Year: 2007

Assessing landmark salience for human navigation

Caduff, David

Abstract: Prominent spatial features play an important role for a plethora of spatially related tasks, including spatial learning, wayfinding and navigation, and the communication of route directions. Human judgment of the prominence or importance of these spatial features, for which the term landmark became popular, is typically based on subjective impressions and experience. The computational assessment of the prominence of these spatial objects is of interest to various scientific disciplines and applications, including spatially related information and navigation systems. Computational salience assessment, however, is highly challenging, as information systems need objective criteria and formalized techniques to reproduce human judgment of landmark salience. We propose a conceptual framework for assessing the salience of landmarks for navigation. Landmark salience is derived as a result of the observer's point of view, both physical and cognitive, the surrounding environment, and the objects contained therein. This is in contrast to the currently held view that salience is an inherent property of some spatial feature. Salience, in our approach, is expressed as a three-valued Saliency Vector. The components that determine this vector are Perceptual Salience, which defines the exogenous (or passive) potential of an object or region for acquisition of visual attention, Cognitive Salience, which is an endogenous (or active) mode of orienting attention, triggered by informative cues providing advance information about the target location, and Contextual Salience, which is tightly coupled to modality and task to be performed. This separation between voluntary and involuntary direction of visual attention in dependence of the context allows defining a framework that accounts for the interaction between observer, environment, and landmark. We identify the low-level factors that contribute to each type of salience and suggest a probabilistic approach for their integration. The framework serves as a bridge between findings from spatial cognition research and practical applications, and forms the basis for a computational model, which is used as test-bed for the evaluation of the concepts and methods developed within the scope of this work. The evaluation includes a comparison with human assessment of salience and provides the evidence for assessing the quality of the model. The results of this comparison suggest that the conceptual framework provides reasonably accurate assessments of saliency for perceptually distinct objects, but also identifies two major issues. The first relates to a systematic weighting issue of low-level components due to the proposed technique for the integrated saliency assessment, and the second aspect is the indication that the model lacks explanatory power due to the limited number of low- level components, in particular for cognitive components.Prominente räumliche Objekte spielen eine wichtig Rolle bei einer Vielzahl von raumbezogenen Aufgaben, wie zum Beispiel beim Erlernen der räumlichen Umgebung, bei der Wegfindung und Navigation, oder auch bei der Kommunikation von Routenbeschreibungen. Menschen beurteilen die Prominenz solcher Objekte, welche oft auch als Landmarken bezeichnet werden, aufgrund subjektiver Eindrücke und Erfahrungen. Die automatische Abschätzung dieser Prominenz mithilfe von Berechnungsmodelle und Algorithmen ist ausschlaggebend für die Entwicklung und Implementierung von Informationssystemen der nächsten Generation. Allerdings ist diese automatische Abschätzung sehr komplex und anspruchsvoll, da Informationssysteme weder subjektive Eindrücke verarbeiten noch über Erfahrungen verfügen, sondern auf formalisierte Methoden und Techniken angewiesen sind. Diese Dissertation befasst sich mit den konzeptuellen Rahmenbedingungen die zu einer akkuraten automatischen Abschätzung der Prominenz von Landmarken notwendig sind, wobei Prominenz als Salienz verstanden wird, also das Hervorspringen oder Hervorstehen eines Objekts aus einer Gruppe von Objekten. Die Salienz von räumlichen Objekten ist abgeleitet von drei zentralen Komponenten, nämlich 1) vom physischen und kognitivem Standpunkt des Beobachters, 2) von den Gegebenheiten der räumlichen Umgebung, und 3) von den einzelnen Objekten die sich im Wahrnehmungsbereich des Beobachters befinden. Die Salienz ist dementsprechend als drei-dimensionaler Vektor definiert, bestehend aus einer Wahrnehmungskomponente, einer Kognitionskomponente, und einer Kontextkomponente. Der konzeptuelle Rahmen diente dazu, Forschungsresultate aus verschiedenen wissenschaftlichen Disziplinen zu integrieren und ein Berechnungsmodel und Prototyp zu erstellen, welches als Testumgebung für die Evaluierung der angewandten Konzepte und Methoden, sowie für weitere Forschungsprojekte benutzt werden kann. Die Evaluierung besteht aus einem Vergleich der Resultate mit den Resultaten einer entsprechenden Umfrage und dient dazu, die Qualität des Berechnungsmodels abzuschätzen. Die Ergebnisse der Evaluierung zeigen dass der konzeptuelle Rahmen und das Berechnungsmodell tendenziell korrekte Abschätzungen der Salienz von Landmarken produzieren. Die Ergebnisse zeigen aber auch auf dass das Model Schwächen und Lücken hat, vor allem in Bezug auf die einzelnen Komponenten die zur Salienz beitragen.

Assessing Landmark Salience for Human Navigation

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

David Caduff

von

Degen GR

Promotionskomitee

Prof. Dr. Sara I. Fabrikant (Vorsitz)
Prof. Dr. Daniel R. Montello
Prof. Dr. Stephan Nebiker
Dr. Sabine Timpf (Leitung der Dissertation)

November, 2007

Abstract

Prominent spatial features play an important role for a plethora of spatially related tasks, including spatial learning, wayfinding and navigation, and the communication of route directions. Human judgment of the prominence or importance of these spatial features, for which the term landmark became popular, is typically based on subjective impressions and experience. The computational assessment of the prominence of these spatial objects is of interest to various scientific disciplines and applications, including spatially related information and navigation systems. Computational salience assessment, however, is highly challenging, as information systems need objective criteria and formalized techniques to reproduce human judgment of landmark salience.

We propose a conceptual framework for assessing the salience of landmarks for navigation. Landmark salience is derived as a result of the observer's point of view, both physical and cognitive, the surrounding environment, and the objects contained therein. This is in contrast to the currently held view that salience is an inherent property of some spatial feature. Salience, in our approach, is expressed as a three-valued Saliency Vector. The components that determine this vector are Perceptual Salience, which defines the exogenous (or passive) potential of an object or region for acquisition of visual attention, Cognitive Salience, which is an endogenous (or active) mode of orienting attention, triggered by informative cues providing advance information about the target location, and Contextual Salience, which is tightly coupled to modality and task to be performed. This separation between voluntary and involuntary direction of visual attention in dependence of the context allows defining a framework that accounts for the interaction between observer, environment, and landmark. We identify the low-level factors that contribute to each type of salience and suggest a probabilistic approach for their integration.

The framework serves as a bridge between findings from spatial cognition research and practical applications, and forms the basis for a computational model, which is used as test-bed for the evaluation of the concepts and methods developed within the scope of this work. The evaluation includes a comparison with human assessment of salience and provides the evidence for assessing the quality of the model. The results of this comparison suggest that the conceptual framework provides reasonably accurate assessments of saliency for perceptually distinct objects, but also identifies two major issues. The first relates to a systematic weighting issue of low-level components due to the proposed technique for the integrated saliency assessment, and the second aspect is the indication that the model lacks explanatory power due to the limited number of low-level components, in particular for cognitive components.

Zusammenfassung

Prominente räumliche Objekte spielen eine wichtig Rolle bei einer Vielzahl von raumbezogenen Aufgaben, wie zum Beispiel beim Erlernen der räumlichen Umgebung, bei der Wegfindung und Navigation, oder auch bei der Kommunikation von Routenbeschreibungen. Menschen beurteilen die Prominenz solcher Objekte, welche oft auch als Landmarken bezeichnet werden, aufgrund subjektiver Eindrücke und Erfahrungen. Die automatische Abschätzung dieser Prominenz mithilfe von Berechnungsmodelle und Algorithmen ist ausschlaggebend für die Entwicklung und Implementierung von Informationssystemen der nächsten Generation. Allerdings ist diese automatische Abschätzung sehr komplex und anspruchsvoll, da Informationssysteme weder subjektive Eindrücke verarbeiten noch über Erfahrungen verfügen, sondern auf formalisierte Methoden und Techniken angewiesen sind.

Diese Dissertation befasst sich mit den konzeptuellen Rahmenbedingungen die zu einer akkuraten automatischen Abschätzung der Prominenz von Landmarken notwendig sind, wobei Prominenz als Salienz verstanden wird, also das Hervorspringen oder Hervorstehen eines Objekts aus einer Gruppe von Objekten. Die Salienz von räumlichen Objekten ist abgeleitet von drei zentralen Komponenten, nämlich 1) vom physischen und kognitivem Standpunkt des Beobachters, 2) von den Gegebenheiten der räumlichen Umgebung, und 3) von den einzelnen Objekten die sich im Wahrnehmungsbereich des Beobachters befinden. Die Salienz ist dementsprechend als drei-dimensionaler Vektor definiert, bestehend aus einer Wahrnehmungskomponente, einer Kognitionskomponente, und einer Kontextkomponente.

Der konzeptuelle Rahmen diente dazu, Forschungsresultate aus verschiedenen wissenschaftlichen Disziplinen zu integrieren und ein Berechnungsmodel und Prototyp zu erstellen, welches als Testumgebung für die Evaluierung der angewandten Konzepte und Methoden, sowie für weitere Forschungsprojekte benutzt werden kann. Die Evaluierung besteht aus einem Vergleich der Resultate mit den Resultaten einer entsprechenden Umfrage und dient dazu, die Qualität des Berechnungsmodels abzuschätzen. Die Ergebnisse der Evaluierung zeigen dass der konzeptuelle Rahmen und das Berechnungsmodell tendenziell korrekte Abschätzungen der Salienz von Landmarken produzieren. Die Ergebnisse zeigen aber auch auf dass das Model Schwächen und Lücken hat, vor allem in Bezug auf die einzelnen Komponenten die zur Salienz beitragen.

