similar technique to give more visual importance to predictions closer to the current time.

Santiago *et al.* [168] present a tool to investigate the accuracy of flight prediction models using bubble plots. This also allows the comparison of multiple different prediction models. Sanyal *et al.* [169] propose a method to visualize uncertainties caused by the use of multiple numerical weather models. They use two methods: First, using glyphs that show the mean and the deviation of values over the different simulations, and second, using uncertainty ribbons to convey the uncertainty of contours where the width of the ribbon is based on the uncertainty.

Matthews *et al.* [133] incorporate uncertainty in a maritime operational picture to enhance situational awareness. The authors propose two types of glyphs to show the uncertainty in location, identity and the timeliness of the information.

7.3 Visual Prediction System

We are interested in a GIS-application that allows an operator to explore an operational view of traffic. The operator should be enabled to interactively investigate the development of the current situation by selecting one or more vessels and generating predictions for future positions. Using the interactive visualization techniques described in this chapter, the operator can investigate where a vessel may be within a certain time period or a certain time interval in the future, what the chance is two or more vessels will collide in the near future, or whether two or more vessels may have the opportunity to interact.

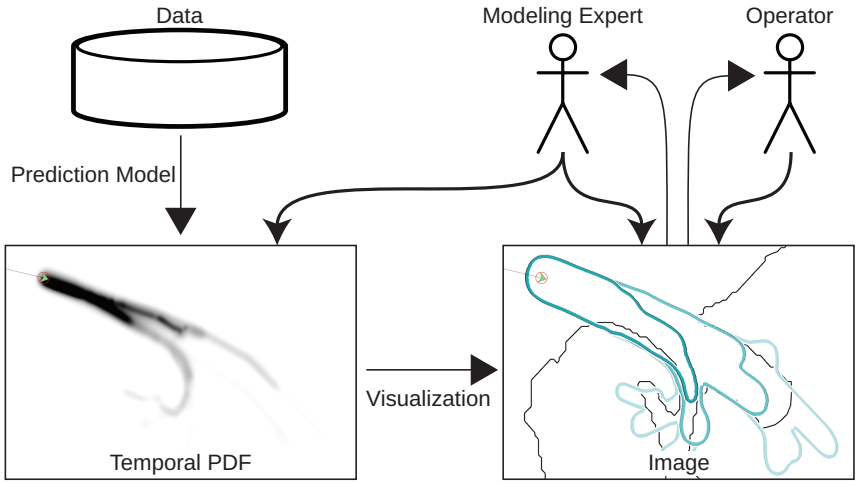


Figure 7.1: The architecture of our prediction system is divided into three major components: The data component is a large database containing historical trajectories and other information such as weather data; the prediction model is a model that produces a temporal distribution; and the visualization component composes one or more temporal distributions into a single visualization. The operator can adjust the visualization, while the modeling expert can adjust both the visualization and the prediction model.

Our prediction visualization has three parts: the data, the prediction model and the visualization component (see Figure 7.1). The data part is a database containing a large volume of historical trajectories and other data such as weather information. The prediction model provides a temporal Probability Density Field (PDF) to the visualization component. In the visualization component, temporal distributions are composed into a picture that allows a user to interactively and intuitively investigate these distributions over time. Operators may want to change parameters of the visualization in their investigation, while a modeling expert may also want to change parameters of the underlying models to investigate the performance or sensitivity of a prediction model. In this chapter we mainly focus on the visualization part, but to be able to show results we also define prediction models for demonstrating our visualization methods.

7.3.1 Data

We use a large data set H containing vessel trajectories on the Dutch continental shelf for four non-consecutive days. Each day contains around 2000 unique vessels, 4000 trajectories, and over 7 million position measurements simplified to approximately 210 000 points using segmentation [219]. The measurements in our data are obtained using AIS, but can also be obtained using conventional radar systems. The attributes of our trajectories are sampled over time, typically in intervals of several seconds. We assume vessels move from a positional sam-

ple to the next using a constant course. In the case of sparse sampling, this may give erroneous results, however, our study area has an excellent coverage. To model the movement between consecutive samples, we use the movement model defined by Willems *et al.* [217].

7.4 Models

A prediction model produces a positional prediction over time as a temporal distribution. The model may or may not be data-driven. It may be a task-driven model, for example, a model that describes the drift of a vessel. A model to predict the drift of a vessel can be based on other data, such as current and wind. Any model that produces a temporal distribution can be plugged into the system.

Our prediction models produce, for a reference object o, a probability density $P_o(\mathbf{x}, t)$ with position $\mathbf{x}$, such that the chance that the object is in area A at time t, is given by:

$$P_o(\mathbf{p}_o \in A, t) = \int_{\mathbf{x} \in A} P_o(\mathbf{x}, t) d\mathbf{x}. \tag{7.1}$$

Besides the probability for a certain time t, also the integrated probability $P_o^{'}(\mathbf{x})$, *i.e.*, the chance that an object has been at a certain position $\mathbf{x}$ in the time interval $[0, T]$ can be interesting and useful. This can be obtained by integration of $P_o(\mathbf{x}, t)$:

$$P_o^{'}(\mathbf{x}) = \frac{1}{T} \int_0^T P_o(\mathbf{x}, t) dt, \tag{7.2}$$

where the current time is considered to be $t = 0$ and T is the length of the prediction period.

We define two general prediction models below, which we apply both for ocean-faring vessels, and pedestrians. Our terminology, however, is inspired by the maritime domain. The first model is based on finding *similar* trajectories in a large historical data set and is more suited to describe normal behavior. The second model estimates kinematic properties of the reference vessel and simulates a large number of trajectories. It is more suited to describe complete potential movement behavior.

For both models we use, for vessel o, trajectory convolution $C_o(\mathbf{x}, r)$ at point $\mathbf{x}$, and radius r, as described by Willems *et al.* [217]:

$$C_o(\mathbf{x}, r) = \frac{1}{T} \int_0^T k_r(\mathbf{x} - \mathbf{p}(t)) dt, \tag{7.3}$$

where $\mathbf{p}(t)$ is the position of o at time t, and k_r is a kernel with radius r.

For our models we describe the confidence in the resulting temporal PDF, by the number of trajectories used to obtain the result and the average of a similarity measure that is defined later.

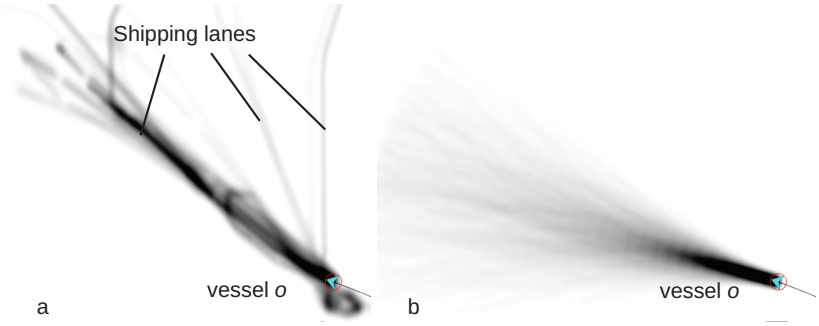


Figure 7.2: An aggregated probability distribution of vessel position over 30 minutes of a selected reference vessel o, using (a) the history based model, and (b) the simulation based model. In the history based model the underlying structure of the shipping lanes is visible.

7.4.1 History Based Model

We use a data set H as a base for our history based model. To get insight into possible future positions of a reference vessel o, we scan H for ships that have been close to the position of vessel o and had similar characteristics. Next, we construct the PDF $P_o(\mathbf{x}, t)$ by convolving and aggregating similar trajectories using Equation (7.3). Equation (7.2) then gives $P_o'(\mathbf{x})$.

First, we construct a list of *similar* trajectories over some time period of length T with, for each trajectory s, a trajectory similarity measure $\phi_s \in [0, 1]$, describing how similar a trajectory is to the reference trajectory of o. To determine the trajectory similarity measure ϕ, we require a similarity measure between two vessel states α and β. In this section, we define such a similarity measure σ, such that $\sigma(\alpha, \beta)$ is in $[0, 1]$, equal to 1 if the states are highly alike, and equal to 0 if the states are not alike at all. Furthermore, two states α and β are called similar if $\sigma(\alpha, \beta) > 0$. We assume a trajectory is similar to the reference trajectory if its first state is similar to the current state of reference vessel o. In other words, the similarity measures to what extent a trajectory is relevant for predictions about the future path of the reference vessel. Let S be the set of sub-trajectories of the historical trajectories H that have duration T and are similar to the reference trajectory, *i.e.*, $\phi_s > 0$ for all $s \in S$. From here on we assume every trajectory in S is shifted into the time interval $[0, T]$.

The less similar a trajectory is to the trajectory of the reference vessel o, the less certain we are that the reference vessel o may behave according to this trajectory. Hence the uncertainty that the reference vessel o behaves as a given historical trajectory s in S increases as the trajectory similarity measure ϕ_s decreases. We use this uncertainty in our model by scaling both the kernel radius and the density value itself to convey the uncertainty due to the similarity into both the area of influence and the total contribution to the probability of the vessel, respectively. The probability P_o (see Equation (7.1)) is then defined as follows:

$$P_o(\mathbf{x}, t) = \frac{1}{\sum \phi_s} \sum_{s \in S} \phi_s k_{r(\phi_s)}(\mathbf{x} - \mathbf{p}(t)), \tag{7.4}$$

where $P_o^{'}$ follows from Section 7.4.

The definition of the kernel radius function is left up to expert modelers, such that they can decide how important the impact of uncertainty is. In this chapter we choose a logarithmic kernel size: $-\frac{1}{2}\log\phi$, however, this can be any decreasing function. Figure 7.2a shows a probability distribution generated using this model.

The confidence in the prediction using this model is described by the number of similar trajectories $|S|$ and the average similarity measure $\sum_{s \in S} \phi_s / |S|$.

Similarity Measure

The similarity measure σ of two vessel states $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ at time t and $t^{'}$, respectively, is determined with a user-configurable function:

$$\sigma(\boldsymbol{\alpha}(t), \boldsymbol{\beta}(t^{'})) = \prod_{a \in A} 1 - w(a) \cdot dist_a(a_{\boldsymbol{\alpha}}, a_{\boldsymbol{\beta}}), \tag{7.5}$$

where A is the set of all available attributes in a state, $w(a)$ is a user definable weight in the range $[0, 1]$ for attribute a, and $dist_a$ is a user-definable function that determines the normalized distance in $[0, 1]$ between two values of attribute a. For example, if a modeling expert considers only vessels with approximately the same velocity as similar, the distance function for velocity can be described as follows:

$$dist_v(v_{\boldsymbol{\alpha}}, v_{\boldsymbol{\beta}}) = \max(0, 1 - \frac{|v_{\boldsymbol{\alpha}} - v_{\boldsymbol{\beta}}|}{\Delta V}), \tag{7.6}$$

where ΔV is a velocity difference within which vessels are considered similar. To get a realistic prediction we also define a distance function for position:

$$dist_{\mathbf{p}}(\mathbf{p}_{\boldsymbol{\alpha}}, \mathbf{p}_{\boldsymbol{\beta}}) = \max(0, 1 - \frac{||\mathbf{p}_{\boldsymbol{\alpha}}(t) - \mathbf{p}_{\boldsymbol{\beta}}(t^{'})||}{\Delta D}), \tag{7.7}$$

where ΔD is the maximum distance at which states are considered similar. Furthermore, a similarity function can be defined on time that relates vessels from the same time of day or in the same season. The modeling expert can define a set of similarity functions and attribute weights in $[0, 1]$. The operators then only have to select the appropriate set for their task. For the predictions visualized in this chapter we use an estimate of 200m for ΔD and 10kn for ΔV. These estimates have been derived in cooperation with our domain experts.

We currently find similar trajectories by only considering the current state $\boldsymbol{\alpha}_o(0)$ of the reference vessel o. To get better results, the model can be extended

to consider more states of the historical trajectory by applying distance measures between trajectories, such as time warping distance [18, 207] and the least common sub-sequence measure [1], or by a technique called lifting [39], which is used for comparing and smoothing trajectories. This should give a better similarity and may help to even out possible errors in measurements in the current state of reference vessel o, such as velocity spikes caused by measurement or transmission errors. Furthermore, similarity can also be determined by using geographical domain knowledge [50] such as information about shipping lanes, or by using geographic context such as obstacles or weather [31].

7.4.2 Simulation Based Model

In our second prediction model, we simulate vessel movements based on vessel characteristics such as maximum acceleration and turning speed. Using a Monte Carlo approach, we simulate a large number of trajectories, apply convolution described by Equation (7.3) for all these trajectories and sum them into a density field as described by Willems *et al.* [217]. To get a probability distribution, we divide by the number of simulated trajectories.

Characteristics of a vessel may be obtained in three different ways. First, by prior knowledge of the capabilities of the reference vessel o, however, we expect this information to be unavailable as it is not present in AIS or radar data. Second, by estimating the capabilities of the reference vessel by looking at its type and size. And, lastly, by extracting movement characteristics from trajectories of similar vessels [53]. In this chapter we have chosen to go for the third approach by deriving a probability distribution from vessel behavior of vessels similar to the reference vessel o in our historical data set H. We find a set of vessels S in the historical data set H that are comparable to the reference vessel o. We say two vessels are comparable if they have the same type and similar size, according to a user-definable function. We have chosen this method because the information available on a target vessel may be very limited, *i.e.*, AIS data may not be available. The type and size can be estimated through other sensors, such as human observation or radar, if required.

Since vessels tend to maintain both their course and velocity, we derive *rates of change* P_c and P_v, respectively, that describe the chance a vessel changes its course or velocity in a time step Δt. We determine the proportion P_c of the samples at all time stamps $i\Delta t$ at which the similar trajectories change course. The amount of change of the course is computed using a normal distribution $\mathcal{N}(\mu_c, \sigma_c^2)$ and $\mathcal{N}(\mu_v, \sigma_v^2)$ with average $\mu_c = \mu_v = 0$ and σ_c^2 and σ_v^2 as found in the historical tracks for the course and velocity, respectively. A single trajectory is then modeled as follows. We assume that at time t a vessel may choose to randomly change its course and velocity based on the above probability distributions and rates of change. The state $\boldsymbol{\alpha}_o(t + \Delta t)$ of the vessel at $t + \Delta t$ is determined using a simple Eulerian scheme:

$$\mathbf{p}(t+\Delta t) = \mathbf{p}(t) + \Delta\mathbf{p}, \tag{7.8}$$
$$\Delta\mathbf{p} = v(t)(\sin(c(t)), \cos(c(t)))\Delta t, \tag{7.9}$$
$$v(t+\Delta t) = v(t) + \Delta v, \tag{7.10}$$
$$c(t+\Delta t) = c(t) + \Delta c, \tag{7.11}$$

where Δv and Δc are a random velocity or course change, respectively, based on the derived probability distributions and defined by:

$$\Delta c = \begin{cases} 0 & \text{if } \mathrm{rand}_{\mathcal{U}}(1) > P_c; \\ \mathrm{rand}_{\mathcal{N}}(\sigma_c^2) & \text{otherwise,} \end{cases} \tag{7.12}$$

$$\Delta v = \begin{cases} 0 & \text{if } \mathrm{rand}_{\mathcal{U}}(1) > P_v; \\ \mathrm{rand}_{\mathcal{N}}(\sigma_v^2) & \text{otherwise,} \end{cases} \tag{7.13}$$

where $\mathrm{rand}_{\mathcal{U}}(x)$ is a random uniform sample taken from the range $[0, x]$ and $\mathrm{rand}_{\mathcal{N}}(\sigma^2)$ is a random sample taken from $\mathcal{N}(0, \sigma^2)$. An example of a probability distribution generated using the simulation based model is shown in Figure 7.2b.