Acknowledgments

Although it might seem that the following dissertation is an individual work, it would never have been possible without the guidance, support, help, understanding, effort, and goodwill of a lot of people. It is to these people, whom I owe highest gratitude, that this section is dedicated.

First of all, I would like to thank Dr. Sabine Timpf for offering me the opportunity to conduct this research within the frame of one of her research project. Despite her many responsibilities with research, teaching, and family, she found the time to guide and support this work in order for me to bring it to a successful end. The Swiss National Science Foundation funded most of this project.

I would also like to thank Prof. Dr. Sara Fabrikant for taking over the job of the responsible faculty member after prof. Kurt Brassel's retirement, and to provide support during the remainder of the dissertation. At the same time, I would also like to thank Prof. Dr. Stephan Nebiker and Prof Dr. Daniel R. Montello, the other members of the advisory committee, for their valuable input and comments.

Many people on the staff of the Department of Geography at the University of Zurich assisted and inspired me in various ways during my course of studies. I am especially grateful to Urs-Jakob Rüetschi, my office mate and reliable source of advice and many good ideas, as well as Jeannette Nöetlzli for her input and careful proofreading. In addition, I would also like to thank Arzu Çöltekin, Tumasch Reichenbacher, Frank Obermann, Stephan Steiniger, and Anna-Katharina Lautenschütz for their valuable contributions and discussions.

Finally, my sincere gratitude goes to my parents, my brother and my sisters for their support, and to all my friends, who helped me keep the balance between academic life and living a life.

Table of Contents

Table of Figures

Chapter 1
Introduction

Landmarks are present throughout history as reference points for navigation and also play an important role in the development of spatial knowledge and for solving spatial reasoning problems. There is a vast body of literature that supports the importance and role of landmarks, among them Siegel and White's (1975) seminal work that introduced a three-phased theory of acquisition of spatial knowledge, which assumes that landmarks are the linking points between Route and Survey Knowledge, and thus, form the base of cognitive maps (Tolman 1948; Downs and Stea 1977). Lynch (1960) investigated human descriptions of urban environments and identified landmarks, along with districts, edges, nodes, and paths as one of the main elements that enhance imageability of city space.

Despite the proven significance of landmarks and the vast literature on its nature, there are only a few attempts to quantify the *quality* of landmarks (Raubal and Winter 2002; Elias 2003; Nothegger, Winter et al. 2004). The objective of this thesis is to complement landmark research with the definition of a conceptual framework that draws from previous research, and to develop a computational model for the assessment of the quality of landmarks. In this scope, we investigate the nature of landmarks and identify the components that define their quality. A software prototype of the computational model is implemented to verify the developed concepts and theories.

1.1 Problem Statement and Motivation

Navigation is defined as coordinated and goal-directed movement through the environment and requires both *planning* of a route and *execution* of movements (Montello 2003) along this route. Planning a route involves reasoning about the immediate and distant environment, as well as active decision-making about possible routes through this environment from a starting location to a destination. Execution of movements, in contrast, is understood as locomotion adapted to the local surrounds. The planning process is also known as *wayfinding* and typically manifests itself in route instructions. The task of emulating this process and producing cognitively adequate route instructions is of great significance for many practical applications, such as navigational aids for various modes of transportation (navigation system, traffic information systems, etc.) or spatially related information systems (route planners, tourist information systems, location based services, etc.).

The automated generation of cognitively adequate route instructions is a highly complex task, as it involves not only metric information about routes, segments, and turns, but also references to prominent spatial features. From the beginning of human history, such prominent spatial features, for which the collective term landmarks became popular, played an important role. They are conceivably the most fundamental pieces of spatial information as they are used for a wide collection of tasks related to the description, understanding of and reasoning about our physical environment (Lynch 1960; Siegel and White 1975; Golledge 1991; Montello 1997; Montello and Freundschuh 2005). Several studies investigated the role of landmarks (Allen 1997; Werner, Krieg-Brückner et al. 1997; Fontaine and Denis 1999; Lovelace, Hegarty et al. 1999; Lee, Tappe et al. 2002; Steck, Mochnatzki et al. 2003) and affirmed their importance (Denis, Pazzaglia et al. 1999; Daniel and Denis 2004; Tom and Denis 2004; Weissensteiner and Winter 2004; Newman, Caplan et al. in press) as essential part of the production and communication of route instruction. Despite this evidence, only few attempts exist to enhance route instructions with landmark knowledge (Raubal and Winter 2002; Nothegger 2003; Winter 2003; Nothegger, Winter et al. 2004; Winter, Raubal et al. 2004) or to incorporate landmarks in route generation algorithms (Caduff and Timpf 2005a; Caduff and Timpf 2005b; Rüetschi, Caduff et al. 2006).

The reason for the lack of such solutions lies in the intricacy of determining what spatial features arise as 'good' landmarks in what context. This complexity is tightly linked to the semantics of the term landmark. The original meaning of the term in a navigational context was that of a distinct geographic feature used by hunters, explorers and others to find their way back through an area on a return trip. The semantics of the term in modern usage differs merely in the type of the objects that are referenced. Hence, a landmark may be any object in the environment that is easily recognizable (e.g., buildings, rivers, specific districts) or even idiosyncratic objects (e.g., a celebrities mansion, my workplace), as long as its primary property is that of a point of reference (Presson and Montello 1988; Couclelis, Golledge et al. 1995).

One of the most important concepts in this context is the notion of *salience* or *saliency*. This term denotes relatively distinct, prominent or obvious features compared to other features. The above definition of a landmark, however, suggests that the assessment of the salience of landmarks is a challenging task. In this thesis, we review literature on the assessment of landmark salience, whereby we focus on the use of landmarks for human navigation, and propose a conceptual framework for the assessment of the importance of potential landmarks. The conceptual framework provides the foundation for a computational model of integrated saliency assessment, which is used as test bed for evaluation.

1.2 Approach

Gaerling (1986) found that three facets of the physical environment are important for successful wayfinding. These facets are 1) degree of (architectural) differentiation, 2) degree of visual access, and 3) complexity of spatial layout, and are essentially the result of the tri-lateral relationship between observer, observed feature, and physical environment. Accordingly, the central assumption of our approach is that this trilateral relationship defines the salience of the observed spatial feature. This approach allows incorporating perceptual, cognitive, and contextual aspects into the assessment of salience, and hence, accounts for all three facets identified by Gaerling.

This definition of salience differs drastically with the traditional definition. The property of being a landmark has so far been attributed to distinct objects, such as facades, churches, or other outstanding buildings (Sorrows and Hirtle 1999; Raubal and Winter 2002; Winter 2003). We argue that salience is not an inherent property of some specific spatial features, but rather is a unique property of the trilateral relation between the feature itself, the surrounding environment, and the observer's point of view, both, cognitively and physically. This view is in accordance with studies of human behavior in urban environments that investigate why environmental features are known or referenced (Lynch 1960; Appleyard 1969). In the following paragraphs we will elaborate this claim and lay out the theoretical framework of our approach.

The most general requirement of a landmark is that it must be perceptually salient in some sense (i.e., visually, auditory, olfactory, or semantically). Specifically for vision, this requires, first of all, a contrast with the environment (e.g., architectural differentiation), either in terms of its attributes (color, texture, size, shape, etc.) or due to its spatial location with respect to the other objects in the scene. Contrast and perceptual distinction of sensory input are key to learning landmarks from spatial environments (Montello and Freundschuh 2005), and hence, are important aspects of salience. Perceptual distinction is also imperative when formulating route instructions that are addressed to navigators unfamiliar with the environment. In contrast, it is of lesser importance if the inquiring navigator is familiar with the environment and relies not only on perceptual input, but also on former experience and knowledge. Hence, the degree of importance of the perceptual input varies as a function of the experience of the navigator.

This subjective selection of spatial references implies that the cognitive abilities of the observer play an important role in selecting appropriate features for reference (Presson and Montello 1988; Stevens 2006), that is, our knowledge, thoughts and preconceptions shape what we perceive and finally select as reference for making decisions. The cognitive processes involved in understanding and reasoning about a spatial scene include knowing, thinking, learning, judging, and problem solving (Montello and Freundschuh 2005). Cognitive abilities vary strongly among observers and directly influence the assessment of the relative importance or salience of potential landmarks. The salience assessment, hence, needs to consider cognitive aspects, along with the perceptual stimuli.

Human perception is always limited to our view of the world and the properties of our sensory system as it is intrinsically tied to our egocentric frame of reference (Parkhurst and Niebur 2003; Marcel and Dobel 2005). The origin of this frame of reference is defined by the current position of the navigator and its orientation exhibits a directional fixation of varying strength. The orientation of our visual frame of reference, for instance, is firmly tied to the plane of progression (Hollands, Patla et al. 2002), while the orientation of the auditory frame of reference is only loosely coupled with the orientation of the body.

Another aspect we consider is that navigation may be performed by different means of transportation (walking, riding, driving, etc.). Each of these modes imposes a different cognitive load on the navigator, which in turn affects the range of perception and amount of visual attention available for wayfinding. Walking, for instance, allows for a greater degree of physical freedom and requires fewer cognitive resources than driving, which in turn affects the range of perception and hence, modulates the salience of features in the environment. The directed goal-oriented nature of navigation together with the means of transportation dictates the perceptual range, which implies that only features that are within this range contribute to the salience.

Landmarks are prominent spatial features, which are often used as points of reference to identify targets or reassure navigators that they are still on track (Denis, Pazzaglia et al. 1999; Montello 2003), whereby emphasis is put on the notion of ‘point of reference’. The statement ”Follow the river,” for instance, is basically an abbreviation of ”Take the path that will lead you along the river.” Such a statement differs considerably from just mentioning that a landmark can be seen from some point of view, as it not only refers to the landmark as a main attraction, but in that it uses the spatial relation between landmark and path in order to identify what path to take next. As a result, the spatial relation between path and spatial feature dictates the degree of salience of a potential landmark. These considerations point out that the circumstances and the purpose of a journey, which we will refer to as Navigation Context, influence the salience of features and need to be considered accordingly.

1.3 Integrated Assessment of Landmark Salience

Considering perceptual characteristics, cognitive aspects, and contextual influence in the assessment of landmark salience promises to produce accurate approximations of the salience of urban objects. This thesis attempts to provide evidence for this assertion. The following sections describe goal and concept as well as the hypothesis and the four central research questions of this research.

1.3.1 Goal

The goal of this thesis is to draw from research results of various scientific disciplines (psychology, spatial cognition, geographic information science, etc.) and to create a framework for the computational assessment of landmark salience based on this evidence. The focus of the work, however, is on the integration of the components of salience, rather than on their individual peculiarities. Such a framework, along with the computational model will bridge the gap between theoretical findings and practical applications. In addition, it can be used as a hypothesis-testing engine for further research related to the assessment of landmark salience.

1.3.2 Hypothesis

This research focuses on the assessment of landmark salience for human navigation in urban environments. How salient urban objects are depends on a plethora of factors, including perceptual, cognitive, and contextual factors. The quality of a model for the quantification of salience depends on the integration of these factors. This integration, however, is intrinsically complex, as it involves a reduction of the factors, as well as formalisms that quantify the mutual influence among the components. Hence, finding a set of factors and a method for their integration without decreasing the model's expressiveness is a fundamental premise for a computational model of salience assessment.

The hypothesis of this thesis, therefore, is concerned with finding a specific subset of contributing factors that, when used for assessing the object's salience, lead to a ranking that is comparable to that obtained by human judgment. The focus is on the first few ranks of the assessed objects, because they are most likely to be used as references for human navigation. This leads to the formulation of the following hypothesis:

> *"If salience of urban objects is a result of the trilateral relationship between observer, environment, and observed object, then a computational model based on this relationship approximates saliency judgments by humans."*

The trilateral relationship between observer, environment, and observed object can be redefined from the perspective of human information processing. The basic assumption is that perception, cognition, and context are the fundamental components of human information processing, whereby the interaction between the components is a crucial aspect of the assessment process. Therefore, we reformulate the general hypothesis into the following two testable hypothesis statements (*HS_1* and *HS_2*):

> ***HS_1:*** *If perceptual, cognitive, and contextual aspects fully explain the trilateral relationship between observer, observed object, and environment then a computational model that integrates these components produces saliency values that approximate saliency judgments by humans.*

> ***HS_2:*** *Perceptual, cognitive, and contextual components contribute equally to landmark salience.*

Proving these hypothesis statements requires comparing results generated by a computational model with results based on human judgment, which, in turn, requires the compilation of appropriate data sets. The focus of this comparison is on the relevant aspects of the integrated saliency assessment, that is, on the contributing components and the interaction between these components. Contributing components and interaction between them, however, are two aspects that are tightly intertwined, rather than independent. Therefore, we will have to design elaborate scenarios that allow assessing the hypotheses and draw the correct conclusion. Should the first testable statement (i.e., *HS_1*) proof true, however, then the focus of research on a computational level will shift from identifying the components to thoroughly analyzing their interaction. If not, additional effort into the set of contributing components will be required.

1.3.3 Research Questions

The hypotheses are embedded in a set of research questions, which will provide the evidence for the evaluation. Specifically, the research questions that we will investigate are the following:

Question 1: *What are the fundamental components of salience?*

We are interested in analyzing the influence of perception, cognition, and context on salience of urban objects. What are the specific factors of perception, cognition, and context that contribute to salience? These questions are relevant, because their answers set the frame for this work.

Question 2: *How do the individual components of salience influence each other?*

In a navigation context, perception is key to the assessment of landmark salience. Perception, however, is influenced by prior knowledge and experience, as well as context. An important issue that we want to examine is the role of each component and their mutual influence. Understanding the role of each component and how they are related is crucial for computationally assessing landmark salience.

Question 3: *Is a computational model for integrated saliency assessment feasible?*

This question is concerned with the practicality of the framework and concepts that are developed within the scope of this thesis. The implementation of a computational model becomes the test bed for integrated salience assessment, empirical tests of the proposed formalisms, and the overall framework.

Question 4: *How well does the computational model replicate saliency rankings by human subjects?*

This question is concerned with the quality of the computational model with respect to real-world scenarios. In particular, we are interested in evaluating if the proposed framework suffices for the assessment of landmark salience. What are the benefits of the framework, and what are its limitations and restrictions? Where are refinements necessary? These questions are important for further development and research.

1.3.4 Concept

The assessment of landmark salience for navigation can be described at four conceptual stages. The first stage is concerned with interpreting the sensory input of the current environment so that perceptually distinct components can be extracted. Once these components have been extracted and assigned to a specific object in the scene, the objects are compared sequentially for cognitive differences. In the third stage, the objects are rated according to their contextual importance. Finally, in the fourth stage, the single components that define salience are integrated and a ranking of the objects is produced, which orders the objects according to their salience. This ranking mirrors the importance of objects with respect to the trilateral relation between observer, environment, and observed object.

The investigation starts with a systematic examination of the most important psychological insights about the nature of salience, its properties in terms of perceptual and cognitive resources, and the implications of these properties for the direction of attention. We follow a bottom-up approach, starting with the identification of the components that contribute to salience and progressing to the structure and dynamics of their interaction. Combining the results of this inquiry produces a solid theoretical foundation for the conceptual framework.

In a second step, the conceptual framework will be used for the definition of a computational model. The implementation of the computational model serves both, as proof for the feasibility of an integrated salience assessment, as well as for evaluating the correctness and performance of framework and model with respect to human judgment of salience. An online survey will provide the benchmark data set against which the results of the computational model will be tested. The final step consists of the discussion and interpretation of the results, along with the conclusion of the research questions and the evaluation of the hypotheses.

1.4 Major Results

The major findings of this thesis suggest that approaching the assessment of salience based on the trilateral relationship between observer, environment, and observed object produces rankings of salience that approximate the rankings produced by humans. The findings also show, however, that the set of components proposed in the framework is not sufficient for accurate predictions of salience for complex scenes. Furthermore, the results show that the interaction between the components of salience in the integrated saliency assessment varies with the content of the scene. That is, cognitive aspects contribute stronger to salience if the objects are perceptually similar.

The computational model that was used for the empirical evaluation of the assessment is based on the conceptual framework for the integrated assessment of landmark salience. The prototype implementation of the concepts proposed in the framework shows that (1) a computational model of integrated saliency assessment for human navigation in urban environments is feasible, (2) the approach based on the trilateral relationship produces reasonably good approximations of salience values for simple scenes, and (3) that the framework needs further refinement, especially in terms of cognitive capabilities and integration of components, in order to produce better results for complex scene configurations.

The online survey about real-world judgment of salience provided the benchmark data set for the evaluation of the computational model. It also provided evidence for the complexity inherent in the assessment of salience. Specifically, judgments of objects' salience showed a high variation for scenes with semantically similar objects, which suggests that there is no consensus among participants on a single rating strategy. Rather, it confirms the results from previous research that assigns a prominent role to cognitive aspects. These observations and findings from the survey are relevant, because they provide the basis for the revision and refinement of the proposed framework and computational model.

1.5 Intended Audience

This work is intended for researchers and developers interested in the computational assessment of salience and the use of the results in geographic information systems, particularly in a navigation context. The proposed framework in combination with the computational model provides a test environment that may be of interest to scientists who want to perform experimental human subject tests or experiments based on integrated assessment of salience.

1.6 Organization of Thesis

This thesis is divided into sections according to the four research questions postulated in Section 1.3.2. One chapter is devoted to the first two questions while questions 2 and 3 are discussed in separate chapters, whereby each chapter builds on observations and findings of previous chapters. The remainder of this thesis is organized as follows:

The second chapter creates the link between previous research and our work by reviewing and summarizing conceptual and computational approaches to the assessment of landmark saliency. For this purpose, related research in various fields, such as psychology and spatial cognition, geographic information science, and artificial intelligence is analyzed and compared. Specifically, this chapter reviews previous approaches and theories of landmarks, investigates the underlying assumptions and concepts, and describes proposed computational frameworks in detail. On the basis of this literature review, we identify the components that contribute to salience and introduce a framework for the integrated assessment of landmark salience that describes the dynamics of the trilateral relationship between observer, observed object, and environment. The framework is based on theories of attention and human information processing, which are reviewed and organized such as to provide a solid foundation for this work.

Chapter three is concerned with the feasibility of the proposed framework, wherefore proof is provided in terms of a computational model. The chapter is composed of two parts. The first part gives an overview of the computational strategy, including the specification of the data model for the representation of urban scenes, the model of human information processing, and the infrastructure for the combination of the two modules. The second part formally defines the quantification of the scene content and the integrated salience assessment process.

The fourth chapter evaluates the framework and the computational model proposed in the previous chapters. The chapter is divided into a part concerned with the verification of the computational model, and a second part, which validates the conceptual framework. Verification is based on test cases using artificial and real-world data, and ensures that the computational model conforms to the specifications. Validation, in contrast, investigates the degree of correlation of rankings generated by the computational model and ranking received by means of an online survey. The chapter presents the setup, methods, and results used in the evaluation process.

Chapter five critically discusses the presented work and draws the conclusions. The discussion includes the scope and limitations of the framework, experiences from the implementation of the computational model, the validity of the results, and an outlook on the expected effects of our work on future research. Finally, the chapter concludes with the assessment of the research questions, the evaluation of the hypothesis, and the scientific and industrial contributions.