In the model described above, we derive statistics based on vessel type and vessel size. This model can, however, be extended to simulate actual vessel behavior more closely by deriving location-based statistics from vessels that are within a certain radius of the reference vessel o. This can be further extended by also taking velocity and heading into account. These statistics have to be derived again for every step, as the current location, heading and velocity may change. More in general, the similarity measure introduced in Section 7.4.1 can be used to find similar vessel states. If additional data, *e.g.*, from AIS, is available, it can be used to more accurately find similar vessels.

The confidence in the prediction using this model is described by the number of similar trajectories $|S|$ and an average similarity measure based on the vessels in S.

7.4.3 Model Comparison and Discussion

The two prediction models described above have their own strengths and weaknesses. The history based model gives a realistic positional probability distribution, implicitly using shipping lanes and other popular routes. Since the model is based purely on averaging historical movements, the resulting distributions can only describe *normal* behavior, which may not be what an operator is looking for. The simulation based model, on the other hand, *can* describe complete potential movement behavior. It does not, however, describe any structures such as shipping lanes. This can be solved by extending the model such that location-based statistics that also take course and velocity into account, are used. Both models can suffer from a lack of similar trajectories. This can be solved by widening the range of what is considered similar if such a case arises. This results, however, in a lower confidence.

Since we mainly focus on visualization, we have kept our models, and especially the simulation model, fairly generic. For a real world system, specialized, mission-specific models can be developed, *e.g.*, a model that specifically deals with the behavior of a vessel engaged in pirate activity. Reasoning on the intent of vessels, as discussed in Chapter 6, may then serve as a guide in selecting an appropriate vessel movement model.

Mobility patterns generally depend on the spatiotemporal context of movement [5], such as the relationship between the vessel and the tides, currents, weather conditions, and physical barriers such as other vessels, platforms, and the coast. Both our models, especially the simulation model, ignore this context to a large degree. The history based model does indirectly deal with fixed physical barriers such as the coast or platforms, and spatial context such as shipping lanes because this information is in the mobility patterns of the historic trajectories themselves. It does not, however, deal with any spatiotemporal context directly.

The required spatial context such as barriers can be retrieved from charts, while spatiotemporal context can be retrieved from weather services, radar, and other sensors. Additionally, spatiotemporal context can also be extracted directly from the historical (AIS) trajectories using visual analytic methods [6, 11], e.g., finding anchorage or fishing zones [179], extracting shipping lanes [176], understanding the dynamics of shipping lanes (see Chapter 3), or even extracting currents [99]. Developing models that take all relevant spatial temporal context into account is, however, not within the scope of this thesis.

7.5 Visualization

Now that we have models to predict vessel position over time, we require a visualization technique to allow users to intuitively recognize patterns in the predictions, enabling them not only to see where a vessel may be, but also when the vessel may be there. There are several methods to visualize a temporal distribution, such as small multiples or volume rendering techniques. Since our users are familiar with maps, however, we have chosen to visualize temporal PDFs in a 2.5D visualization that reveals both the space and the time component.

We use the visualizations described in this section in a prototype mimicking an operational environment (see Figure 7.3). We present the geographical context to the user with land mass shown in grey and the ocean shown in white. The current position of each vessel is marked using a colored arrow shape. The color of the arrow represents the type of the vessel. The direction of the arrow is based on the heading of the vessel. A small faded trail is used to show the position of the vessel during the last 10 minutes.

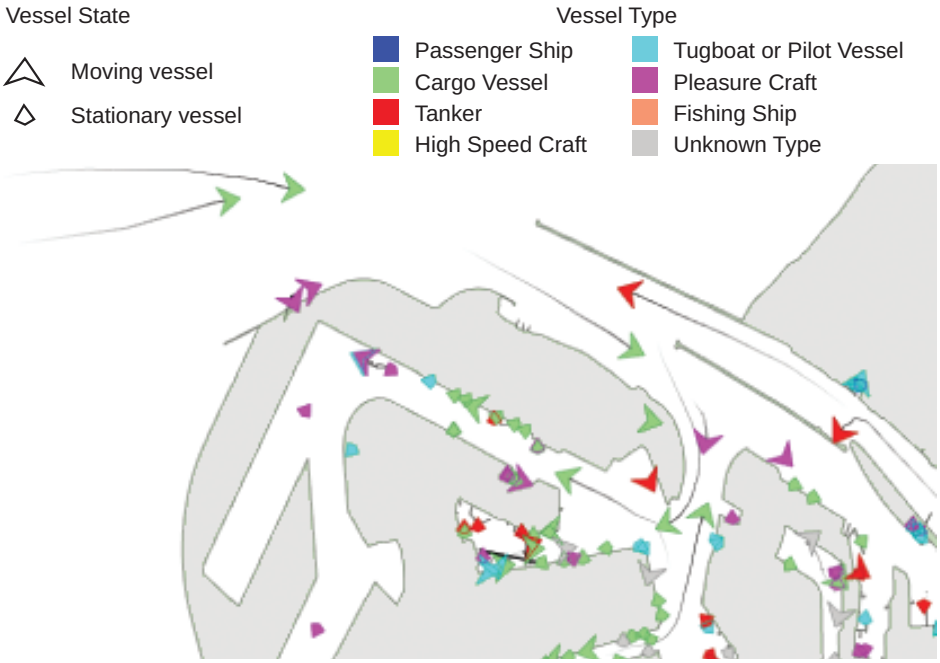


Figure 7.3: The Rotterdam harbor, a busy area near the Dutch coast (green) with many vessels docked in the harbor area. The current vessel positions are shown using a colored shape. Their color is based on vessel type. An arrow shape is used for moving vessels and a diamond-like shape is used for stationary vessels, where the orientation of the shapes shows the orientation of the vessel. Using a fading trail, the positions of each vessel during the last 10 minutes is shown.

7.5.1 Temporal Probability Density Fields

The PDF $P_o'(\mathbf{x})$ can be visualized using the techniques described by Scheepens *et al.* [179]. This, however, means aggregating probability over time, and, while this gives a good overview of the area utilization distribution, the influence of time is not or only slightly visible. Since we work with prediction models, typically for a relatively short period of time, such as one or two hours, the relationship between time and area utilization distribution becomes important. An operator does not only need an estimate of *where* a vessel o may be within the time range $[0, T]$, but also *when* the vessel may be in that area.

To visually maintain the relationship between time and space, we propose to divide our PDFs into a number n of time segments. This can be done in three ways: For $t_i = iT/n$ with $0 < i \leq n$, we have time intervals $[t_{i-1}, t_i]$, time instances t_i, and increasing time intervals $[0, t_i]$ (see Figure 7.4a, b and c, respectively). We find that intervals are best suited for investigating the development of area usage over time, while instances are best suited for estimating the position of vessel

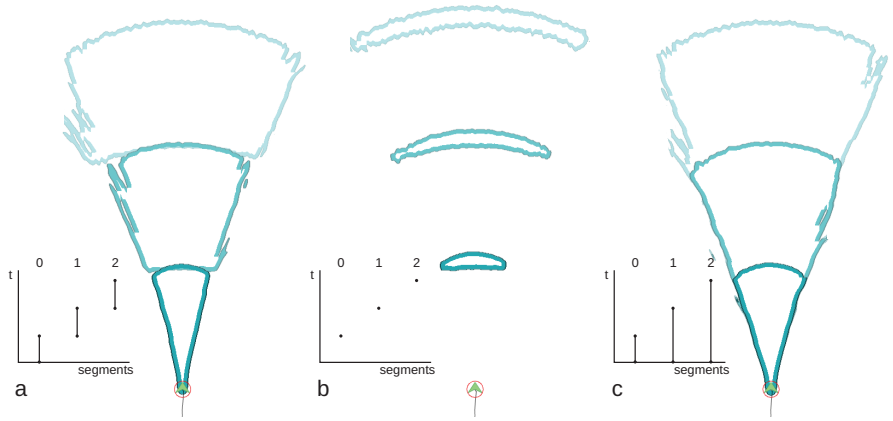


Figure 7.4: A visualization of a thirty minute prediction using the simulation based model with (a) three time intervals, (b) three time instances, and (c) three increasing time intervals.

o changing over time. The advantage of the increasing time intervals is that subsequent segments are visually-geometrically linked.

7.5.2 Contours

We have n, potentially overlapping, PDFs for which we want to visualize the probability distributions, the order in time, and the confidence in the prediction model. Visualizing these PDFs as a density map may lead to an information overload and to overlap problems. Another common method to visualize PDFs is by showing contours. Given a percentile E, a contour encloses a spatiotemporal zone, such that the chance that the vessel is inside the area in the given time segment is E and the value of the PDF at the boundary is constant.

To easily identify the inside of the spatiotemporal zones, we mark the inside of the contours with a thin, colored edge (see Figure 7.5a). Furthermore, to help the user to recognize overlapping contours we add a white halo [13, 75] around the contours such that contours seem to stop when they cross other contours (see Figure 7.5a). We render our contours ordered from later in time to earlier in time such that spatiotemporal zones closer to the current time are on top and later spatiotemporal zones may be occluded. For each time segment, we draw the white haloes for all contours first, followed by the colored inner lining and the contour lines themselves. This means haloes only interrupt the contours of predictions further in time, which puts contours that belong to the same time segment visually closer together.

The spatiotemporal zones will have increasing uncertainty as the vessel position is predicted further into the future. We visually convey more importance and thus more confidence in predictions closer to the current time by the ordering of the contours described above, and also by fading the contour lines to white and

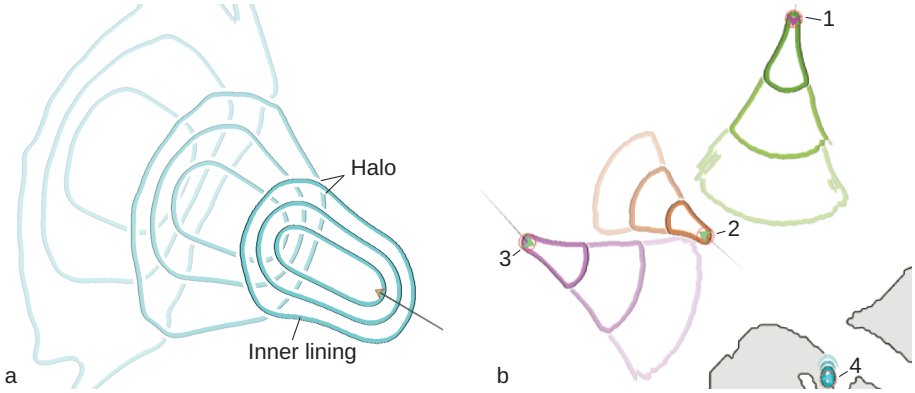


Figure 7.5: (a) A visualization of three time intervals of a vessel prediction over 15 minutes using contours at the 15%, 25%, and 95% percentiles. The further into the future an interval is, the more faded the contour appears. And, (b) thirty minute predictions of the simulation based model, divided into three time intervals. Four vessels are selected: a pleasure craft lined in green (1), a cargo vessel lined in orange (2), another cargo vessel lined in purple (3), and a special craft that is just departing from its harbor lined in teal (4).

reducing the saturation of the contour linings further in time (see Figure 7.5a). For each time segment of the prediction, one or more of the above described contours are rendered for a selected set of percentiles.

Since multiple vessels may be selected at once, it is possible that contours of different vessels overlap and confuse the user. We therefore encode the identity of the selected vessels by the hue of the contour lining, see Figure 7.5b. We use the iso-luminant color map used by Scheepens *et al.* [179] for our color coding.

7.5.3 Interaction

We say two vessels o and $\bar{o}$ have a chance to interact in their predictions $p_o(\mathbf{x}, t)$ and $p_{\bar{o}}(\mathbf{x}, t)$, at some position $\mathbf{x}$ and at time t if the interaction chance $I(\mathbf{x}, t) = p_o(\mathbf{x}, t)p_{\bar{o}}(\mathbf{x}, t)$ is non-zero. We render the potential interaction per pair of contours sharing the same time segment using a blue to red color map. For a pair of contours, we show the interaction chance only where the individual areas overlap. The interaction probability is rendered after the haloes, but before the colored inner linings, per time segment. This visually links the potential interaction area to the contours of its time segments (see Figure 7.6a). In Figure 7.6 we show an example interaction prediction between two tanker vessels. For both vessels we predict 12 minutes, divided into two time intervals, using the history based model. We see an area of potential interaction during the second time interval. The interaction probability $I(\mathbf{x}, t)$ can also be treated as a separate PDF, visualized in a similar way to the PDFs of the individual vessel prediction. We found, however, that this introduces more visual overload.

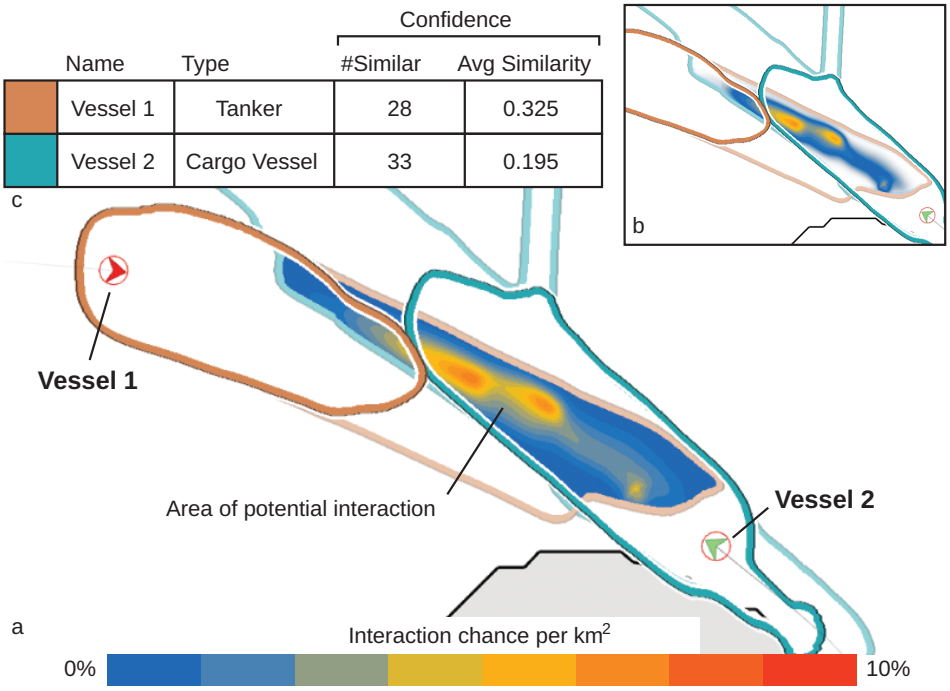


Figure 7.6: The interaction between predictions of two vessels. Both predictions are over 12 minutes, divided into two time intervals. The interaction chance is shown using a discrete color map ranging from blue (low chance), to red (high chance). We see a large area of potential interaction in the second time interval. The range of the color map can be changed to suppress low probabilities (b). In (c) we show a table showing the confidence of the predictions.