Chapter six concludes the thesis with a summary and the presentation of the major results. The chapter also provides an outlook on possible enhancements to the model and future research activities enabled through this research. The outlook focuses on conceptual extensions of the framework, potential refinements of the computational model, and possible applications in different fields. The thesis closes with a portrayal of an integrated route generation system that includes the automatic assessment of landmark saliency, which provided the main motivation for this work.

Chapter 2
Background and Framework

In this chapter, we will set our work in relation to previous landmark-based research by reviewing relevant literature and by defining a conceptual framework for the assessment of landmark salience. The main purpose of the conceptual framework is to set the base for a computational model for an integrated assessment of salience. The framework for the assessment of landmark salience is based on the assumption that salience of landmarks can only be determined when taking into consideration situatedness along with perceptual and cognitive abilities of the traveler. In a navigation context, hence, salience of geographic objects is a property of the trilateral relationship between observer, environment and geographic object. The overall salience of geographic features is defined as a three-valued vector, whereby the components capture perceptual, cognitive, and contextual aspects of geographic objects.

The chapter is organized as follows: In the first section, we will review the literature and background of landmarks research in terms of theoretical and computational approaches. In the second section, we conceptualize our understanding of landmark salience for human navigation and introduce a strategy for quantifying the components that contribute to salience. In the third section we investigate the integration of the components in a single assessment process, which is based on a probabilistic approach. Finally, we conclude with a summary of the main points of this chapter.

2.1 Related Work

The nature of landmarks has been investigated from various points of view (Presson and Montello 1988; Golledge 1991; Couclelis, Golledge et al. 1995; Denis, Pazzaglia et al. 1999), but despite the vast amount of evidence for the prominent role landmarks play in spatial behavior and navigation, few attempts have been made to formally characterize the qualities of landmarks and to computationally assess their salience. In the following sections we review landmark-related work in terms of formal descriptions and computational frameworks.

2.1.1 Landmark Theory

Sorrows and Hirtle (1999) proposed one of the most influential descriptions of the characteristics of landmarks in the domain of Geographic Information Science (GIScience). The authors compare commonalities between real and electronic space and propose three different characteristics of a 'good' landmark. These aspects are: 1) Visual Prominence, which describes the visual importance of a spatial feature, 2) Semantic Salience, which describes the cultural or historical importance of the feature, and 3) Structural Significance, which explains the role that the feature plays in the configuration of the environment. The approach is an attempt to generically describe the nature of landmarks for real and electronic space in a comprehensive way, but no formalization is proposed.

An alternative characterization of landmarks and their properties was proposed by Burnett et. al. (2000), who suggest permanence, visibility, location in relation to a decision point, uniqueness, and brevity as the main aspects of 'good' landmarks. The main objective of the study was to investigate the properties of landmarks in terms of usability for car navigation. The study revealed that the significance of landmarks for car navigation (e.g., traffic lights, pedestrian crossings, and petrol stations) was dependent on the mentioned aspects, whereby two of these aspects correlate with the aspects proposed by Sorrows and Hirtle (i.e., visual salience as equivalent to visibility and structural salience as equivalent to location in relation to a decision point). Both approaches are restricted to a qualitative characterization of landmarks and lack an answer on how to assess landmark salience for navigation.

2.1.2 Proposed Computational Frameworks

The enumeration of the quantitative and qualitative parameters that define a landmark is the first step in the assessment of its salience. The second step is the computational evaluation of these parameters. The computational assessment of landmark salience is of interest to many scientific fields (GIScience, Robotics and Artificial Vision, Remote Sensing, etc.). For the purpose of this review, we focus on reviewing approaches in the fields of GIScience and Artificial Vision.

2.1.2.1 Geographic Information Science

Sorrows and Hirtle's (1999) characterization of landmarks provides the foundation for various computational approaches for the determination of the salience of landmarks in the

GIScience domain. Raubal and Winter (2002) propose a model of landmark salience that addresses the question of enriching route instructions with local landmarks. The authors suggest a set of measures for each aspect (i.e., visual, semantic, and structural) to formally specify the landmark salience of a feature. The model was developed with a specific set of urban features in mind, namely facades, and was further refined and tested by Nothegger (2003; 2004). The results suggest that the model is a viable assessment of the salience of landmarks. However, as the approach focuses on facades and landmarks are treated as point-like structures, prominent spatial features, such as rivers or districts, which are essential for wayfinding tasks, are not considered.

Elias (2003) proposes an approach for the extraction of landmarks from large datasets that is based on Sorrows and Hirtle's (1999) definition of a landmark and on Raubal and Winter's salience model (2002). From a computational point of view, the main objective of Elias' approach is to automatically extract landmarks from existing data using a data mining approach (Elias 2003). Although the approach considers a variable point of view of the wayfinder and different modes of transportation, it lacks a detailed investigation of the cognitive peculiarities involved with navigation, such as cultural differences, experience of navigators, and relative importance of certain features to observers. Yet the investigation provides useful insights about the collection and processing of suitable data, particularly when data collection involves large sets of data.

A similar approach was taken by Galler (2002) in her attempt to identify landmarks in urban environments. The goal of this work was to use the existing theoretical framework (Sorrows and Hirtle 1999; Raubal and Winter 2002; Elias 2003) for the characterization of landmark attributes and to propose an automated solution for the assessment of landmark salience in 3D city models. An interesting aspect of this work is that a reference set of visible urban features (i.e., facades) is evaluated using descriptive statistics and Shannon's information theory (Shannon 1948), with the evident goal of singling out those features that contrast most within the set. The results show that this approach for the characterization of urban space is promising, despite the fact that the type of features is constrained to facades and the number of attributes for which measures are derived is restricted to a set of eight attributes (i.e., accessibility, height, width, curvature, color, signs and marks, and relief).

2.1.2.2 Synergetics and Self-organizing Systems

Similarly, Haken and Portugali (2003) propose a synergetic approach for the assessment of landmark salience that uses information theory to define the amount of information externally represented in urban environments. Based on Lynch's elements of the city (i.e., nodes, paths, edges, landmarks, and districts), the authors introduce a process of grouping and categorization, which gives meaning to the urban environment and thus forms its semantic information. This approach, however, takes a global view at the urban environment as it is based on Shannon entropy (Shannon 1948), which is a measure of the average information content of a system. Analogous to Galler's (2002) approach and as a result of the holistic nature of information theory, this approach does not allow deducing values of single features in relation to observer and navigation task, and hence, is inadequate for our purpose.

2.1.2.3 World Wide Web

Tezuka and Tanaka (2005) investigated the World Wide Web as a source for landmarks and suggest web mining as a new, vision-independent way of acquiring knowledge about landmarks. The central focus of this work is on the way humans express knowledge of geographic objects, rather than how objects are perceived. The expression of spatial knowledge is assessed by means of statistical and linguistic measures, which also take spatial context into account, and result in the generation of new geographic knowledge not present in conventional Geographic Information Systems. First results suggest that this approach matches with human judgment of landmarks. Nevertheless, the relevance of this approach for the evaluation of landmark saliency for navigation is marginal, as the approach does not account for the goal-oriented nature of navigation.

2.1.2.4 Landmarks and the Generation of Route Instructions

Klippel et al. (2005) introduce a model of structural salience that complements landmark research with an approach to formalize the structural salience of objects along routes. The structural salience of point-like objects is approached with taxonomic considerations and with respect to their positions along a route. The results are used to extend the wayfinding choreme theory, which is a formal language of route knowledge (Klippel 2004; Klippel, Richter et al. 2005). Analogous to Raubal and Winter's approach (2002), this approach treats landmarks as point-like features and does not consider spatially extended objects as potential landmarks. However, it provides a solid foundation for the incorporation of locomotion into the assessment process.

Moulin and Kettani (1999) developed a system that uses the influence area of spatial objects to generate route descriptions. The system uses a spatial model to represent neighborhood, orientation, and distance between wayfinder and spatial objects, based on which prominent spatial entities, i.e., landmarks, are deduced and integrated in route directions. The system produces route directions that correspond to descriptions given by humans. However, the system does not consider cognitive aspects, such as memory, knowledge, and familiarity with the environment.

2.1.2.5 Artificial Vision and Robotics

Analogous to approaches in GIScience, where the focus is on human navigation, landmarks also play an important role in the field of Robotics and Artificial Vision. An open problem in the field of robotics is the challenge of developing robots or agents that are able to learn their geographic environment, reason about it, and navigate through it autonomously in order to achieve some task (rovers for planetary exploration missions, search and rescue robots, etc.). This challenge raises many questions related to navigation and the interaction between agent and environment, and therefore obviously correlates with the aim of our work. Space perception for autonomous robot navigation comes in many styles (Escrig and Toledo 2000). Straightforward approaches, such as the use of pre-designed and pre-selected landmarks (Busquets, Sierra et al. 2002; Kosmopoulos and Chandrinos 2002; Busquets, Sierra et al. 2003), are complemented by more complex approaches involving visual attention and automatic extraction of salient features (Trahanias, Velissaris et al. 1999).

Attention-based models of landmark extraction are typically bottom-up as they extract a set of pre-attentive features (i.e., intensity, color, contrast, etc.), which are assessed in terms of their salience and used to direct the focus of attention. Unlike the primitive approaches using pre-designed and pre-selected landmarks, attention-based approaches promise to answer many questions related to the determination of landmark saliency. Typically, attention-based approaches consider visual stimuli only, which works well for robot navigation. For human navigation, however, cognitive and contextual aspects need to be considered, and hence, the methods need to be adapted accordingly.