If the user is not interested in low probabilities, the start range of the color map can be increased. All probabilities below the start of the color map range are faded away. This allows a user to focus only on probabilities that are considered significant (see Figure 7.6b). This may, however, break the visual link between the potential interaction area and its contours, making it harder to interpret.

7.5.4 Confidence

For each selected vessel in Figure 7.6 the confidence measures of Section 7.4 are displayed in a table widget (Figure 7.6c) such that the operator can determine the relative reliability of the displayed positional predictions. The confidence of the predictions may be used to support the decision of the operator. For example, the prediction of Vessel 1 in Figure 7.6 is computed using a smaller number of similar tracks than Vessel 2, however, Vessel 1 has a higher average similarity. This may lead the operator to conclude that the prediction of Vessel 1 is more reliable than that of Vessel 2.

7.6 Implementation

We find similar trajectories for the history based model and simulate the trajectories for the simulation based model using the CPU. The PDFs are computed and visualized using a Graphics Processing Unit (GPU), using a grid of cells of equal area and stored as textures in video memory. We use image-based techniques to render our contours and make heavy use of shaders on the GPU. The PDFs are computed only when required in approximately 200 ms for a resolution of approximately 1000 by 1000 cells on an Intel Core i7 CPU with 6GB of RAM and an Nvidia GeForce GTX 285. The fields are stored and visualized in real-time.

7.7 Use Cases

In the following sections we demonstrate our visualization methods with real-world use cases in the maritime domain taken from Hendriks and van de Laar [93]. Our visualizations are intended for operators in the maritime domain, but are also applicable to other domains concerned with moving objects, such as air traffic control, ecology, or urban law enforcement. We show how our visualizations can aid an operator in collision avoidance, and how smuggling operations and pirate activity can be investigated. Additionally, we show how our method can be applied to a different domain, namely urban law enforcement.

Many more use cases exist for our visualization method. For example, the visualization of a normal behavior model can help a user to visually detect anomalies [162].

7.7.1 Collision Avoidance

Ocean faring vessels are often large vessels with low maneuverability. A collision between two vessels can lead to considerable economic damage or loss of life. Therefore, even if the risk of collision is low, the potential impact may be great.

An operator, whether on board of one of the vessels involved in the possible collision, or on board of some law enforcement vessel monitoring traffic, may be notified of a possible collision by an alarm or notification powered by some collision avoidance algorithm. Our method can help by showing how likely a collision is and also where and when this collision may occur.

We look at a real situation in which a collision has occurred between two cargo vessels. We visualize a prediction of the situation approximately 10 minutes before the collision in Figure 7.7a. Using the simulation based model we generate a prediction for 30 minutes, divided into two time intervals. A large chance of interaction is shown in a relatively small area within the first 15 minutes. In Figure 7.7a, the location where the two vessels actually collided is marked.

Based on this visualization, an operator may decide to instruct one of the vessels to change course, avoiding a collision. In Figure 7.7b we show a search

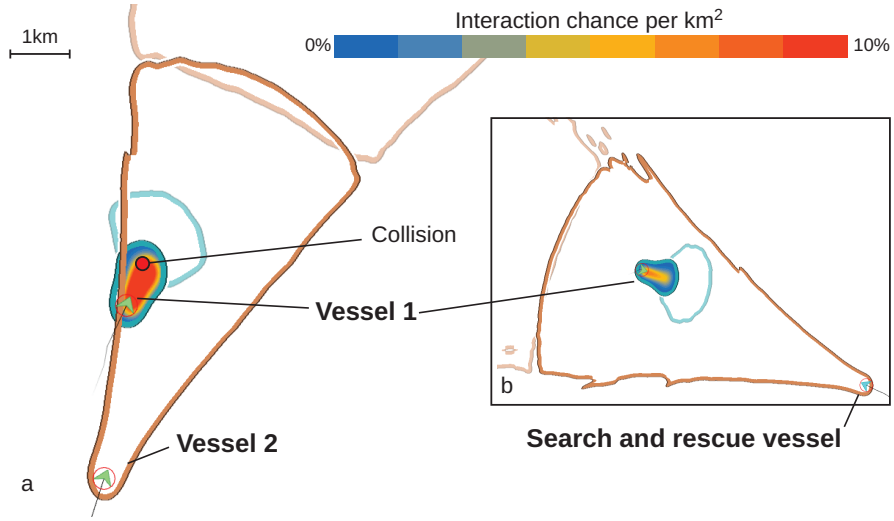


Figure 7.7: (a) A real situation approximately 10 minutes before a collision between vessel 1 and 2. Using our prediction visualization we can predict the area in which the vessels may collide. We predict the vessel positions using the simulation based model with two time intervals of 15 minutes each. (b) A search and rescue vessel intercepts vessel 1 after the accident, visualized using the same settings as (a).

and rescue vessel intercepting Vessel 1, which has gone adrift due to the damage caused by the collision. This has been visualized using the same settings as above.

7.7.2 Smuggling

Criminals may use vessels to smuggle illegal goods such as drugs or weapons [93]. These goods are often transferred between vessels in an attempt to avoid detection. Such transfers can be done without the vessels stopping or significantly slowing down. Operators investigate individual vessels or pairs of vessel based on intelligence. A law enforcement vessel may be sent to intercept the vessels to catch the criminals in the act. However, sending out a vessel to intercept is costly and multiple objectives may simultaneously require interception which leads to a conflict in resources. Therefore, using our visualization, the operator can investigate whether the suspect vessels may actually meet and where they are most likely to do so, and can take an informed decision on whether to dispatch a law enforcement vessel.

An operator looks at a situation in Figure 7.8 in which two selected vessels (1 and 2) are suspected of being involved in a smuggling operation. The operator wants to know if and when the two suspect vessels may meet to transfer illegal goods. Since smugglers will try not to be noticed, we use the history based model to represent normal behavior. We show two time intervals in a total period of

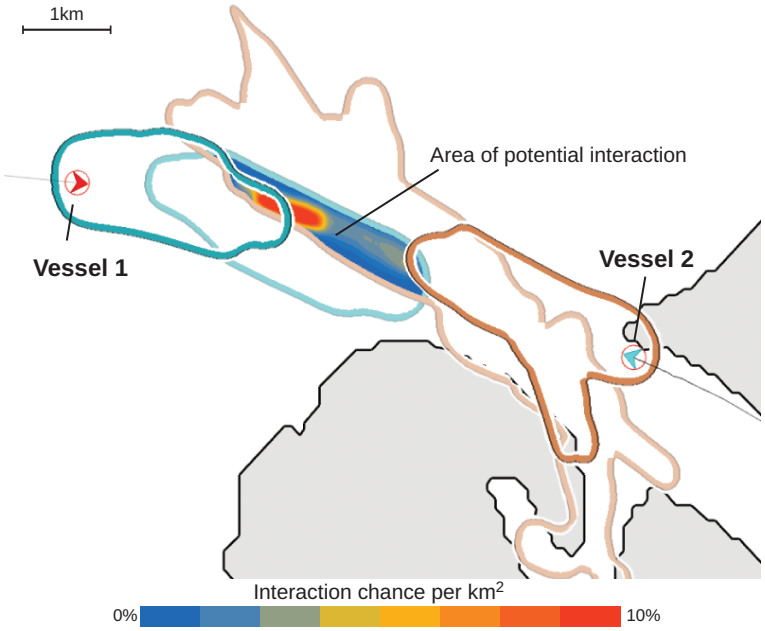


Figure 7.8: Two vessels suspected of smuggling. We predict the vessel positions using the history based model with two time intervals of 6 minutes each.

12 minutes. As we can see, vessel 3 is expected to continue moving in an eastern direction, while vessel 4 is expected to either head south into the harbor area of Rotterdam or keep moving in a north-western direction. We can now see an area during the second 6 minutes, where illegal goods may be exchanged. Based on the visualization, the operator can decide to monitor the area of potential interaction during the given time period, or send a law enforcement vessel to intercept.

7.7.3 Piracy

Our visualization method can also be used to investigate the possible future movements of a vessel suspected of piracy. For instance in the gulf of Aden, pirates attack and hold merchant vessels for ransom. There are only a limited number of navy vessels in the area to prevent this. This lack of resources as in Section 7.7.2 means that a certain level of confidence is required before a vessel is dispatched, which can be attained by using our method.

Where the merchant vessels follow their usual shipping routes, pirate vessel will generally try to move fast and unexpectedly. We look at a situation in which a vessel suspected of being involved in piracy is heading towards a tanker. To decide whether or not to take action, an operator may use our visualization. We use the simulation based model to describe the complete potential movement behavior of the pirate vessel trying to attack a merchant vessel and the history

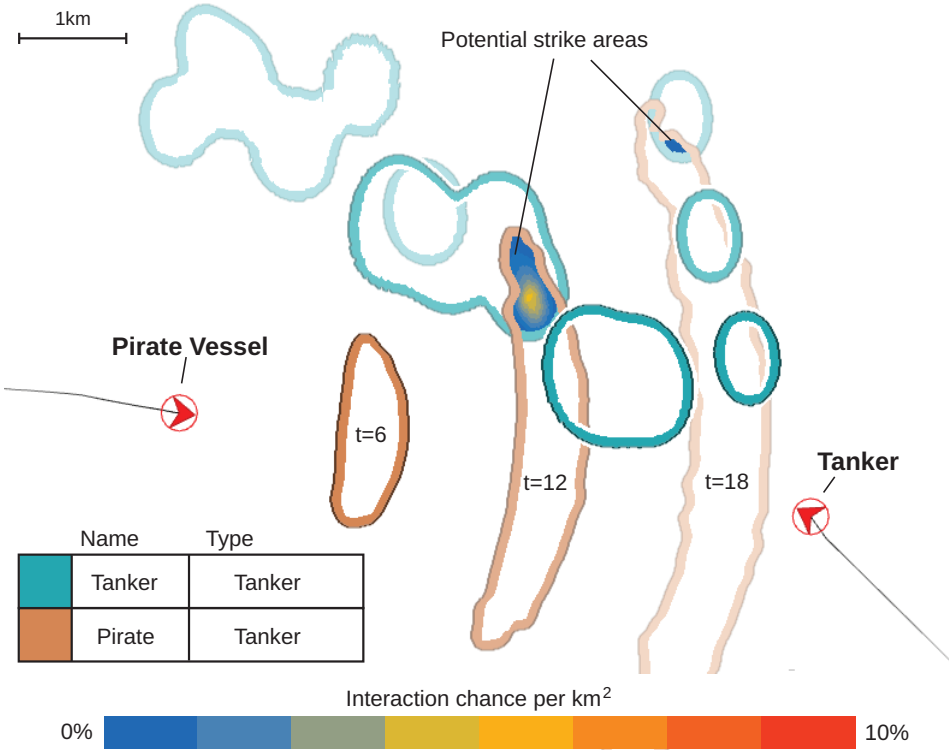


Figure 7.9: Prediction of the interaction between a tanker and a vessel suspected of piracy. The prediction is divided into three time instances, 6 minutes apart. There are two areas where the pirate vessel may strike; around 12 minutes and around 18 minutes.

based model to describe the normal behavior of a tanker, which appears to be likely to choose one of two routes. It can be argued that once the crew of the tanker realizes it is at risk, it will no longer follow normal behavior, however, in this example we assume the tanker is unaware. We set T to 18 minutes and select three time instances, the 6th minute, the 12th minute and the 18th minute. In Figure 7.9 we show the possible interaction between the tanker (teal) and the pirate vessel (brown). There appear to be two areas where the pirate vessel may strike. This may prompt the operator to decide to send in a law enforcement vessel.

7.7.4 Pedestrians

We show our method can be used for law enforcement in an urban environment, in this case the Dutch city of Delft. We look at a situation in which a law enforcement official is provided with a sighting at a certain point in time in the center

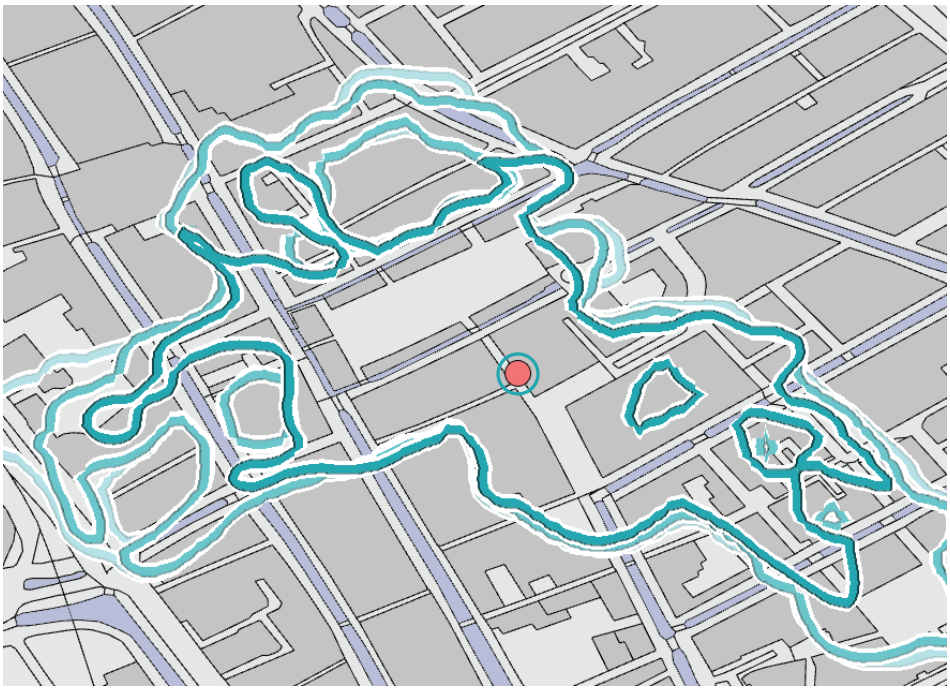


Figure 7.10: The possible future positions of a person suspected of pick-pocketing over three growing time intervals of 10 minutes each. The latest sighting of the suspect is marked by a red dot. We can see that the suspect will either move in a north-western direction, or in a south-eastern direction. Law enforcement officials can use this visualization to determine where to deploy scarce resources to apprehend the suspect.

of Delft of a person suspected of pick-pocketing. The assumption is made that the suspect maintains a low profile and moves around like a normal pedestrian. Using a large set of historical pedestrian trajectories the future positions of the suspect can be predicted with our method using the model based on history. This allows the law enforcement officials to deploy their scarce resources at the right place at the right time to increase the chance that the suspect is caught. For our historical data set H we use a large number of pedestrian trajectories in the city center of Delft [201].

In Figure 7.10 we show where the suspect may go in three growing time intervals of 10 minutes each. As can be seen in the visualization, the suspect is likely to either move in a north-western or in a South-Eastern direction.

7.8 Discussion

Our method is designed for the interaction of pairs of vessels as, according to our domain experts, this is the standard scenario. For interaction between three or more vessels visual clutter becomes a problem, and a more involved interaction model is required.

We presented our method for vessel position prediction and interaction to a group of six domain experts with naval backgrounds. They greeted the presentation with enthusiasm and ideas, which sparked an interesting discussion on the practical applications of our method. A number of experts suggested that the spatiotemporal zones visualized by the contours could directly be used as operational zones for surveillance using Unmanned Aerial Vehicles (UAVs), or search and rescue areas. We found that the domain experts find the concept of time instances the most intuitive of the three alternatives discussed in Section 7.5.1. The time intervals and growing time intervals appear to be interpreted as a description of movement through time instead of area usage within the time interval.

Additionally, we presented and discussed our method with a group of six naval operators and system developers. The experts were positive on the operational usefulness of our proposed method. The piracy case was of special interest to this group. The experts viewed our visualization as a rationale that not only shows when and where a pirate vessel may strike, but also *why* it may strike at that point and at that time. However, in an operational environment, the experts prefer that only the interaction zones are visualized and the entire visualization is only shown when an operator wants to see why the pirate vessel may strike there.

We also discussed our method with two ecologists, one interested in the migratory patterns of birds, and the other interested in the interaction of groups of primates. Due to both unpredictable movement of the animals and the precision of the tracking instruments used, animal movement models such as the BBMM generally contain a large degree of uncertainty. While the former ecologist was interested in the time aspect of the visualization, the latter was interested in the interaction visualization as applied to probabilistic movement models used to estimate the movement of groups of primates. The ecologist studying the interaction of groups of primates was interested in the probability, area and approximate time of interaction between groups and found our visualization method very suitable for this task.

7.9 Conclusion and Future Work

We have introduced a contour based visualization method to visualize models that predict future positions of selected moving objects. The prediction time period is divided into a number of time segments. For each of these segments, contours for one or more percentiles are rendered. These contours have an internal colored

lining to help identify the selected vessel and haloes to avoid confusion in the case of overlap. The saturation of the colored lining is reduced and the lightness of the contour is increased for spatiotemporal zones later in time to both help distinguish between time segments, and to convey the notion of increasing uncertainty for predictions further in time. We add an additional, color-mapped cue to highlight interaction between the predictions of multiple selected vessels. This interaction is shown per pair of contours for each time segment.

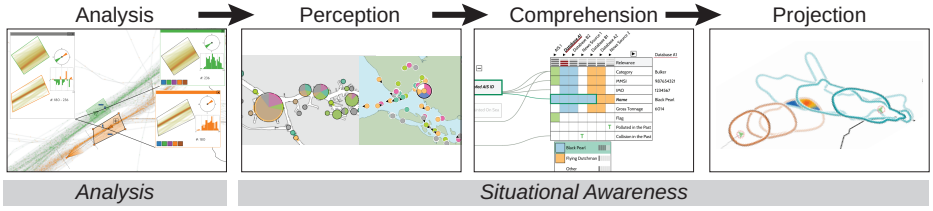
We have described two models that predict future vessel positions of some reference vessel within a given prediction period: A model based on comparing the current state of the reference vessel with a large set of historical trajectories and building a PDF from these trajectories, and another model based on simulating a large number of trajectories based on movement statistics derived from vessels similar to the reference in a large set of historical trajectories.

We presented the results of a discussion with domain experts and demonstrated the use of our method with a number of real-world maritime use cases: avoiding collision, and investigating smuggling and piracy. By application in the urban law enforcement domain, we demonstrated that our models and visualizations are not restricted to the maritime domain.

In future work we would like to apply our visualizations to additional domains with more uncertainty, such as animal movement. Another important topic is the scalability of our visualization method. Currently we have been looking at a selection of two vessels, but have not studied yet how it performs for a larger group of vessels. Lastly, we would like to do a user study with expert users to test which of the visualizations introduced in this chapter perform best under which conditions, and how they compare to similar visualizations. As an example, we would not only like to test which of the time segmentation methods defined in Section 7.5.1 performs best, but also like to find optimal visualization parameters for operators in the maritime domain.

8

Conclusion



In the previous chapters of this thesis we have attempted to answer our research question: "How can operators be supported in attaining situational awareness using interactive visualization?" In this chapter we look back at how we answered this question and we look forward towards what still needs to be done.

8.1 Contributions

The main contributions of this thesis are visualization techniques for analyzing and summarizing patterns to enable analysts to build and verify models of normal and anomalous behavior, and to aid in the three levels of Situational Awareness (perception, comprehension, and projection) as discussed in Chapter 1–see Figure 8.1. All visualization techniques in this thesis have been validated, either in close cooperation with domain experts or by performing a quantitative user study.

Traffic Flow Analysis

In Chapter 3 we have presented our approach to visualize traffic flows and provide interaction tools to support their exploration. We show an overview of the traffic using a density map. The user can discover and identify traffic flows using a visualization that is a combination of a density map and a particle system. Next, using a selection widget, these traffic flows can be selected. When selected, the user can explore the dynamics of traffic flows using annotation windows, which can be dragged on top of each other to compare multiple traffic flows.

Glyphs

In Chapter 4 we have presented a method to remove visual overlap of the glyph representations of moving objects. The set of all objects is partitioned into subsets, which are visualized with visually non-overlapping aggregated multivariate glyphs that show the distribution of attributes such as heading, type and velocity. The size of the glyph represents the size of the subset. The partition is updated as the objects move such that the change in subset configuration is approximately minimized. Changes in subsets in between frames are smoothly animated. A quantitative user study has been performed to compare the effectiveness of our method compared to alternative methods. We can conclude from this study that our method is competitive overall, and better than its competitors for density comparison tasks in the presence of a lot of overlap. Additionally, the participants of the user study have a preference overall for our method.

In Chapter 5 we have demonstrated the design and evaluation process of a new multivariate glyph to visualize vessel information with uncertainty. The

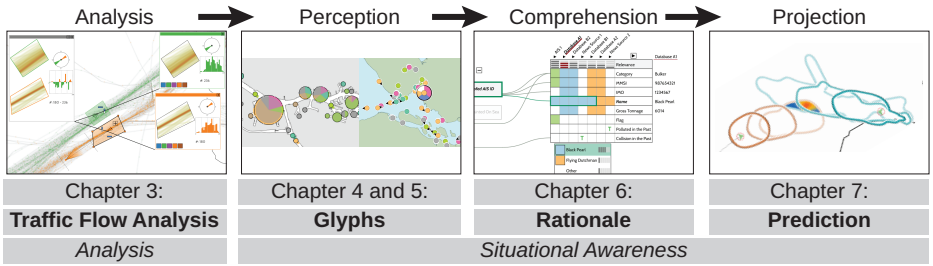


Figure 8.1: An overview of the contents of this thesis and their relationships.

glyph has been designed and evaluated in close cooperation with domain experts and potential users. It visualizes the uncertain distribution of the vessel type according to an automated reasoning and attracts attention from the user when required.

Rationale
In Chapter 6 we have presented a method to visually explain how an automated reasoner came to its conclusion. The visualization uses an abstraction of an automated reasoning engine model to visualize the rationale behind its reasoning, enabling the user to understand the reasoning process and confidently take appropriate action. The rationale is visualized by showing a directed graph of connected hypotheses, which are in turn connected to observed attributes visualized in an evidence matrix. Despite their widely varying heterogeneous attribute values, we show the observations in the evidence in a compact and quick to read matrix. The visualization has been designed in close cooperation with both automated reasoning and MSSD experts.

Prediction
Finally, in Chapter 7 we have introduced a contour based visualization method to visualize models that predict future positions of selected moving objects. The prediction time period is divided into a number of time segments. For each of these segments, contours for one or more percentiles are rendered. These contours have an internal colored lining to help identify the selected vessel and haloes to avoid confusion in the case of overlap. The saturation of the colored lining is reduced and the lightness of the contour is increased for spatio-temporal zones later in time to both help distinguish between time segments, and to convey the notion of increasing uncertainty for predictions further in time. We add an additional, color-mapped cue to highlight interaction between the predictions of multiple selected vessels. This interaction is shown per pair of contours for each time segment. We presented the results of a discussion with domain experts and demonstrated the use of our method with a number of real-world maritime use cases: avoiding collision, and investigating smuggling and piracy.

8.1.1 Research Question

In this section we discuss how the above methods have contributed to answering our research question. In Chapter 1 we divided our research question into four parts–see Figure 8.1. In the following we discuss how the presented visualization methods answer our research question and how these methods are interconnected:

i Analysis: *How can we provide tools to analyze and summarize patterns, enabling domain experts to find critical areas and to verify what normal or anomalous behavior is?*

The way vessels typically behave depends, among other things, on the environment and other traffic. To be able to build some model of normal behavior an analyst needs to understand patterns in traffic and their

relation to certain areas of interest. An analyst can do this using the visualization and interaction tools discussed in Chapter 3. The analyst can extract traffic flows of interests and study how their behaviour evolves over time to understand how vessels normally behave.

ii Perception: *How can we help the operator in perceiving a situation?*

Being able to perceive elements and their dynamics in the environment, especially those that are relevant, is the first level of situational awareness. In the maritime domain, the operational picture can contain hundreds of vessels. These vessels may be in close proximity, especially in busy harbor areas, leading to a clutter of their visual representations in this operational picture. This clutter is very harmful to acquiring and maintaining situational awareness as the operator may miss relevant elements due to overlap in visual representations. We discuss in Chapter 4 how we partition the vessels in view into subsets and visualize these using aggregated glyphs. We have shown that using this method, we can reduce the clutter problem in an operational picture even with moving objects. This enables the operator to perceive the situation better, especially in a busy area of interest.

To further improve perception and to enable the operator to more easily find and recognize elements that are relevant, we have designed a specialized glyph for the maritime domain as discussed in Chapter 5. This glyph also displays the status, attributes, and dynamics of the vessels. This information is fused from multiple, heterogeneous sources which leads to uncertain distributions [137]. Additionally, the glyph is designed to attract attention from the operator when required, according to an automated reasoning, by increasing both size and salience.

iii Comprehension: *How can we help the operator in comprehending a situation?*

The maritime environment can contain a large number of vessels. Due to the number of vessels and the amount of information that can be found about these vessels, it is not feasible for the operator to reason on each of these vessels to determine whether they are relevant and require attention. This is where automated methods, such as automated reasoning, can help the operator [139]. These automated reasoners, however, typically only output a conclusion and this conclusion alone does not help the operator in comprehending the situation. Therefore, we have presented a way to visually explain how the automated reasoner comes to its conclusion in Chapter 6. The automated reasoner now acts as a sort of filter that alerts the operator when a vessel requires attention, *i.e.*, using the glyphs of Chapter 5. For vessels requiring attention, the operator can then comprehend why the reasoning believes the vessel requires attention using our rationale visualization, leading to a better understanding of the situation.

iv Projection: *How can we help the operator in projecting a current situation into the future?*

Once relevant vessels have been identified, the operator must be enabled to project the situation into the future, *i.e.*, what will or can the

relevant vessels do next. This can be done using the prediction visualization of Chapter 7, which allows the operator to see what the potential next positions are of some vessel of interest. For example, the operator can assess the threat level of a suspected pirate by seeing whether the pirate can potentially strike nearby vessels and how likely these potential strikes are.

8.2 Discussion

As in many application domains, the amount of available information in the Maritime Safety and Security Domain is growing rapidly. Especially in situational awareness applications, where real-time processing of information on all actors in the environment is vital, this poses an enormous challenge. We have shown in this thesis that visualization is uniquely suited to help the operator or analyst to more easily process large quantities of information and find what is relevant (see Chapters 3, 4, and 5). Visualization alone, however, is not enough. The current trend, running counter to the rapid growth of the quantity of information, is to reduce ship's complements and their education and training time [204]. Therefore, automated methods are required to fill in the widening gap between available information and available human resources. These automated methods are required to support the operators in their tasks by filtering and making sense of the vast quantity of information from sensors and other sources, *e.g.*, by reasoning on the information as shown in Chapter 6, or by making sense of a situation, *e.g.*, by projecting it into the future as shown in Chapter 7. Visualization is then the means by which the operator can interpret and understand the reasoning and projection and to put new knowledge back into the system. Visualization remains essential as the link between the operator and the automated systems, keeping the operator in the loop.

To give an idea of how the visualization methods described in this thesis can be used in practice, we envision the following hypothetical future situation:

Using the visualization and interaction tools of Chapter 3, a domain analyst has defined and verified a normal model for vessel behavior. An operator is monitoring a maritime environment for potential pirates. Using the aggregated glyphs of Chapter 4, the operator can perceive all merchant vessels and fishing vessels in close proximity in the area. An automated reasoning engine that reasons on the intent of vessels and on whether vessels are involved in piracy raises an alarm to the operator, which the operator can quickly perceive due to the glyphs of Chapter 5. This reasoning engine uses, among other things, a model of normal behavior as defined by the domain analyst to decide the suspect vessel is behaving anomalously. The operator now knows which vessel is a potential threat and where it is. To understand why it is a potential

threat the operator uses our rationale visualization of Chapter 6. Having verified the vessel is indeed likely a pirate using said visualization, the operator wants to know how big the threat is by investigating what the vessel might do next. To this end, the operator uses the prediction visualization discussed in Chapter 7. The prediction model is based on behavioral models whose design and verification have been aided by the visualization and interaction tools discussed in Chapter 3. Through the prediction visualization, the operator can see that the suspected pirate can potentially strike a merchant vessel in about 30 minutes. The operator has now perceived and comprehended the situation and has projected it into the future. As a response, the operator sends in a patrol vessel to prevent the pirate attack.

8.2.1 Applicability

Despite the emphasis on the Maritime safety security domain, we have shown most of our visualization solutions are also applicable to other domains.

We have shown that our visualization and interaction tools for traffic analysis as discussed in Chapter 3 are also applicable to air traffic. Moreover, we argue that said tools are applicable to any kind of moving object data and most suitable for moving objects that have factors in common, such as destination or rules governing traffic, but do not have a functional relationship.

The non-overlapping aggregated multivariate glyphs of Chapter 4 can be applied to any moving objects visualization that does not aggregate over time, such as in an operational view where moving objects are being tracked in real-time, or in a scatter plot where the user can browse through a timeline. The glyph design of Chapter 5 is specific to the maritime safety and security domain. Lessons learned can, however, be applied to any glyph design in any domain.

The rationale visualization discussed in Chapter 6 is mainly intended for safety and security systems, such as the Maritime safety and security domain this thesis focusses on, and other law enforcement domains. We argue, however, that the rationale visualization can be used to explain the rationale of any automated reasoner in a decision-support system in various contexts, such as machine diagnostics or medical diagnostics. The main requirement is that an automated reasoning system is available that can reason on evidence from heterogeneous and possibly conflicting sources. Additionally, the evidence is expected to contain attributes with varying domains.