2.2 Conceptual Framework

The main contribution of this thesis is a framework for the integrated assessment of the salience of spatial or geographic features. We will first conceptualize our understanding of salience and introduce the terms *Perceptual Salience*, *Cognitive Salience*, and *Contextual Salience*, which constitute a *Saliency Vector* corresponding to the overall salience of spatial objects. Next, we will discuss the components of the saliency vector in more detail and investigate their contributing factors. Finally, we propose a computational approach for the assessment of the contributing factors and their integration.

2.2.1 Conceptualizing Salience for Navigation

The central assumption is that in the domain of navigation, salience emerges from the trilateral relationship between *Observer*, *Environment*, and *Geographic Feature* (Figure 1). As a result, it cannot be attributed to a geographic feature per se. We assume that during navigation, the observer is located in the environment, which is perceived through sensory input. Based on this sensory input and on the task at hand (e.g., sightseeing, driving or walking to some destination), navigators are able to discriminate salient spatial features (i.e., geographic features that highly contrast with the surrounding environment, either perceptually or cognitively) and refer to them as landmarks. These geographic features can be districts, edges or barriers, rivers or lakes, or unique objects (i.e., the classical global landmark), or any feature of the environment that is recognizable and may serve as spatial reference.

The implications of this central assumption are manifold. First, it means that since the observer is located in the environment, only a limited part of the whole environment is perceived. This fact is important because it also means that only those properties of an object that are directly perceived can be used for memorizing, referencing, and identifying potential landmarks from that specific point of view. Reducing the set of properties for the assessment of salience to those that are directly perceived by the sensed stimuli detaches direct experience from prior experience, and hence, draws the line between navigators that have no knowledge of the environment and those who are familiar with the environment. This distinction is important for communication as humans adjust the description of spatial configurations depending on the level of knowledge of the inquirer (Couclelis, Golledge et al. 1995).

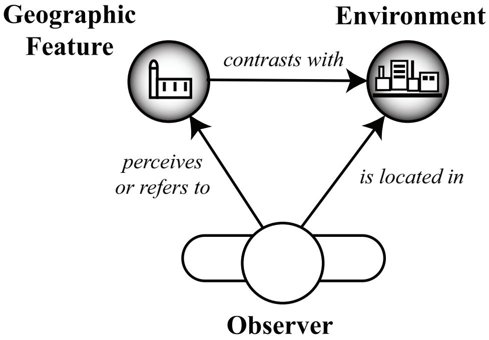


Figure 1 The trilateral relationship between *Observer*, *Environment*, and *Geographic Feature*. The Observer is located in the environment and perceives or refers to some geographic feature, which contrasts with the environment. This configuration defines the basic assumption of our framework.

Second, the assumption that salience is defined by a trilateral relationship also requires that for a feature to be salient, the perceived properties need to contrast with the environment. This requirement implies that in order to assess the salience of a feature, only the perceived physical properties of the geographic features need to be compared, rather than the total sum of their attributes.

Third, the trilateral relationship also accounts for the cognitive abilities of the observer. These include comprehension and use of speech, visual perception and construction, attention and information processing, memory, and executive functions such as planning, problem-solving, and self-monitoring (Newell and Simon 1972; Posner 1998). The amount of cognitive resources being allocated for discriminating potential landmarks depends on various factors, such as the task at hand or the mode of transportation (walking, driving, etc.).

Based on these considerations, we conclude that salience may also be described as the allocation of attention to a salient object, and hence, we base our assessment of the salience of landmarks for navigation on models of attention (Miller 1956; Eriksen and Yeh 1985) and theories of human information processing (Newell and Simon 1972; Gaerling 1999). Attention is a psychological construct that describes detection, selection, discrimination of stimuli, as well as allocation of limited cognitive resources to competing attentional demands (Scholl 2001). Research in cognitive processing has shown that attention is either exogenous (i.e., passive or involuntary) or endogenous (i.e., active or voluntary) (Funes, Lupianez et al. 2005), and that it is influenced by the amount of resources that can be allocated. Figure 2 illustrates the three factors that influence the overall salience of potential landmarks.

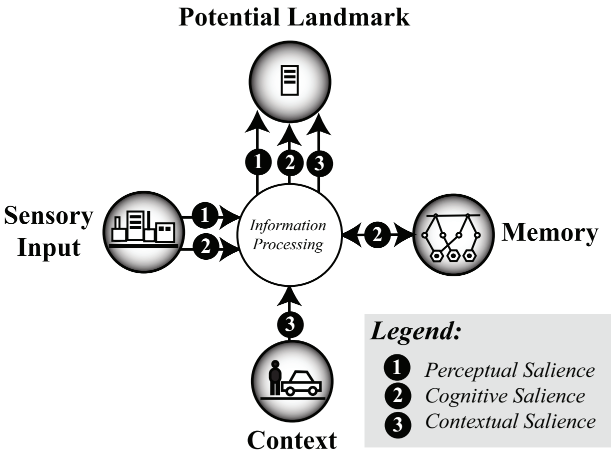


Figure 2 The three different types of salience that contribute to the overall salience of geographic objects: A part of the sensory input contributes directly to the salience of the landmark (Perceptual Salience). Former experience and memory modulates sensory input in a top-down manner and contributes indirectly to salience, and finally, the given context acts as a filter for both perception and cognition, as it defines how much processing resources may be allocated.

Attentional Capture, or the exogenous allocation of attention is described as a bottom-up process in which attention is captured by salient properties of the environment, independent of the observer's intentions (James 1890). Sensory input, such as light, sound waves, or touch is transduced from environmental energy to neuro-chemical energy. If perceptually salient features are received, a capturing effect occurs and attention is automatically directed towards these. For example, if a tall bright building looms in the horizon, probability is high that attention is directed towards this highly salient object, even though it may be irrelevant for the task at hand (Ruz and Lupianez 2002). Control of attention is exerted in a bottom-up manner, as perceived stimuli are directly analyzed for salient properties (Scholl 2001). We will use the term *Perceptual Salience* to refer to effects of attentional capture on a feature's salience.

The endogenous mode of attention is also known as *Attentional Orienting* and is characterized by being initiated actively by the person in a top-down manner (Eriksen and Yeh 1985). Top-down, in this context, refers to the modulation of neural processing via back-projections (i.e., Prefrontal - Parietal - Sensory Control) (Soto and Blanco 2004). Modulation of neural processing occurs when attention is deployed to a stimulus because it is important for achieving some goal. That is, if any of the features are recognized or otherwise considered relevant in the navigation context, we recall them and orient our attention towards them. Hence, the processing of information is based on prior knowledge, while intentions and strategies of the observer are in control of the allocation of attention. In our framework, we will use the term *Cognitive Salience* to refer to the endogenous factors that influence salience.

Finally, the deployment of attention is also based on the amount of attentional resources that can be allocated. If a task is such that it requires full attention of a person, the threshold that separates relevant from irrelevant environmental information is higher than if the task does not require full attention. For example, a tourist on a sightseeing tour is able to discriminate objects in the environment on a higher level of granularity than a bus driver, who needs to allocate much of his attention to traffic. As a result, trip purpose and modality influence the assessment of the salience of geographic features and need to be considered accordingly. In our assessment of salience, we will refer to this kind of influence on attention as *Contextual Salience*.

In summary, our framework (Figure 2) for the assessment of the salience of geographic features introduces three types of salience, namely Perceptual Salience, Cognitive Salience, and Contextual Salience. Perceptual Salience accounts for attentional capture of attention through direct interpretation and discrimination of data received from sensors. Cognitive Salience involves the processes of problem-solving, decision-making, memory, and other aspects of integrative performance into the assessment. Finally, Contextual Salience modulates the assessment in terms of resources that may, or may not determine the salience of geographic features. Within the scope of our framework, we will treat the total salience of a geographic feature as a variable quantity that can be resolved into these three components. As a result, we will use the term *Saliency Vector* to expresses the overall potential of a spatial feature of attracting navigator's attention. In the following sections, we will discuss the components of the saliency vector in more detail and investigate their contributing factors.

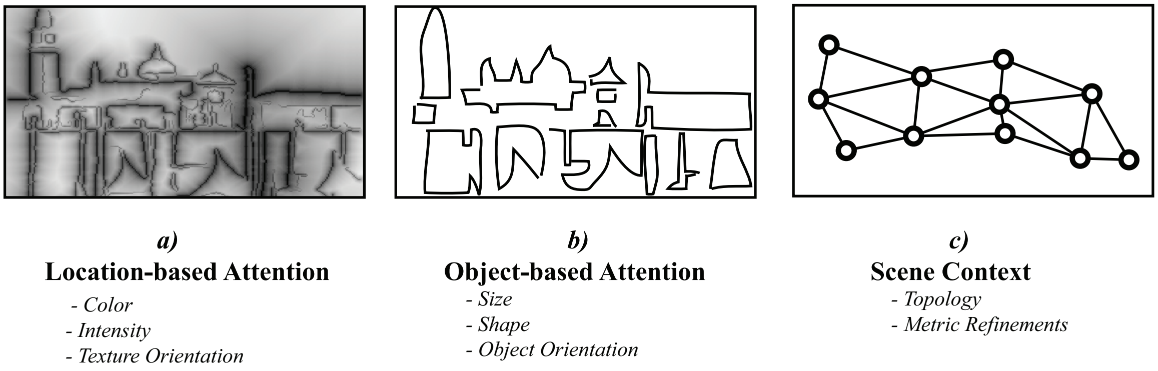
2.2.2 Quantifying the Saliency Vector

The Saliency Vector describes the total salience of a feature or static element of the physical environment. For the purpose of navigation, we restrict the range of spatial features to those that correspond to the definition of landmark as point of reference. Such spatial features include, but are not restricted to the elements of urban environments, such as those described by Lynch (1960). Note that for the rest of this thesis, we refer to spatial features that are potential landmarks as *Spatial Objects*. The following sections define the components of salience of such spatial objects and describe ways to computationally quantify them.