Finally, we have shown that the prediction visualization of Chapter 7 is applicable to other law enforcement domains as well, for example where the possible future positions of a suspect individual are predicted. We further argue that this method is applicable to any prediction of movement or interaction of moving objects in other domains such as ecology. The only requirement is that a movement model is available to model the mobility of the objects of interest.

We can conclude that the work in this thesis is applicable to a broader context of safety and security systems and in some cases to an even broader context of moving object analysis and visualization in general. The work in this thesis, especially Chapters 5, 6, and 7, has partly been integrated into an existing maritime safety and security product as a technology demonstrator and was received well by domain experts and potential users.

8.2.2 Scalability

In general, scalability is a major factor in visualization [222]. In this section we discuss the scalability of the research discussed in this thesis in some more detail with respect to the following factors [62]–see Figure 8.2:

- **Human perception**: the way the human user processes the visualization;
- **Display constraints**, such as the size or resolution of the screen;
- **The visual metaphor**: the way information is mapped to the visual display, *e.g.*, the visual design choices;
- **Computational Complexity** and efficiency of algorithms, data structures and other computational infrastructure.

As can be seen in Figure 8.2, the scalability issues vary as we approach situational awareness from different angles across the different chapters. Each unique challenge requires its own unique solution and therefore has no scalability issues in common with the other solutions.

Traffic Flow Analysis

For the traffic flow analysis of Chapter 3, the number of particles forms the major limiting factor in scalability and does so in both a computational and perceptual sense. We find that, on modern hardware, the particle system runs smoothly for both the vessel and the aircraft data set we studied. When the number of trajectories in the data set increases, the number of particles per trajectory may need to be reduced to avoid traffic flows becoming invisible due to clutter in particles. This means the computational complexity does not relate to the number of trajectories in a straightforward way. The effect of the number of trajectories on scalability may need further investigation.

Glyphs

For the aggregated glyphs of Chapter 4, the number of moving objects is not so much a limiting factor in computational complexity as is the density, or the degree of overlap between the visual representations of the moving objects. The glyphs designed in Chapter 5 may suffer under visual clutter as the number of moving objects visualized on screen increases. This can, however, be solved by partitioning the moving objects into subsets as in Chapter 4. For both glyphs, scalability issues in the visual metaphor may arise if additional attributes need to be visualized. There are a limited number of visual parameters and viable

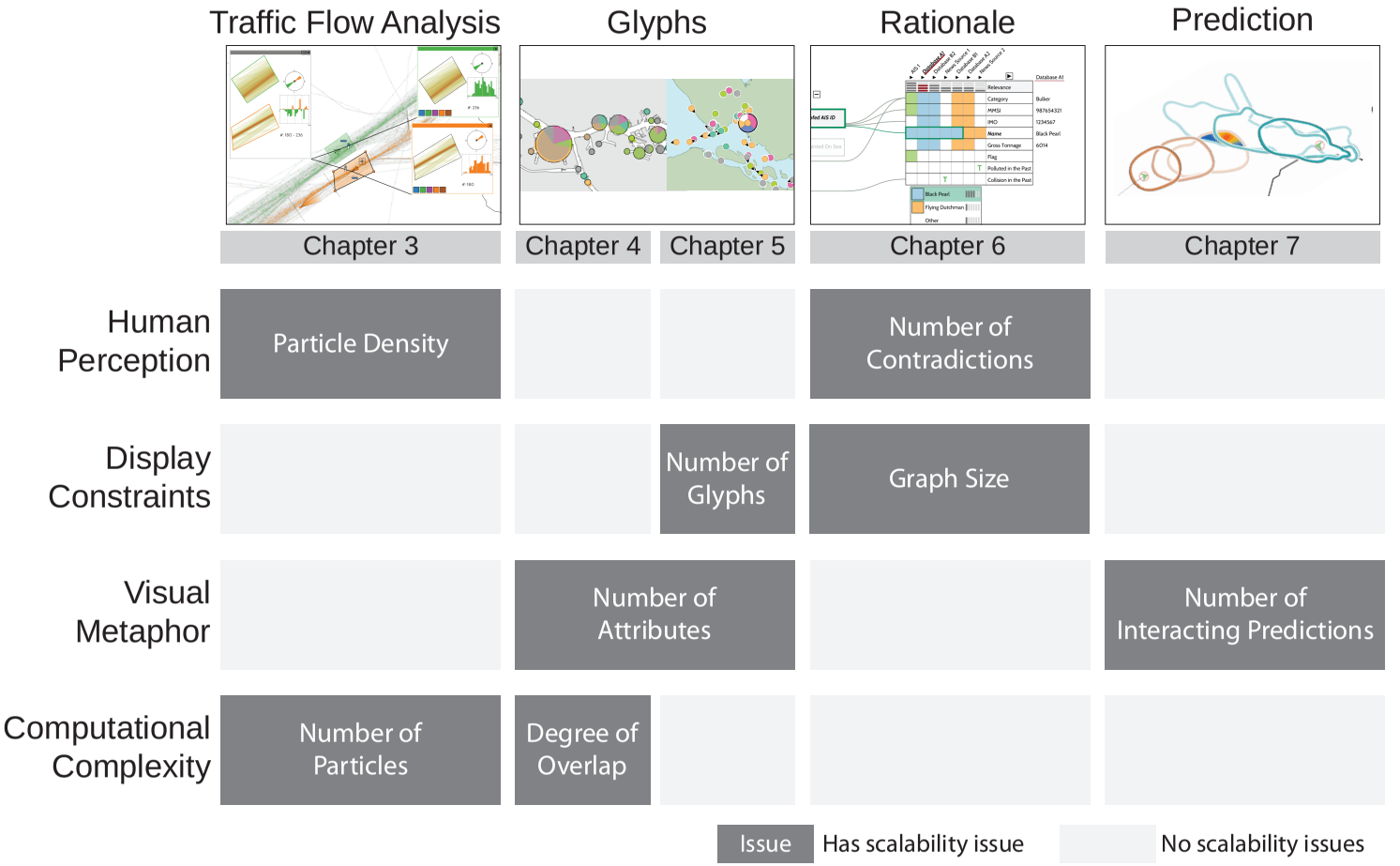


Figure 8.2: An overview of the scalability issues in the work presented in this thesis over four factors: Human perception, display constraints, the visual metaphor, and computational complexity.

combinations of these parameters that can be used to clearly convey attributes, especially aggregated attributes, in a single glyph.

Rationale

The major scalability factor for our rationale visualization of Chapter 6 is the display. The visualization scales with both the number of hypotheses and the number of sources and attributes in the evidence matrix. The size of the evidence matrix can be reduced significantly by reducing the size of the cells, however, the color differences need to remain perceivable [186]. We have already shown that the visual representation of the hypotheses can be reduced in size. For larger graphs these can be reduced further in size. There is, however, a limit to the size of the rationale that our visualization can handle in one display. In these cases hierarchical visualization methods or semantic zoom may prove beneficial. The scalability of the number of conflicts in the evidence matrix is limited by human perception, as the number of different colors that can be distinguished is limited [62].

Prediction

The computational complexity of the prediction visualization discussed in Chapter 7 depends on the complexity of the underlying movement model. The major scalability issue of this method, however, is in the visual metaphor. While the interaction of two predictions is already not straightforward to interpret, the interaction of three or more predictions becomes problematic and may require alternative methods. We have not found any practical case in the maritime domain in which the interaction between three or more objects is of interest. This may, however, be more common in other domains, such as ecology.

8.3 Looking Forward

Even though we have gone a long way in answering our research question, there is still much to be done. Suggestions for future work for each of the discussed methods individually can be found at the end of their respective chapters. In this section we describe a number of opportunities for future work to further aid the operator in acquiring and maintaining situational awareness in general.

Combining Glyphs

We have shown a method for dealing with clutter and overlap of glyphs and a glyph especially designed for the maritime domain to display uncertain distributions and attract attention when required. The next step is to combine these two methods into a single approach. For example, the visual representations of glyphs that do not require attention could be partitioned and visualized according to the approach described in Chapter 4, while the vessels that require attention could be visualized using the attention attracting glyph of Chapter 5. This makes sense as the basic, non-attention attracting glyph of our specialized maritime glyph of Chapter 5 strongly resembles the basic glyphs of Chapter 4. However, the more salient and larger glyph intended to attract attention may no longer have the de-

sired effect as the aggregated glyphs also become larger and more salient as they represent more vessels. This requires additional research.

Extracting Domain Knowledge

One of the main challenges for designing visualizations for situational awareness is extracting domain knowledge from operators and domain experts [144]. This is also a major challenge for the ongoing development of rule-based automated methods aimed at supporting situational awareness. This is referred to as the *knowledge acquisition bottleneck* [112]. Visualization could be used as a tool to *discover* this domain knowledge and communicate it to others. An option could be to let domain experts visually explore large volumes of data of real maritime situations and mark what is important and what is not. This would require a significant amount of research.

Growing volume of Information

The growing gap between available information and the available human resources gives rise to a number of challenges. Not all information is pushed by sensors, but a lot of relevant information is available elsewhere, waiting to be found. This information needs to be found, retrieved, and processed in a reasonable amount of time. Additionally, to make sense of the information, and mostly to filter what is relevant for the operator, automated reasoning has to be applied to this information. As the available information grows in number and volume, this costs an increasing amount of resources. A major research challenge remains in integrating information retrieval and reasoning into a situational awareness system such that it remains functional and responsive.

Scalability

The scalability issues discussed in Section 8.2.2 do not present a problem in the Maritime Safety and Security Domain. When our visualizations are applied to other domains, however, these scalability issues may become problematic. Additional research is required to find where and how the individual scalability issues may become problematic in other domains and what can be done to resolve them.

References

[1] R. Agrawal, K.-I. Lin, H. S. Sawhney, and K. Shim. Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In *Proceedings of the 21th International Conference on Very Large Data Bases*, pages 490–501, San Francisco, CA, USA, 1995.

[2] G. Andrienko and N. Andrienko. Visual exploration of the spatial distribution of temporal behaviors. In *Proceedings of the 9th International Conference on Information Visualisation (IV)*, pages 799–806, July 2005.

[3] G. Andrienko and N. Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 51–58, Oct. 2008.

[4] G. Andrienko, N. Andrienko, J. Dykes, S. I. Fabrikant, and M. Wachowicz. Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. *Information Visualization*, 7(3-4):173–180, 2008.

[5] G. Andrienko, N. Andrienko, and M. Heurich. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, 25(9):1347–1370, 2011.

[6] G. Andrienko, N. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel. Scalable analysis of movement data for extracting and exploring significant places. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1078–1094, July 2013.

[7] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti. Interactive visual clustering of large collections of trajectories. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 3–10, oct. 2009.

[8] G. Andrienko, N. Andrienko, and S. Wrobel. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorer Newsletter*, 9(2):38–46, december 2007.

[9] N. Andrienko and G. Andrienko. Designing visual analytics methods for massive collections of movement data. *Cartographica*, 42:117—138, 2007.

[10] N. Andrienko and G. Andrienko. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, 17(2):205–219, feb 2011.

[11] N. Andrienko and G. Andrienko. A visual analytics framework for spatio-temporal analysis and modelling. *Data Mining and Knowledge Discovery*, 27(1):55–83, 2013.

[12] N. Andrienko and G. Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.

[13] A. Appel, F. J. Rohlf, and A. J. Stein. The haloed line effect for hidden line elimination. *SIGGRAPH Computer Graphics*, 13(2):151–157, Aug. 1979.

[14] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A review of temporal data visualizations based on space-time cube operations. In *Eurographics State of the Art Reports*, EG STARs. Eurographics Association, 2014.

[15] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. The state of the art in visualizing dynamic graphs. In *Eurographics State of the Art Reports*, EG STARs. Eurographics Association, 2014.

[16] K. Been, E. Daiches, and C. Yap. Dynamic map labeling. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):773–780, Sep 2006.

[17] P. Bereuter and R. Weibel. Generalisation of point data for mobile devices: A problem-oriented approach. In *Proceedings of the 13th Workshop on Progress in Generalisation and Multiple Representation*, pages 1–8, Zurich, Switzerland, 2010.

[18] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *Proceedings Knowledge Discovery and Delivery Workshop*, pages 359–370, 1994.

[19] J. Bertin. *Semiology of Graphics*. University of Wisconsin Press, Madison, WI, 1983.

[20] E. Bertini and G. Santucci. By chance is not enough: preserving relative density through nonuniform sampling. In *Proceedings of the 8th International Conference on Information Visualisation (IV)*, pages 622–629, 2004.

[21] E. Bertini and G. Santucci. Improving 2d scatterplots effectiveness through sampling, displacement, and user perception. In *Proceedings of the 9th International Conference on Information Visualisation (IV)*, pages 826–834, 2005.

[22] J. Blaas, C. Botha, E. Grundy, M. Jones, R. Laramee, and F. Post. Smooth graphs for visual exploration of higher-order state transitions. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):969–976, Nov 2009.

[23] N. Bomberger, B. Rhodes, M. Seibert, and A. Waxman. Associative learning of vessel motion patterns for maritime situation awareness. In *9th International Conference on Information Fusion (FUSION)*, pages 1–8, July 2006.

[24] S. C. Boraz. Maritime domain awareness: Myths and realities. *Naval War College Review*, 62(3), 2009.

[25] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics State of the Art Reports*, EG STARs, pages 39–63. Eurographics Association, May 2013. http://diglib.eg.org/EG/DL/conf/EG2013/stars/039-063.pdf.

[26] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2271–2280, Dec 2014.

[27] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, Dec 2013.

[28] S. Bruckner and M. E. Gröller. Style transfer functions for illustrative volume rendering. *Computer Graphics Forum*, 26(3):715–724, sep 2007.

[29] B. Buchanan and E. Shortliffe. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, MA, 1984.

[30] K. Buchin, B. Speckmann, and K. Verbeek. Flow map layout via spiral trees. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2536–2544, Dec 2011.

[31] M. Buchin, S. Dodge, and B. Speckmann. Context-aware similarity of trajectories. In N. Xiao, M.-P. Kwan, M. Goodchild, and S. Shekhar, editors, *Geographic Information Science*, volume 7478 of *Lecture Notes in Computer Science*, pages 43–56. Springer Berlin Heidelberg, 2012.

[32] F. Bullard. *Estimating the Home Range of an Animal: A Brownian bridge Approach (Thesis)*. Phd thesis, University of North Carolina, Chapel Hill, North Carolina, USA, 1999.

[33] S. Buschmann, M. Trapp, and J. Dollner. Real-time animated visualization of massive air-traffic trajectories. In *Proceedings of the International Conference on Cyberworlds (CW)*, pages 174–181, Oct 2014.

[34] A. Cairo. *The Functional Art: An introduction to information graphics and visualization*. New Riders, 2012.

[35] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[36] M. B. Carmo, A. P. Afonso, P. Pombinho De Matos, and A. Vaz. Movisys - a visualization system for geo-referenced information on mobile

devices. In *Proceedings of the 10th International Conference on Visual Information Systems: Web-Based Visual Information Search and Management*, pages 167–178, Berlin, Heidelberg, 2008. Springer-Verlag.

[37] M. A. Carmody and J. P. Gluckman. Task specific effects of automation and automation failure on performance, workload and situational awareness. In *Proceedings of the Seventh International Symposium on Aviation Psychology*, volume 1, pages 167–171, 1993.

[38] M. S. T. Carpendale. Considering visual variables as a basis for information visualisation. Technical report, University of Calgary, January 2003.

[39] F. Chazal, D. Chen, L. Guibas, X. Jiang, and C. Sommer. Data-driven trajectory smoothing. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 251–260, Chicago, Illinois, 2011.

[40] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(342):361–368, 1973.

[41] E. Chlan and P. Rheingans. Multivariate glyphs for multi-object clusters. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 141–148, oct. 2005.

[42] R. E. Christ. Review and analysis of color coding research for visual displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 17(6):542–570, 1975.

[43] W. S. Cleveland. *Visualizing Data*. Hobart Press, Summit, New Jersey, U.S.A., 1993.

[44] A. Cockburn, A. Karlson, and B. B. Bederson. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Computing Surveys (CSUR)*, 41(1):2:1–2:31, January 2009.

[45] A. Colmant and J. Van Vuuren. Prerequisites for the design of a maritime law enforcement resource assignment decision support system. In *Proceedings of the 42nd Annual ORSSA Conference*, pages 90–101, 2013.

[46] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, Dec 2014.

[47] M. De Berg, P. Bose, O. Cheong, and P. Morin. On simplifying dot maps. *Computational Geometry*, 27(1):43–62, 2004.

[48] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf. *Computational Geometry: Algorithms and Applications*. Springer-Verlag, 2nd edition, 2000.

[49] M. de Oliveira and H. Levkowitz. From visual data exploration to visual data mining: a survey. *IEEE Transactions on Visualization and Computer Graphics*, 9(3):378–394, July 2003.

[50] G. de Vries, W. van Hage, and M. van Someren. Comparing vessel trajectories using geographical domain knowledge and alignments. In *Proceedings of the IEEE International Conference on Data Mining Workshops*, pages 209–216, 2010.

[51] U. Demšar and K. Virrantaus. Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, 24(10):1527–1542, 2010.

[52] A. Dix and G. Ellis. By chance enhancing interaction with large data sets through statistical sampling. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 167–176, Trento, Italy, 2002.

[53] S. Dodge, R. Weibel, and E. Forootan. Revealing the physics of movement: comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, 33(6):419–434, November 2009.

[54] S. Dodge, R. Weibel, and A.-K. Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, 7(3-4):240–252, 2008.

[55] W. Dommelen, P. van de Laar, and L. Noldus. Extending track analysis from animals in the lab to moving objects anywhere. In P. van de Laar, J. Tretmans, and M. Borth, editors, *Situation Awareness with Systems of Systems*, pages 89–103. Springer New York, 2013.

[56] S. dos Santos and K. Brodlie. Gaining understanding of multivariate and multidimensional data through visualization. *Computers & Graphics*, 28(3):311–325, 2004.

[57] M. J. Druzdzel and M. Henrion. Using scenarios to explain probabilistic inference. *Working notes of the AAAI-90 Workshop on Explanation*, pages 133–141, 1990.

[58] J. Dykes, A. M. MacEachren, and M.-J. Kraak. *Exploring geovisualization*. Elsevier, 2005.

[59] P. Eades, W. Lai, K. Misue, and K. Sugiyama. Preserving the mental map of a diagram. In *Proceedings of the COMPUGRAPHICS conference*, pages 34–43, 1991.

[60] J. Edlund, M. Grönkvist, A. Lingvall, and E. Sviestins. Rule-based situation assessment for sea surveillance. In *Proceedings SPIE*, volume 6242, pages 624203–624203–11, 2006.

[61] A. Edwardes, D. Burghardt, and R. Weibel. Portrayal and Generalisation of Point Maps for Mobile Information Services. *Map-based Mobile Services*, pages 11–30, 2005.

[62] S. G. Eick and A. F. Karr. Visual scalability. *Journal of Computational and Graphical Statistics*, 11(1):22–43, 2002.

[63] G. Ellis, E. Bertini, and A. Dix. The sampling lens: making sense of saturated visualisations. In *Extended Abstracts on Human Factors in Computing Systems*, pages 1351–1354, Portland, OR, USA, 2005.

[64] G. Ellis and A. Dix. Density control through random sampling: an architectural perspective. In *Proceedings of the 6th International Conference on Information Visualisation (IV)*, pages 82–90, 2002.

[65] G. Ellis and A. Dix. The plot, the clutter, the sampling and its lens: occlusion measures for automatic clutter reduction. In *Proceedings of the working Conference on Advanced visual interfaces*, AVI '06, pages 266–269, New York, NY, USA, 2006. ACM.

[66] N. Elmqvist, P. Dragicevic, and J. Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1539–1148, Nov 2008.

[67] Embedded Systems Innovations (ESI). Metis: Dependable cooperative systems for maritime safety and security. http://www.esi.nl/research/research-in-projects/metis.dot, 2015.

[68] A. Endert, P. Fiaux, and C. North. Semantic interaction for sensemaking: Inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2879–2888, Dec 2012.

[69] M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

[70] M. R. Endsley. Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making*, 9(1):4–32, 2015.

[71] M. R. Endsley and C. A. Bolstad. Human capabilities and limitations in situation awareness. In *Combat automation for airborne weapon systems: Man/machine interface trends and technologies (AGARD-CP-520)*, pages 19/1–19/10, 1993.

[72] M. R. Endsley and E. S. Connors. Situation awareness: State of the art. In *IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–4, 2008.

[73] M. R. Endsley and W. M. Jones. A model of inter- and intrateam situation awareness: Implications for design, training and measurement. *New trends in cooperative activities: Understanding system dynamics in complex environments*, 7:46–67, 2001.

[74] L. Etienne, T. Devogele, and A. Bouju. Spatio-temporal trajectory analysis of mobile objects following the same itinerary. In *Proceedings of the International Symposium on Spatial Data Handling*, page 6, 2010.

[75] M. Everts, H. Bekker, J. Roerdink, and T. Isenberg. Depth-dependent halos: Illustrative rendering of dense line data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1299–1306, nov.-dec. 2009.

[76] Fairplay. Maritime intelligence and publications: IHS Fairplay. http://www.ihs.com/products/maritime-information/index.aspx?pu=1&rd=ihsfairplay_com, july 2013.

[77] D. Feng, L. Kwock, Y. Lee, and R. Taylor. Matching visual saliency to confidence in plots of uncertain data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):980–989, 2010.

[78] N. Ferreira, J. Poco, H. Vo, J. Freire, and C. Silva. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158, Dec 2013.

[79] Y. Fischer and A. Bauer. Object-oriented sensor data fusion for wide maritime surveillance. In *International Waterside Security Conference (WSS)*, pages 1–6, Nov 2010.

[80] J. Gao, L. J. Guibas, J. Hershberger, L. Zhang, and A. Zhu. Discrete mobile centers. In *Discrete and Computational Geometry*, pages 188–196. ACM Press, 2001.

[81] M. Glandrup. Improving situation awareness in the maritime domain. In P. van de Laar, J. Tretmans, and M. Borth, editors, *Situation Awareness with Systems of Systems*, pages 89–103. Springer New York, 2013.

[82] M. Green. Toward a perceptual science of multidimensional data visualization: Bertin and beyond. *ERGO/GERO Human Factors Science*, 8, 1998.

[83] D. Guo. Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1041–1048, Nov 2009.

[84] D. Guo and X. Zhu. Origin-destination flow data smoothing and mapping. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2043–2052, Dec 2014.

[85] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan. Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *Proceedings of the 4th IEEE Pacific Visualization Symposium (PacificVis 2011)*, pages 163–170, 2011.

[86] R. B. Haber and D. A. McNabb. Visualization idioms: A conceptual model for scientific visualization systems. *Visualization in scientific computing*, 74:93, 1990.

[87] S. J. Hansen. Piracy in the greater Gulf of Aden. Technical report, Norwegian Institute for Urban and Regional Research, London, UK, 2009.

[88] S. Har-Peled. Clustering motion. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 84–93, oct. 2001.

[89] M. Harrower and C. A. Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.

[90] H. Hauser, F. Ledermann, and H. Doleisch. Angular brushing of extended parallel coordinates. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 127–130, 2002.

[91] C. G. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, July 2012.

[92] J. S. Helmick. Port and maritime security: A research perspective. *Journal of Transportation Security*, 1(1):15–28, 2008.

[93] T. Hendriks and P. van de Laar. Metis: Dependable cooperative systems for public safety. *Procedia Computer Science*, 16:542–551, 2013.

[94] J. Hoeksema and W. R. van Hage. A case study on automated risk assessment of ships using newspaper-based event extraction. *Detection, Representation, and Exploitation of Events in the Semantic Web*, 1123:42–52, 2013.

[95] R. Hoffman. Origins of situation awareness: Cautionary tales from the history of concepts of attention. *Journal of Cognitive Engineering and Decision Making*, 9(1):73–83, 2015.

[96] J. Hollands and I. Spence. The discrimination of graphical elements. In *Applied Cognitive Psychology*, pages 413–431, 2001.

[97] J. S. Horne, E. O. Garton, S. M. Krone, and J. S. Lewis. Analyzing animal movements using Brownian bridges. *Ecology*, 88(9):2354–2363, 2007.

[98] B. Huijbrechts, M. Velikova, R. Scheepens, and S. Michels. Metis: An integrated reference architecture for addressing uncertainty in decision-support systems. *Procedia Computer Science*, 44:476–485, 2015.

[99] C. Hurter, R. Alligier, D. Gianazza, S. Puechmorel, G. Andrienko, and N. Andrienko. Wind parameters extraction from aircraft trajectories. *Computers, Environment and Urban Systems*, 47(0):28–43, 2014.

[100] C. Hurter, O. Ersoy, S. Fabrikant, T. Klein, and A. Telea. Bundled visualization of dynamicgraph and trail data. *IEEE Transactions on Visualization and Computer Graphics*, 20(8):1141–1157, Aug 2014.

[101] C. Hurter, O. Ersoy, and A. Telea. Moleview: An attribute and structure-based semantic lens for large element-based plots. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2600–2609, dec 2011.

[102] C. Hurter, O. Ersoy, and A. Telea. Graph bundling by kernel density estimation. *Computer Graphics Forum*, 31(3):865–874, 2012.

[103] C. Hurter, O. Ersoy, and A. Telea. Smooth bundling of large streaming and sequence graphs. In *Proceedings of the 6th IEEE Pacific Visualization Symposium (PacificVis 2013)*, pages 41–48, Feb 2013.

[104] C. Hurter, B. Tissoires, and S. Conversy. FromDaDy: Spreading aircraft trajectories across views to support iterative queries. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1017–1024, Nov 2009.

[105] ICC International Maritime Bureau (IMB). Piracy and armed robbery against ships: Annual report 2006. Barking: ICC Publishing, 2006.

[106] A. Inselberg and B. Dimsdale. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st Conference on Visualization*, pages 361–378, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

[107] International Maritime Organization (IMO). International Convention for the Prevention of Pollution From Ships (MARPOL). Signed on 17 February 1973, entered into force 2 October 1983, 1972.

[108] International Maritime Organization (IMO). International Regulations for Preventing Collisions at Sea 1972 (COLREGs). Signed on 20 Oktober 1972, entered into force 15 July 1977, 1972.

[109] International Maritime Organization (IMO). International ship & port facility security (ISPS) code, 2003 and december 2002 amendments to SOLAS. Technical Report IMO-I116E, IMO, London, 2003.

[110] International Maritime Organization (IMO). *SOLAS 2014: Consolidated Text of the International Convention for the Safety of Life at Sea, 1974, as Amended*. International Maritime Organization, Sep 2014.

[111] International Telecommunication Union (ITU). Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile band. *Recommendation ITU-R M.1371-4*, 2010.

[112] P. Jackson. *Introduction to expert systems*. Addison-Wesley Publishing Company, Reading, MA, 3rd edition, 1999.

[113] W. Javed and N. Elmqvist. Exploring the design space of composite visualization. In *Proceedings of the 5th IEEE Pacific Visualization Symposium (PacificVis 2012)*, pages 1–8. IEEE, 2012.

[114] W. Javed, B. McDonnel, and N. Elmqvist. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):927–934, Nov 2010.

[115] A. Kaufman and K. Mueller. Overview of volume rendering. *The visualization handbook*, pages 127–174, 2005.

[116] B. Köbben and M. Yaman. Evaluating dynamic visual variables. In *Proceedings of the seminar on teaching animated cartography, ACI/ICA, Madrid*, pages 45–51, 1996.

[117] R. Kosara, S. Miksch, and H. Hauser. Semantic depth of field. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 97–104, 2001.

[118] A. Koussoulakou and M.-J. Kraak. Spatio-temporal maps and cartographic communication. *The Cartographic Journal*, 29(2):101–108, 1992.

[119] B. Kranstauber, R. Kays, S. D. LaPoint, M. Wikelski, and K. Safi. A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology*, 81(4):738–46, 2012.

[120] J. M. Krisp and S. Peters. Directed kernel density estimation (DKDE) for time series visualization. *Annals of GIS*, 17(3):155–162, 2011.

[121] R. Krüger, D. Thom, M. Wörner, H. Bosch, and T. Ertl. TrajectoryLenses – a set-based filtering and exploration technique for long-term trajectory data. *Computer Graphics Forum*, 32(3):451–460, 2013.

[122] B. Kuijpers, H. J. Miller, T. Neutens, and W. Othman. Anchor uncertainty and space-time prisms on road networks. *International Journal of Geographical Information Science*, 24(8):1223–1248, 2010.

[123] B. Kuijpers, H. J. Miller, and W. Othman. Kinetic space-time prisms. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 162–170, Chicago, IL, 2011.

[124] O. Lampe and H. Hauser. Interactive visualization of streaming data with kernel density estimation. In *Proceedings of the 4th IEEE Pacific Visualization Symposium (PacificVis 2011)*, pages 171–178, march 2011.

[125] R. O. Lane, D. A. Nevell, S. D. Hayward, and T. W. Beaney. Maritime anomaly detection and threat assessment. In *13th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2010.

[126] V. Lavigne, D. Gouin, and M. Davenport. Visual analytics for maritime domain awareness. In *Proceedings of the IEEE International Conference on Technologies for Homeland Security*, pages 49–54, nov. 2011.

[127] J. Li, J.-B. Martens, and J. J. van Wijk. A model of symbol size discrimination in scatterplots. In *Proceedings 28th International Conference on Human Factors in Computing Systems*, pages 2553–2562, 2010.

[128] A. M. MacEachren. *How maps work: representation, visualization, and design*. Guilford Press, 1995.

[129] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141, Apr. 1986.

[130] D. Madigan, K. Mosurski, and R. G. Almond. Graphical explanation in belief networks. *Journal of Computational and Graphical Statistics*, 6(2):160–181, 1997.

[131] E. Maguire, P. Rocca-Serra, S.-A. Sansone, J. Davies, and M. Chen. Taxonomy-based glyph design–with a case study on visualizing workflows of biological experiments. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2603–2612, 2012.