2.2.2.1 Perceptual Salience

Perceptual salience models the bottom-up guidance of attention as it is derived from the part of the environment that is perceived by the navigator from one specific position. The continuous stream of stimuli may be analyzed based on a myriad of criteria (e.g., auditory, olfactory). For our purpose, however, we analyze a snapshot of the visual stream of stimuli. Note that the restriction of the analysis to one stream of stimuli does not affect the basic assumption of the framework. The restriction is due to results from spatial cognition and psychology, which state that the visual stream is the main contributor for the identification of landmarks in the context of navigation (Janzen and Turennout 2004).

The motivation for attention-based assessment of landmarks is the simple hypothesis that landmarks attract attention. There are two dominant divisions of theories in the vast literature of *Visual Attention* research that investigate this hypothesis. The first theory is based on Treisman's model (1980) of S*pace-* or *Location-based Attention* and the second is the developing theory of *Object-based Attention* (see Scholl (2001) for a review).



a)
Location-based Attention
- Color
- Intensity
- Texture Orientation

b)
Object-based Attention
- Size
- Shape
- Object Orientation

c)
Scene Context
- Topology
- Metric Refinements

Figure 3 The three components of Perceptual Salience: a) Location-based Attention, b) Object-based Attention, and c) Scene Context. Each of the components has its own set of attributes, which contributes to the degree of salience of the object.

The main difference between location-based attention and object-based attention is that they use different fundamental units of attention. The focus of location-based attention is on continuous spatial areas of the visual field while the theory of object-based attention holds that visual attention can directly select discrete objects. Although the question of the underlying units has not been definitely answered up to date, it is evident that these two notions, i.e. objects and locations, should not be treated as mutually exclusive (Kubovy, Cohen et al. 1999; Müller and Kleinschmid 2003). Attention may well be object-based in some context, location-based-based on others, or even both at the same time.

In addition to location- and object-based attention, research has shown that attention is also dependent on the concept of the scene, which defines the structure and global semantic characteristics of the scene (see Henderson and Hollingworth (1999) for a review). Results support the idea that *Scene Context* is employed not only for scene recognition and object identification, but also for guiding eye movement, and hence focus of attention (Hayhoe, Shinoda et al. 2000; Shinoda, Hayhoe et al. 2001; Aivar, Hayhoe et al. 2005). We will base our assessment of perceptual landmark salience on these three factors.

Location-based attention assesses the potential for attraction of attention of regions across spatial scenes, that is, attention selects regions in space like a spotlight (Soto and Blanco 2004). All visual stimuli across the visual field are processed in parallel, and the most salient regions are attended. There are many well-known models of spatial attention, such as the guided search model of Wolfe (1994), the spotlight or zoom lens model of Eriksen et. al. (1986), the saliency map model of Koch and Ullman (1985), or the dynamic routing model of Olshausen et. al.

(1992). Common to these approaches is their bottom-up nature and that the visual stimuli are processed in parallel.

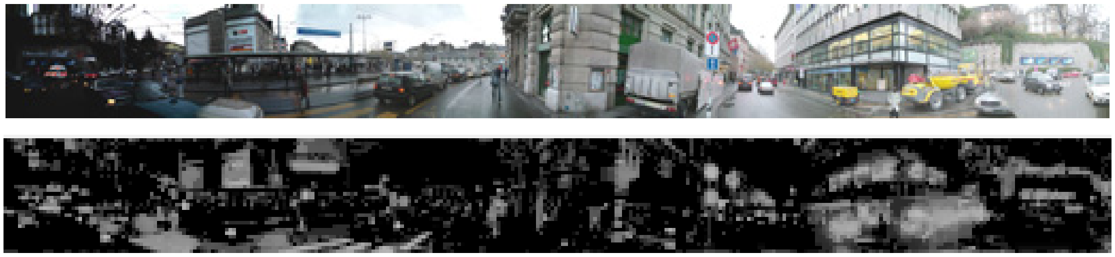


Figure 4 The picture on top shows a typical urban scene and the picture below shows the corresponding saliency map, as generated by Itti and Koch's saliency-based model of spatial attention. Each salient or conspicuous location in an image or a scene is evaluated with respect to its surrounding.

A highly successful implementation of location-based attention is Itti and Koch's saliency-based spatial attention model (Itti, Koch et al. 1998). A saliency map (cf. Figure 4) is used to encode and combine information about each salient or conspicuous location in an image or a scene in order to evaluate how different a given location is from its surrounding. In this biologically-inspired system, an input image is decomposed into a set of multi-scale neural *Feature Maps,* which extract local spatial discontinuities in the modalities of color, intensity and orientation. All feature maps are then combined into a unique scalar *Saliency Map,* which encodes for the salience of a location in the scene irrespectively of the particular feature that detected this location as conspicuous. This model has been shown to perform well on natural scenes, which are at the focus of our research. Therefore we will use the same approach for the determination of location-based attention in our framework.

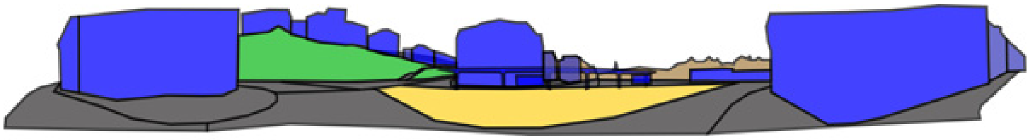


Figure 5 Object-based Attention is influenced by the structure of spatial objects. We base our assessment on the similarity of shape, size, and orientation of objects across the scene.

Object-based attention defines the salience of single objects or groups of objects contained in a scene (Figure 5). In terms of attention theory, the object-based view suggests that attention is directed to objects or perceptual groups based on their structure, instead of locations of particular discontinuities of the visual scene (see Scholl (2001) for a review). Furthermore, location-based attention is blind to geometric properties of spatial objects, which means that features of salience may occur at different scales. The assessment of object-based attention accounts for these properties as it is derived from the object's geometric attributes. Specifically, we derive measures of shape, size, and orientation for objects in the scene, which provide the basis for the assessment of the geometric similarity among objects. We consider location-based and object-based attention in an integrative way. This approach is consistent with results from psychology that state that the two types complement, rather than exclude each other (Soto and Blanco 2004).

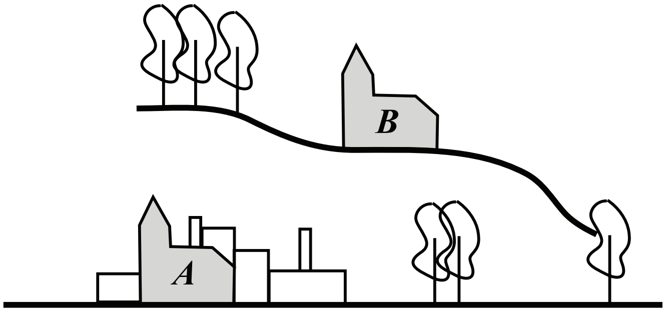


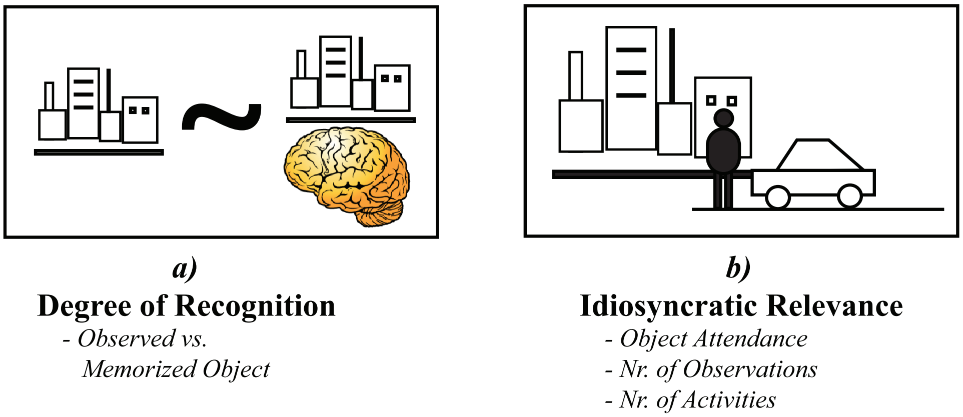
Figure 6 An example of a spatial scene, where objects A and B have the same perceptual attributes, but the spatial configuration provides additional information about the salience of the object.

Scene context focuses on the global type and configuration of a visual scene (Biederman 1972), rather than on single objects. Location-based attention and object-based attention ignore contextual information provided by the type of the scene and the resulting correlation between environment and objects. In our framework, we account for this correlation by assessing scene-based salience and integrating it with perceptual salience. For example, given the case of two perceptually identical objects in a visual scene (Figure 6), their spatial context provides the additional information that object B is further away and higher up than object A. The resulting salience of the objects, hence, needs to be weighted accordingly.

Research results suggest that feature proximity and connectedness are essential elements supporting memorization of the objects (Xu 2006). Accordingly, we assess scene-based salience by means of the binary relations among the objects contained in the spatial scene. The binary relations capture the configuration of the scene, which are then analyzed in term of topology (i.e., adjoin, disjoint), distance, and direction. The result of this assessment is a measure of salience for each binary relation, which, summed up and adjusted with perceptual salience, contributes to the total salience of the object.