[132] A. Malik, R. Maciejewski, B. Maule, and D. Ebert. A visual analytics process for maritime resource allocation and risk assessment. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 221–230, Oct 2011.

[133] M. L. Matthews, L. Rehak, A.-L. S. Lapinski, and S. McFadden. Improving the maritime surface picture with a visualization aid to provide rapid situation awareness of information uncertainty. In *Proceedings of the IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, pages 533–538, sept. 2009.

[134] T. McLoughlin, R. S. Laramee, R. Peikert, F. H. Post, and M. Chen. Over two decades of integration-based, geometric flow visualization. *Computer Graphics Forum*, 29(6):1807–1829, 2010.

[135] W. Merzkirch. *Flow visualization*. Elsevier, 2012.

[136] L. Micallef, P. Dragicevic, and J. Fekete. Assessing the effect of visualizations on Bayesian reasoning through crowdsourcing. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2536–2545, Dec 2012.

[137] S. Michels, M. Velikova, A. Hommersom, and P. J. Lucas. A probabilistic logic-based model for fusing attribute information of objects under surveillance. Technical Report ICIS–R12006, Radboud University Nijmegen, Nijmegen, Netherlands, 2012.

[138] S. Michels, M. Velikova, A. Hommersom, and P. J. Lucas. A decision support model for uncertainty reasoning in safety and security tasks. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 663–668. IEEE, 2013.

[139] S. Michels, M. Velikova, and P. Lucas. Probabilistic model-based assessment of information quality in uncertain domains. In M. Thielscher and D. Zhang, editors, *AI 2012: Advances in Artificial Intelligence*, volume 7691 of *Lecture Notes in Computer Science*, pages 890–901. Springer Berlin Heidelberg, 2012.

[140] M. Migut and M. Worring. Visual exploration of classification models for risk assessment. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 11–18. IEEE, 2010.

[141] S. Mittelstädt and D. A. Keim. Efficient Contrast Effect Compensation with Personalized Perception Models. *Computer Graphics Forum*, 34(3):211–220, 2015.

[142] K. E. Moffitt. *An Empirical Test of Expert System Explanation Facility Effects on Incidental Learning and Decision-making*. Phd thesis, Arizona State University, Tempe, AZ, USA, 1989. AAI9018520.

[143] P. Muigg, M. Hadwiger, H. Doleisch, and E. Gröller. Visual coherence for large-scale line-plot visualizations. *Computer Graphics Forum*, 30(3):643–652, 2011.

[144] M. Nilsson, J. van Laere, T. Ziemke, and J. Edlund. Extracting rules from expert operators to support situation awareness in maritime surveillance. In *11th International Conference on Information Fusion (FUSION)*, pages 1–8, June 2008.

[145] D. A. Norman. A psychologist views human processing: Human errors and other phenomena suggest processing mechanisms. In *Proceedings of the 7th international joint conference on Artificial intelligence-Volume 2*, pages 1097–1101. Morgan Kaufmann Publishers Inc., 1981.

[146] P. Novák and C. Witteveen. Reconfiguration of large-scale surveillance systems. In J. Leite, T. Son, P. Torroni, L. Torre, and S. Woltran, editors, *Computational Logic in Multi-Agent Systems*, volume 8143 of *Lecture Notes in Computer Science*, pages 325–339. Springer Berlin Heidelberg, 2013.

[147] P. Novák and C. Witteveen. Context-aware reconfiguration of large-scale surveillance systems: argumentative approach. *Argument & Computation*, 6(1):3–23, 2015.

[148] H. Ögmen, T. U. Otto, and M. H. Herzog. Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research*, 46(19):3234–3242, 2006.

[149] F. Onuoha. Sea piracy and maritime security in the Horn of Africa: The Somali coast and Gulf of Aden in perspective. *African Security Review*, 18(3):31–44, 2009.

[150] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[151] T. Pham, R. Hess, C. Ju, E. Zhang, and R. A. Metoyer. Visualization of diversity in large multivariate data sets. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1053–1062, 2010.

[152] W. Pike, R. May, B. Baddeley, R. Riensche, J. Bruce, and K. Younkin. Scalable visual reasoning: Supporting collaboration through distributed analysis. In *Proceedings of the International Symposium on Collaborative Technologies and Systems (CTS)*, pages 24–32, May 2007.

[153] G. Pilato, A. Augello, M. Missikoff, and F. Taglino. Integration of ontologies and Bayesian networks for maritime situation awareness. In *Proceedings of the IEEE Sixth International Conference on Semantic Computing (ICSC)*, pages 170–177, Sept 2012.

[154] Publieke Dienstverlening Op de Kaart (PDOK). Het Nationaal Georegister (NGR). http://nationaalgeoregister.nl/, 2015.

[155] C. Z. Raymand and A. Morriën. Security in the maritime domain and its evolution since 9/11. In P. L. Rupert Herbert-Burns, Sam Bateman, editor, *Lloyd's MIU Handbook of Maritime Security*, pages 3–12. CRC Press, 2008.

[156] B. Rhodes, N. Bomberger, M. Seibert, and A. Waxman. Maritime situation monitoring and awareness using learning mechanisms. In *Proceedings of the IEEE Military Communications Conference (MILCOM)*, volume 1, pages 646–652, Oct 2005.

[157] B. Rhodes, N. Bomberger, and M. Zandipour. Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness. In *10th International Conference on Information Fusion (FUSION)*, pages 1–8, July 2007.

[158] P. Riehmann, M. Potthast, B. Stein, and B. Froehlich. Visual Assessment of Alleged Plagiarism Cases. *Computer Graphics Forum*, 34(3):061–070, 2015.

[159] B. Ristic, B. La Scala, M. Morelande, and N. Gordon. Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. In *11th International Conference on Information Fusion (FUSION)*, pages 1–7. IEEE, 2008.

[160] M. Riveiro. Evaluation of normal model visualization for anomaly detection in maritime traffic. *ACM Transactions on Interactive Intelligent Systems*, 4(1):5:1–5:24, Apr. 2014.

[161] M. Riveiro and G. Falkman. Interactive visualization of normal behavioral models and expert rules for maritime anomaly detection. In *Sixth International Conference on Computer Graphics, Imaging and Visualization (CGIV)*, pages 459–466, Aug 2009.

[162] M. Riveiro and G. Falkman. Evaluating the usability of visualizations of normal behavioral models for analytical reasoning. In *7th International*

Conference on Computer Graphics, Imaging and Visualization, pages 179–185, aug. 2010.

[163] M. Riveiro, G. Falkman, and T. Ziemke. Improving maritime anomaly detection and situation awareness through interactive visualization. In *11th International Conference on Information Fusion (FUSION)*, Cologne, Germany, 2008.

[164] M. Riveiro, T. Helldin, and G. Falkman. Influence of meta-information on decision-making: Lessons learned from four case studies. In *Proceedings of the IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pages 14–20, March 2014.

[165] R. Rosenholtz, Y. Li, and L. Nakano. Measuring visual clutter. *Journal of Vision*, 7(2):17, 2007.

[166] R. E. Roth. Visual variables. *The International Encyclopedia of Geography: People, The Earth, Environment, and Technology*, [To appear], 2015.

[167] J. Roy. Anomaly detection in the maritime domain. In *Proceedings of the SPIE Defense and Security Symposium*, volume 6945, pages 69450W–69450W–14, 2008.

[168] C. Santiago, A. Rusu, L. Falconi, B. Petzinger, and A. Crowell. Overview and flight-by-flight analysis of trajectory prediction systems using a 2d galaxy visualization. In *IEEE/AIAA 28th Digital Avionics Systems Conference*, pages 3.B.3–1–3.B.3–8, oct. 2009.

[169] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. J. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1421–1430, nov.-dec. 2010.

[170] R. Scheepens, C. Hurter, H. van de Wetering, and J. J. van Wijk. Visualization, selection, and analysis of traffic flows. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):[To Appear], January 2016.

[171] R. Scheepens, S. Michels, H. van de Wetering, and J. J. van Wijk. Rationale visualization for decision support [poster]. *IEEE Symposium on Information Visualization*, October 2014.

[172] R. Scheepens, S. Michels, H. van de Wetering, and J. J. van Wijk. Rationale Visualization for Safety and Security. *Computer Graphics Forum*, 34(3):191–200, 2015.

[173] R. Scheepens, H. van de Wetering, and J. J. van Wijk. Contour based visualization of vessel movement predictions. *International Journal of Geographical Information Science*, 28(5):891–909, 2014.

[174] R. Scheepens, H. van de Wetering, and J. J. van Wijk. Design of glyphs for uncertain maritime data [poster]. *Eurographics Conference on Visualization (EuroVis)*, June 2014.

[175] R. Scheepens, H. van de Wetering, and J. J. van Wijk. Non-overlapping aggregated multivariate glyphs for moving objects. In *Proceedings of the 7th IEEE Pacific Visualization Symposium (PacificVis 2014)*, pages 17–24, March 2014.

[176] R. Scheepens, N. Willems, H. van de Wetering, G. Andrienko, N. Andrienko, and J. J. van Wijk. Composite density maps for multivariate trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2518–2527, dec. 2011.

[177] R. Scheepens, N. Willems, H. van de Wetering, and J. J. van Wijk. Density based, visual anomaly detection. *Proceedings Maritime Anomaly Detection (MAD) 2011 Workshop*, pages 11–12, 2011.

[178] R. Scheepens, N. Willems, H. van de Wetering, and J. J. van Wijk. Interactive visualization of multivariate trajectory data with density maps. In *Proceedings of the 4th IEEE Pacific Visualization Symposium (PacificVis 2011)*, pages 147–154, 2011.

[179] R. Scheepens, N. Willems, H. van de Wetering, and J. J. van Wijk. Interactive density maps for moving objects. *IEEE Computer Graphics and Applications*, 32(1):56–66, jan.-feb. 2012.

[180] H. Shahir, U. Glasser, N. Nalbandyan, and H. Wehn. Maritime situation analysis. In *Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 230–232, June 2013.

[181] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, Sep 1996.

[182] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Number 26 in Monographs on Statistics and Applied Probability. Chapman & Hall, 1992.

[183] D. J. Simons and D. T. Levin. Change blindness. *Trends in Cognitive Sciences*, 1(7):261–267, 1997.

[184] D. H. Slade. Meteorology and atomic energy, 1968. Technical report, Environmental Science Services Administration, Silver Spring, Md. Air Resources Labs., 1968.

[185] R. Spence. *Information Visualization, an introduction.* Springer, 3rd edition, 2014.

[186] M. Stone. In color perception, size matters. *IEEE Computer Graphics and Applications*, 32(2):8–13, 2012.

[187] S. Stumpf, V. Rajaram, L. Li, W.-K. Wong, M. Burnett, T. Dietterich, E. Sullivan, and J. Herlocker. Interacting meaningfully with machine learning systems: Three experiments. *International Journal of Human-Computer Studies*, 67(8):639–662, 2009.

[188] H. J. Suermondt and G. F. Cooper. An evaluation of explanations of probabilistic inference. *Computers and Biomedical Research*, 26(3):242–254, 1993.

[189] K. Sugiyama, S. Tagawa, and M. Toda. Methods for visual understanding of hierarchical system structures. *IEEE Transactions on Systems, Man and Cybernetics*, 11(2):109–125, Feb 1981.

[190] B. Tetreault. Use of the automatic identification system (AIS) for maritime domain awareness (MDA). In *Proceedings of MTS/IEEE OCEANS*, volume 2, pages 1590–1594, Sep 2005.

[191] The Dutch Coastguard. Jaarverslag kustwacht 2012. http://www.kustwacht.nl/, 2012.

[192] The Dutch Ministry of Infrastructure and the Environment, and The Dutch Ministry of Economic Affairs. Policy document on the north sea 2016-2021. http://www.noordzeeloket.nl/en/projects/north-sea-policy-in-the-national-water-plan/Beleidsnota_2016_2021/, 2014.

[193] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2565–2574, Dec 2012.

[194] M. Tory and T. Möller. Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):72–84, Jan 2004.

[195] A. Treisman. Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31(2):156–177, 1985.

[196] M. Trutschl, G. Grinstein, and U. Cvek. Intelligently resolving point occlusion. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 131–136, 2003.

[197] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 2nd edition, May 2001.

[198] United Nations Conference on Trade and Development. UNCTAD review of maritime transport. http://unctadstat.unctad.org/, 2014.

[199] United Nations Treaty Series. United Nations Convention on the Law of the Sea (UNCLOS). Opened for signature 10 December 1982, entered into force 16 November 1994, 1984.

[200] S. van den Elzen and J. J. van Wijk. Multivariate network exploration and presentation: From detail to overview via selections and aggregations. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2310–2319, Dec 2014.

[201] S. van der Spek, C. van Langelaar, and C. Christiaan Kickert. Evidence-based design: satellite positioning studies of city centre user groups. In *Proceedings of the ICE-Urban Design and Planning*, volume 166, pages 206–216, 2013.

[202] T. van Gelder. The rationale for Rationale™. *Law, Probability and Risk*, 6(1-4):23–42, 2007.

[203] J. van Laere and M. Nilsson. Evaluation of a workshop to capture knowledge from subject matter experts in maritime surveillance. In *12th International Conference on Information Fusion (FUSION)*, pages 171–178, July 2009.

[204] W. van Norden. *Sensing What Matters*. Phd thesis, Technische Universiteit Delft, 2010.

[205] J. van Wijk. The value of visualization. In *IEEE Visualization*, pages 79–86, Oct 2005.

[206] M. Velikova, P. Novak, B. Huijbrechts, J. Laarhuis, J. Hoeksma, and S. Michels. An integrated reconfigurable system for maritime situational awareness. In *Proceedings of the conference on Prestigious Applications of Intelligent Systems (PAIS)*, 2014.

[207] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh. Indexing multi-dimensional time-series with support for multiple distance measures. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge discovery and Data Mining*, pages 216–225, Washington, D.C., 2003.

[208] P. Šutovský and G. F. Cooper. Hierarchical explanation of inference in Bayesian networks that represent a population of independent agents. In *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI)*, pages 214–218, Amsterdam, The Netherlands, 2008. IOS Press.

[209] A. D. Wall, A. G. Bole, and W. O. Dineley. *Radar and ARPA Manual: Radar and Target Tracking for Professional Mariners, Yachtsmen and Users of Marine Radar*. Butterworth-Heinemann, 2nd edition, 2005.

[210] M. O. Ward. Multivariate data glyphs: Principles and practice. In *Handbook of Data Visualization*, pages 179–198. Springer Berlin Heidelberg, 2008.

[211] M. O. Ward, G. G. Grinstein, and D. A. Keim. *Interactive Data Visualization - Foundations, Techniques, and Applications*. A K Peters, 2010.

[212] C. Ware. Designing with a 2 1/2d attitude. *Information Design Journal*, 10(3):255–262, 2001.

[213] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[214] M. Wijffelaars, R. Vliegen, J. J. van Wijk, and E.-J. van der Linden. Generating color palettes using intuitive parameters. *Computer Graphics Forum*, 27(3):743–750, 2008.

[215] N. Willems. *Visualization of Vessel Traffic*. Phd thesis, Technische Universiteit Eindhoven, 2010.

[216] N. Willems, R. Scheepens, H. van de Wetering, and J. J. van Wijk. Visualization of vessel traffic. In *Situation Awareness with Systems of Systems*, pages 73–87. Springer, 2013.

[217] N. Willems, H. van de Wetering, and J. J. van Wijk. Visualization of vessel movements. *Computer Graphics Forum*, 28(3):959–966, 2009.

[218] N. Willems, H. van de Wetering, and J. J. van Wijk. Evaluation of the visibility of vessel movement features in trajectory visualizations. *Computer Graphics Forum*, 30(3):801–810, 2011.

[219] N. Willems, W. R. van Hage, G. de Vries, J. H. Janssens, and V. Malaisé. An integrated approach for visual analysis of a multisource moving objects knowledge base. *International Journal of Geographical Information Science*, 24(10):1543–1558, 2010.

[220] J. Wood, A. Slingsby, and J. Dykes. Visualizing the dynamics of London's bicycle-hire scheme. *Cartographica*, 46(4):239–251, 2011.

[221] H.-M. Wu, S. Tzeng, and C.-h. Chen. Matrix visualization. In *Handbook of Data Visualization*, pages 681–708. Springer Berlin Heidelberg, 2008.

[222] B. Yost and C. North. The perceptual scalability of visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):837–844, 2006.

[223] K. Zhou, Q. Hou, R. Wang, and B. Guo. Real-time kd-tree construction on graphics hardware. *ACM Transactions on Graphics*, 27(5):126:1–126:11, Dec. 2008.

Summary

Visualization for Maritime Situational Awareness

In the maritime safety and security domain, operators are monitoring the maritime situation for threats to the safety and security of the maritime environment. To be able to do this, the operators require a state of situational awareness, which is defined as: "The **perception** and **comprehension** of the current situation and the **projection** of the current situation into the future." In this thesis, we have attempted to answer the following research question:

> *How can operators be supported in attaining situational awareness using interactive visualization?*

Visualization plays an essential role in aiding operators to acquire and maintain Situational Awareness. We have answered our research question in the following four parts.

Analysis: How can we provide tools to *analyze* and summarize patterns, enabling domain experts to find critical areas and to verify what normal or anomalous behavior is? We have presented a way to gain more domain knowledge by analyzing movement patterns. We do this by visualizing traffic flows and provide interaction tools to support their exploration. The user can intuitively select and filter traffic flows from an overview visualization. The dynamic behaviors of selected flows may then be shown in annotation windows in which they can be interactively explored and compared. We have demonstrated the effectiveness of our method through a number of use cases in the air traffic domain and the maritime safety and security domain.

Perception: How can we help the operator in *perceiving* a situation? We have shown how to support perception in situational awareness by reducing the cognitive overload and visual overlap caused by dense populations in the maritime picture. In moving object visualization in general, objects and their attributes are commonly represented by glyphs on a geographic map. In areas on the map densely populated by these objects, visual clutter and occlusion of glyphs occur. We have proposed a method to solve this problem and through a user study we found that, for a set of representative tasks, our method does not perform significantly worse than competitive visualizations with respect to correctness. Furthermore, it performs significantly better for density comparison tasks in high density data sets. We also found that the participants of the user study have a preference for our method. Additionally, we have shown how to support percep-

tion in situational awareness using a specialized multi-variate glyph designed in cooperation with domain experts. The glyph is designed to help the operator to perceive quickly which vessels are relevant and require more attention. Starting from maritime domain requirements, a number of design parameters and feasible design choices for them have been determined. We have determined the best choices by showing the glyphs in a sandbox environment and allowing the domain experts to vary the parameters.

Comprehension: How can we help the operator in *comprehending* a situation? We have shown how to support comprehension in situational awareness by visually explaining conclusions of a reasoning engine that raises an alarm if a certain situation is reached. We offer an improvement over current visualization methods, where only a list of evidence is shown. Two groups of domain and operational experts have been used to evaluate our system by testing a number of use cases in the maritime domain based on real data. Experts could easily follow the reasoning structure, and could quickly understand and find complicated patterns in an evidence matrix and relate these to real-world situations.

Projection: How can we help the operator in *projecting* a current situation into the future? We have shown how to support projection into the future of a current situation using a visualization method for the interactive exploration of predicted positions of moving objects, in particular, ocean-faring vessels. Users, investigating and exploring the possible development of a situation, can see where a vessel will be in the near future according to a given prediction model. Through a number of real-world use cases and a discussion with users, we have shown our methods can be used in monitoring traffic for collision avoidance, and detecting illegal activities, like piracy or smuggling. By applying our methods to pedestrian movements, we have shown that our methods can also be applied to a different domain.

Samenvatting

Visualisatie voor maritiem situatiebewustzijn

In het maritiem veiligheidsdomein monitoren operators de maritieme situatie op bedreigingen voor de veiligheid van de maritieme omgeving. Om dit te kunnen doen hebben operators een situatiebewustzijn nodig, wat gedefinieerd is als: “De **perceptie** en het **begrip** van de huidige situatie en de **projectie** van de huidige situatie tot in de toekomst.” In dit proefschrift hebben wij geprobeerd de volgende onderzoeksvraag te beantwoorden:

> *Hoe kan een operator ondersteund worden in het verwerven van situationeel bewustzijn met behulp van interactieve visualisatie?*

Visualisatie speelt een essentiële rol in het ondersteunen van operators in het verwerven en in stand houden van situatiebewustzijn. We hebben nze onderzoeksvraag in de volgende vier delen beantwoord.

Analyse: Hoe kunnen we het analyseren en samenvatten van patronen zodanig instrumenteren dat domeinexperts belangrijke gebieden kunnen vinden en kunnen verifiëren wat normaal of afwijkend gedrag is? We hebben een manier besproken om meer domein kennis te verkrijgen door het analyseren van bewegingspatronen. We doen dit door verkeersstromen te visualiseren en interactief te verkennen. De gebruiker kan op intuïtieve wijze verkeersstromen selecteren en filteren vanuit een overzichtsvisualisatie. Het dynamische gedrag van geselecteerde stromen kan weergegeven worden in annotatievensters waarin deze interactief verkend en vergeleken kunnen worden. We hebben de effectiviteit van onze methode aangetoond middels een aantal gebruiksvoorbeelden uit de luchtvaart en het maritieme veiligheidsdomein.

Perceptie: Hoe kunnen we de operator ondersteunen in het *waarnemen* van een situatie? We hebben laten zien hoe we perceptie kunnen ondersteunen voor situatiebewustzijn door het terugdringen van cognitieve overbelasting en visuele overlap veroorzaakt in een overbevolkte maritieme presentatie. In bewegende object visualisatie in het algemeen worden objecten en hun attributen gewoonlijk gerepresenteerd middels gliefen op een geografische kaart. Dichtbevolkte delen van de kaart zijn niet alleen rommelig, maar lijden ook onder occlusie van gliefen. We hebben een methode gepresenteerd om dit probleem op te lossen en hebben via een gebruikersexperiment bevonden dat, voor een set representatieve taken, onze methode niet significant slechter presteert dan concurrerende visualisaties met betrekking tot correctheid. Bovendien presteert de methode

significant beter in dichtheidsvergelijkingstaken in dichtbevolkte gegevensverzamelingen. We hebben ook bevonden dat de deelnemers van het experiment een voorkeur hebben voor onze methode. Daarnaast hebben we laten zien hoe we perceptie kunnen ondersteunen in situatiebewustzijn door het gebruik van een multivariate glief te gebruiken die in samenwerking met domeinexperts is ontworpen. Deze glief is ontworpen om de operator te ondersteunen in het snel waar te nemen welke schepen relevant zijn en meer aandacht vereisen. Vanuit maritieme domeinvereisten zijn een aantal ontwerpparameters en levensvatbare invullinge hiervan vastgesteld. We hebben de beste ontwerpkeuzes vastgesteld door de gliefs in een zandbakomgeving aan domeinexperts te laten zien en aan de ontwerpparameters te laten sleutelen.

Begrip: Hoe kunnen we de operator ondersteunen in het *begrijpen* van een situatie? We hebben laten zien hoe we het begrip kunnen ondersteunen in situatiebewustzijn door visueel de conclusies uit te leggen van een automatische redeneerder die alarm slaat als aan een bepaalde situatie wordt voldaan. We bieden een verbetering aan boven huidige visualisatiemethoden waar alleen een lijst van bewijzen wordt getoond. Twee groepen met domein en operationele experts zijn ingeschakeld om ons system te evalueren door een aantal gebruiksgevallen in het maritieme domein gebaseerd op echte gegevens te testen. De experts konden de structuur van de redenering eenvoudig volgen en konden snel gecompliceerde patronen begrijpen en vinden in de bewijsmatrix en deze relateren aan echte situaties.

Projectie: Hoe kunnen we de operator ondersteunen in het *projecteren* van een huidige situatie in de toekomst? We hebben laten zien hoe we het projecteren in de toekomst van een huidige situatie kunnen ondersteunen door middel van een visualisatiemethode voor het interactief verkennen van voorspelde posities van bewegende objecten, in het bijzonder van schepen. Gebruikers, die de mogelijke ontwikkeling van een situatie onderzoeken en verkennen, kunnen zien waar een schip kan zijn in de nabije toekomst volgens een gegeven voorspellingsmodel. Door middel van een aantal gebruiksvoorbeelden uit de echte wereld hebben wij laten zien dat onze methode gebruikt kan worden voor het monitoren van verkeer voor het vermijden van aanvaringen en het detecteren van illegale activiteiten zoals piraterij en smokkel. Door onze methode toe te passen op voetgangersbewegingen hebben wij laten zien dat onze methode ook toegepast kan worden op een ander domein.

O Sumário

Visualização para consciência de situação marítima

Na área de segurança marítima, os operadores monitorizam a situação marítima em relação às ameaças ao ambiente marítimo. Para serem capazes de o fazer, os operadores precisam de ter um estado de consciência de situação, o qual é definido como: "A **percepção** e **compreensão** da situação actual e da **projecção** da situação actual no futuro." Nesta tese, tentou-se responder à seguinte questão de pesquisa.

> *Como podem os operadores ser apoiados em alcançar a consciência de situação recorrendo a uma visualização interativa?*

A visualização desempenha um papel importante na ajuda aos operadores para adquirirem e manterem uma consciência de situação. Respondeu-se à questão da pesquisa nas seguintes quatro partes.

Análise: Como é que se pode fornecer ferramentas para *analisar* e resumir padrões, permitindo a especialistas encontrar áreas críticas e verificar o que é um comportamento normal ou anómalo? Discutiu-se uma forma de obter mais conhecimento da área através da análise dos padrões de movimento de objetos em movimento. Fez-se isto por meio da visualização de fluxos de tráfego e fornecendo ferramentas de interação para apoiar a sua exploração. O utilizador pode seleccionar e filtrar de forma intuitiva os fluxos de tráfego a partir de uma visualização geral. Os comportamentos dinâmicos de fluxos seleccionados podem então ser mostrados em janelas de anotação em que os mesmos podem ser exploradas e comparados de forma interactiva. Demonstrou-se a eficácia deste método através de uma série de estudos de caso nas áreas do tráfego aéreo e da área da segurança marítima.

Percepção: Como é que se pode ajudar o operador a *tomar consciência* de uma situação? Demonstrou-se como apoiar a percepção da consciência de situação, reduzindo a sobrecarga cognitiva e a sobreposição visual causadas por populações densas na imagem marítima. Na visualização dos objectos em movimento em geral, os objetos e seus atributos são comummente representados por glifos num mapa geográfico. Em áreas do mapa, densamente povoadas por estes objetos ocorrem a confução visual e oclusão de glifos. Propôs-se um método para resolver este problema e, através de um estudo com utilizadores, descobriu-se que, para algumas tarefas representativas, este método não tem um desempenho significativamente pior do que as visualizações competitivas no que respeita a

exatidão. Além disso, tem um desempenho significativamente melhor nas tarefas de comparação de densidade em conjuntos de dados de alta densidade. Também se descobriu que os participantes neste estudo têm uma preferência por este método. Além disso, demonstrou-se como apoiar a percepção da consciência de situação usando um glifo multivariado especializado, concebido em colaboração com especialistas da área. O glifo é criado para ajudar o operador a tomar rapidamente consciência de quais os navios são relevantes e que precisam de mais atenção. A partir de requisitos da área marítimo, determinou-se um número de parâmetros para a criação e opções de desenho viáveis. determinou-se quais as melhores escolhas, mostrando os glifos num ambiente *sandbox* e permitindo aos especialistas da área variar os parâmetros.

Compreensão: Como é que se pode ajudar o operador a *compreender* a situação? Demonstrou-se como apoiar a compreensão da consciência de situação explicando visualmente as conclusões de um sistema informático de raciocínio que dispara um alarme se uma situação determinada situação é atingido. Oferece-se uma melhoria sobre métodos de visualização actuais, onde apenas uma lista de evidências é mostrada. Foram utilizados dois grupos de especialistas da área e operacionais para avaliar este sistema testando uma série de estudos de caso na área marítimo com base em dados reais. Os especialistas podiam seguir facilmente a estrutura do raciocínio, e podiam compreender e encontrar rapidamente padrões complicados e relacioná-los com situações no mundo real numa matriz de evidência.

Projecção: Como é que se pode ajudar o operador a *projectar* uma situação actual para o futuro? Demonstrou-se como apoiar uma projecção para o futuro de uma situação actual, usando um método de visualização para a exploração interativa de posições previstas de objetos em movimento, em particular, os navios. Os utilizadores, ao investigarem e explorarem a desenvolvimento possível de uma situação, podem ver onde um navio estará num futuro próximo de acordo com um modelo determinado de previsão. Através de uma série de estudos de caso do mundo real e uma discussão com os utilizadores, demonstrou-se que este método podem ser utilizados no monitorização do tráfego para evitar colisões, e detectar atividades ilegais, como a pirataria ou contrabando. Através da aplicação deste método a movimentos de pedestres, demonstrou-se que o mesmo pode também ser aplicado a uma área diferente.

Curriculum Vitæ

Roeland Scheepens was born on the 22nd of March 1983 in Cuckfield, the United Kingdom. He completed his secondary education at the Christiaan Huygens College in Eindhoven. Following this, he studied Computer Science & Engineering at the Eindhoven University of technology. In 2010 he received his master's degree with his thesis titled "GPU-based track visualization of multivariate moving object data". For his master's thesis he implemented a previously developed visualization technique for generating density maps on the GPU and extended this with interactive visualization and interaction techniques to reveal spatial and temporal distributions as well as relations between subsets of the data that are otherwise not visible in the traditional density map.

From October 2010 until until March 2011, he worked as a researcher in the visualization group at the Eindhoven University of technology to further expand on the above work.

In April 2011, he started on a COMMIT supported PhD project at the same university under the supervision of prof.dr.ir. Jarke J. van Wijk and dr.ir. Huub van de Wetering. This research was carried out as part of the Metis project under the responsibility of Embedded Systems Innovation by TNO with Thales Nederland B.V. as the carrying industrial partner. The results of this work are presented in this thesis.