2.2.2.2 Cognitive Salience

Cognitive Salience, in contrast to perceptual salience, modulates attention in a top-down manner, as it is dependent on the observer's experience and knowledge (Silva, Groeger et al. 2006). In psychology, the term cognition is often used to refer to the mental processes of an individual. For the context of navigation, we abstract these mental processes to the degree that the mind has an internal representation of the spatial environment and that objects are retrieved from this representation based on the *Degree of Recognition* and the *Idiosyncratic Relevance* of individual objects. We assume that objects with a high degree of recognition are more likely to be used as points of reference than objects with low recognition value. Likewise, we also assume that familiar objects are preferred over unfamiliar objects.



a)

Degree of Recognition

- *Observed vs. Memorized Object*

b)

Idiosyncratic Relevance

- *Object Attendance*
- *Nr. of Observations*
- *Nr. of Activities*

Figure 7 The two components of Cognitive Salience: a) The Degree of Recognition, and b) the Idiosyncratic Relevance. The Degree of Recognition measures how well an object can be identified by an observation, while the Idiosyncratic Relevance indicates the object's personal importance to the observer.

The internal representation of the spatial environment consists of a sequence of waypoints representing a route map, a set of observations for each waypoint along the route, and a set of mental spatial objects defined by a non-empty set of observations from multiple waypoints to this mental object (Figure 8). The motivation for this abstraction of the mental representation of navigational space is the incremental nature of route learning (Siegel and White 1975; Kuipers 1982; Golledge 1992). Observations of specific objects are acquired while navigating and stored in long-term memory, from where they are retrieved if necessary.

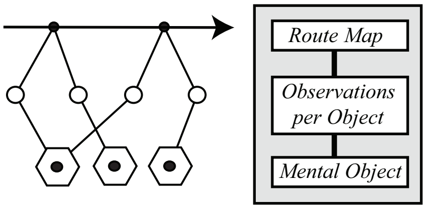
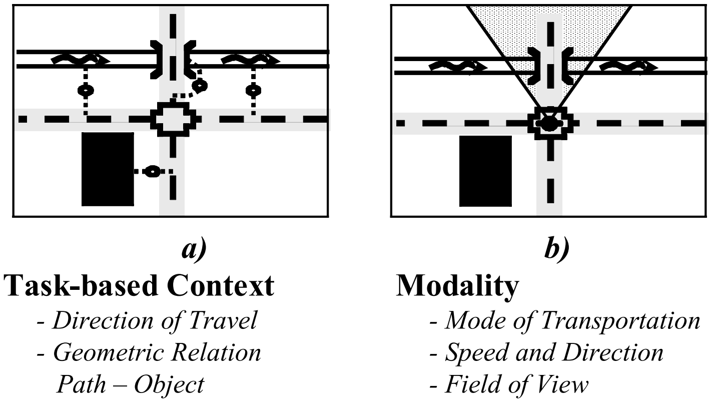


Figure 8 The structure of the route map that is created when navigating: At each waypoint along the route observations to geographic objects are collected. The sum of observations to a single geographic object constitutes a mental object, which we will use in the assessment of cognitive salience.

During the process of reasoning about salience of spatial objects, stored instances of mental objects are considered based on the degree of recognition and idiosyncratic relevance. Recognition occurs when some pattern or object recurs. The basic rule is that recognition is more likely to occur if the current observation matches with the previously stored attributes of that spatial object and vice versa. In order for a spatial object to be recognized, it must be familiar in the sense that it must be linked to at least one observation. Degree of recognition and familiarity, however, are fundamentally different. Recognition, in our framework, is a match between a single observation and a description obtained from a stored instance of a mental spatial object, and as such, is a measure for the degree to which observations from specific points of view support identification of previously observed objects. Analogous to Lacroix (2006), who proposes modeling recognition memory using the similarity structure as input, we will use the similarity between observed object features and mental object features for assessing the degree of recognition.

Idiosyncratic relevance or familiarity, on the other hand, increases with the number of recurrences of a specific object, which basically quantifies the relation individual observers have to specific objects. The term idiosyncrasy is typically defined as a behavioral attribute that is distinctive and peculiar to an individual. In the context of navigation, this behavioral attribute may be defined as the individual familiarity of an observer with respect to a specific object. For example, if the observer recognizes the building where he or she used to work, the relative importance of this object grows compared to other objects. The same pattern applies for public buildings, shopping malls, etc. The idiosyncratic relevance, hence, is determined by the type and number of activities that are associated with individual objects and the frequency by which these activities are performed. The activities and their frequencies are recorded for single objects and set in relation to the objects in the scene. The result of this assessment is a measure of the observer's familiarity with the objects in the scene.

2.2.2.3 Contextual Salience



a)
Task-based Context
- *Direction of Travel*
- *Geometric Relation Path – Object*

b)
Modality
- *Mode of Transportation*
- *Speed and Direction*
- *Field of View*

Figure 9 The two components of Contextual Salience: a) Task, and b) Modality.

Context during navigation plays an important role, as it defines how much attention can be allocated to the recognition and assessment of potential landmarks (Wood, Cox et al. 2006). In our framework, we distinguish between two types of context: 1) *Task-based Context*, which includes the type of task to be performed in the assessment, and 2) *Modality-based Context*, which describes the mode of transportation and the amount of resources that need to be allocated.

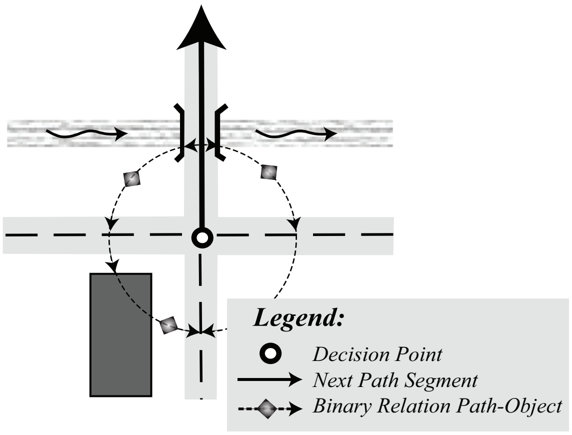


Figure 10 A spatial scene including four possible paths and three potential landmarks (e.g., a river, a bridge, and a building) as experienced by observers during navigation. The binary relation between path and geographic feature defines how valuable geographic features are when considering a specific path.

A definition of the task to be performed during navigation is to state what the goal is, namely to find the route from start to destination. This includes the identification of possible paths and an assessment of the relevance of these paths for achieving the goal (Golledge 1999). This simple definition also points out that navigation is obviously different from tasks such as sightseeing, where navigators follow a route connecting points of interest. In such tasks, the points of interest may overlap with landmarks required to find the way, but this is merely a coincidence rather than a requirement, as the route may well be described only by a subset of the points of interest along the route. In this framework, we consider that navigation itself is the task based on which we assess the salience of spatial objects.

Route instructions that refer to landmarks may take several different forms, as for example ”Walk along the river” or ”Cross the bridge”. Such instructions typically use spatial features to identify the path that is to be followed. Hence, in the context of wayfinding, the choice of landmark is optimized for the identification of the path to be followed. We will use the binary relation between paths and potential landmarks to derive the task-based salience. The binary relation between paths and landmarks is analyzed in terms of topology and metric refinements, where the focus is on distance and orientation between landmark and path. Spatial objects that are located far from the next route segment are of lesser importance than spatially close objects. This approach is analogous to Klippel’s (2005) structural salience of landmarks. In fact, Klippel’s approach captures the idea of task-based salience perfectly and may well be incorporated in future implementations based on this framework. The result of this assessment is a saliency value for each pair of path and potential landmark contained in the visual field. This value describes how salient an object is to a navigator standing at a specific decision point and considering the options available.

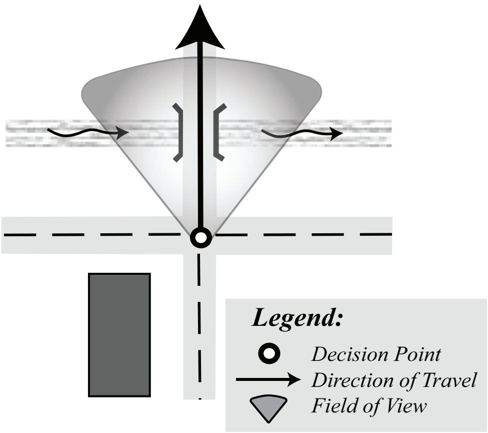


Figure 11 The modality of travel (i.e., walking, driving, or riding) influences both, the cognitive load put on the observer, as well as the degree of physical freedom. The remaining physical and cognitive resources are allocated accordingly, which influences the focus of attention and field of view and hence, the prominence of surrounding geographic features.

Navigation is defined as the combination of wayfinding and locomotion (Montello 2003), whereby locomotion may be achieved through different modes, such as walking, riding, or driving. Each of these modalities has its own requirements in terms of allocation of attention (May, Ross et al. 2003a; May, Ross et al. 2003b; Staal 2004). As a result, each modality will force the navigator to adapt the selection process of spatial objects so that sufficient attention is still allocated to active locomotion. We will assess this type of salience based on the field of view navigators may have when moving about (Figure 11). The field of view is mainly dependent on the speed of the modality and whether locomotion is active or passive (i.e., driving a car vs. riding the bus). These two components allow the definition of a virtual field of view in terms of direction and range, which can be used to assess the importance of potential landmarks. For instance, pedestrians have a field of view that with little effort includes all objects, independent of their spatial location. Car drivers, on the other hand, have a much more limited field of view, since their focus is directed in the direction of locomotion and the range is adjusted to the speed at which they are traveling. The result of this assessment is a ranking of potential landmarks in a scene based on the field of view navigators have when using different modes of transportation.

2.3 Integrated Salience Assessment

So far, we have identified three types of *high-level saliency components* (i.e., perceptual, cognitive, and contextual salience) that define the saliency vector, a set of *auxiliary components* that capture important aspects of salience in terms of attention (i.e., location- and object-based attention, scene context, degree of recognition, and idiosyncratic relevance), and a set of *low-level components* (contrast, size, distance, etc.) that contribute to them (cf. Figure 13). In order to assess the overall salience of spatial objects, these components need to be integrated into a single computational model.

There are a range of cognitive activities that may occur between the time a person first gazes at some feature to the time that relevant information is extracted (Kosslyn 1989). For instance, we know that attentional guidance is a two-stage top-down process whereby the high-level cognitive process of attending alters the low-level processing of visual inputs. The two main questions that arise in this context are how the single components of our framework influence each other and how they may be computationally integrated. We tackle these questions by modeling the human information processing cycle and by integrating a probabilistic approach to describe the interdependence among components this process.

2.3.1 Model of Human Information Processing

One of the most influential theories of visual search is the guided search theory (Wolfe 1994). It suggests a two-stage model of visual processing. In the *pre-attentive stage*, feature maps are computed in parallel in several feature dimensions (e.g., red, blue, green, and yellow feature for color; steep, shallow, left, and right maps for orientation). In the second stage, top-down factors modulate the bottom-up values, and the weighted feature maps are combined additively to form an activation map that eventually guides visual attention in a sequential manner.

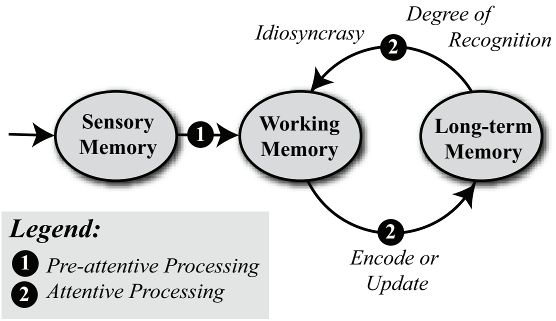


Figure 12 Model of human information processing: Each stage holds a refined perceptual representation of the spatial scene. Pre-attentive processing of the data in sensory memory results in a perceptual scene representation in working memory. Objects in the perceptual scene representation are then assessed sequentially for salient features, and finally, objects in long-term memory are updated with new facts.

In our approach, we propose a similar model for the assessment of salience. Specifically, we propose a model of human information processing that divides the assessment of salience in three stages and that accounts for the characteristics of landmarks as discussed before (Figure 12). The three stages correspond to the types of memory involved, namely *Sensory Memory*, *Working Memory*, and *Long Term Memory*, and are linked together by a set of computational processes (i.e., pre-attentive, attentive processing, encoding, update, recognition, and familiarity).

Each stage is a refinement of the former in terms of salience assessment. In the first phase, the visual stimuli are perceived and stored in Sensory Memory. At this stage, no processing is involved yet. Before reaching the second phase, i.e. working memory, the stimuli undergo the process of pre-attentive processing, which simulates the ability of the low-level human visual system to rapidly discriminate objects and identify certain basic visual properties (Treisman, Vieira et al. 1992). Pre-attentive processing, hence, produces a *Perceptual Representation* of the spatial scene in working memory that contains the spatial objects and quantifies their low-level components (e.g. size, length, color, intensity).

The objects in the Perceptual Representation of the scene are now ready for further processing. Unlike in sensory memory, where stimuli are processed in parallel, objects in working memory are processed sequentially. Sequential processing in working memory simulates the process of attentional orienting and includes top-down factors (i.e., degree of recognition and familiarity with object) and contextual factors (i.e., task and modality), which modulate the perceptual salience of the object. Finally, the objects are either encoded in memory (i.e., a new mental object is created in long-term memory) or, if the object is already present, updated with the new information (i.e., the new observation is attached to the object). Updating objects in long-term memory ensures that the saliency of objects evolves over time and varies with the level of experience of observers.

2.3.2 Integration of Components

In our model, pre-attentive processing is understood as the process of discriminating spatial features and extracting low-level components from a set of visual stimuli. Attentive processing, in contrast, describes the process of sequentially assessing the salience of spatial objects in the scene by integrating the low-level components and computing the three components of the saliency vector. While we assume that the low-level components are independent and contribute equally to the auxiliary components (e.g., location-based attention, object-based attention, scene context), we need to analyze and find a way to model the mutual influence auxiliary components have on the high-level components of salience, that is, how they contribute to perceptual salience, cognitive salience, and contextual salience. For this purpose, we propose to apply a probabilistic inference model, which is able to deal with the complexity and uncertainty of human information processing.

Probabilistic inference models are increasingly becoming important theoretical tools for understanding cognition (Scholl and Tremoulet 2000; Kersten and Yuille 2003; Kersten, Mamassian et al. 2004; Chater, Tenenbaum et al. 2006). Following this trend, we propose to use a *Bayesian* or *Belief network* to model the interdependence of the auxiliary components and assess the overall saliency. The main reason for this approach is that Bayesian methods allow the development of quantitative theories at the information processing level and that they are able to model *Causality*, which plays an important role in human reasoning (Gigerenzer and Murray 1987). Furthermore, recent work has shown that the Bayesian perspective yields a uniform framework for studying object perception (Kersten 2002).

The concept of causality or causation refers to the set of all particular *causal* or *cause-and-effect* relations (Lewis 1973). For better understanding consider the following simple example: When a building stands out among other buildings, it will be salient. The core idea of Bayesian networks, hence, is that based on causal knowledge we are able to causally explain probable outcomes given known relationships between certain actions and consequences, i.e. "a taller building is more likely of attracting attention" is based on the probable cause (taller building) of the effect (attracting attention).

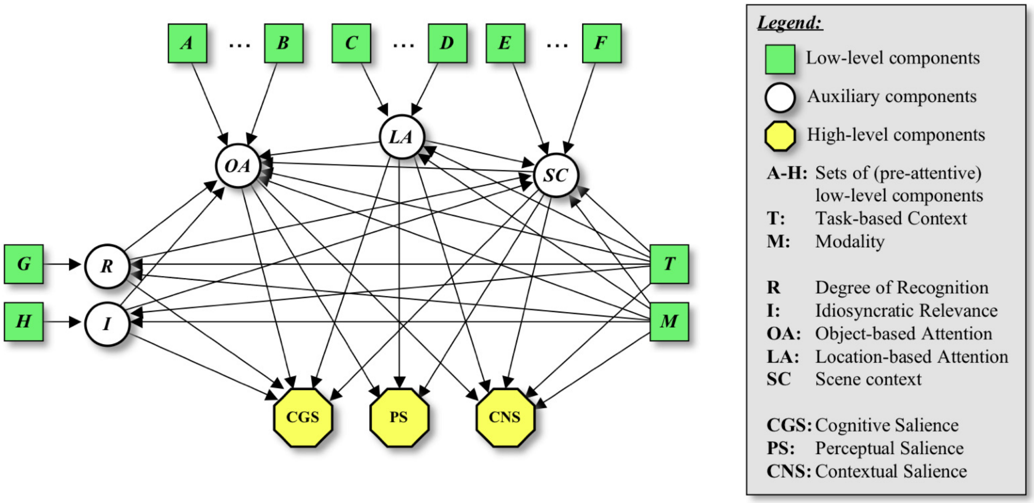


Figure 13 The structure of the Bayesian network used for simulating the salience assessment process. The low-level components are derived directly from input data and serve as evidence. The auxiliary components account for the different types of attention, and the high-level components describe the resulting saliency of the observed spatial object.

Bayesian networks describe conditional independence among subsets of variables or concepts and allow combining prior knowledge about independencies and dependencies among variables with observed data. Formally, a Bayesian network is a directed acyclic graph that contains a set of nodes, which represent random variables, and a set of directed links connecting pairs of nodes and denoting causal dependencies between variables (Jensen 2001). The strengths of the dependencies are expressed by *conditional probability distributions* attached to every node. Nodes can represent any kind of variable, be it a measured parameter (e.g., color, shape), a latent variable (e.g., location- or objects-based attention), or a hypothesis.

In our model, we have a set of low-level components, a set of auxiliary components, and a set of high-level components (Figure 13). We will employ these components as nodes of the Bayesian network. The next step is to define the structure of the Bayesian network, that is, to identify the dependencies among the nodes. Although the interaction between the single components of our model has not been fully investigated and answered yet, available evidence provides a basic idea of the causal structure among the nodes of the Bayesian network. The most important aspects are listed below:

- Task and modality function like a filter for perceptual and cognitive abilities and hence, influence all other components, including what is currently perceived (Williams 1988),
- Location-based attention is the result of attentional capture, and therefore, only dependent on available perceptual input (Treisman and Gormican 1988),
- Object-based attention and scene context are influenced by top-down factors (i.e., degree of recognition and idiosyncrasy) and by the amount of available resources (task and modality) (Serences, Schwarzbach et al. 2004; Staal 2004), and finally,
- Scene context influences the allocation of attention to specific objects (De Graef, Lauwereyns et al. 2000).

Furthermore, we assume the following to complete the structure of the Bayesian network:

- Both types of attention (i.e., location-based attention and object-based attention) and scene context influence the high-level components equally,
- The degree of recognition and idiosyncratic relevance influence cognitive salience, and finally,
- Task and modality modulate contextual salience.

These results from previous research and our own assumptions yield the Bayesian network depicted in Figure 13. The next step is to assign values to the nodes of the network. All low-level components are either observed directly or computationally derived from input data, and hence, serve as evidence. For each node holding evidence, we derive the probability of salience from the corresponding sets of object attributes, that is, we compute the likelihood of salience for each low-level component as a statistical function of all objects in the scene.

In order to fully specify the Bayesian network and thus fully represent the joint probability distribution, it is necessary to further specify for each node X (i.e., auxiliary and high-level components) the probability distribution for X conditional upon X's parents. The distribution of X conditional upon its parents may have any form. It is common to work with discrete or Gaussian Distributions since that simplifies calculations (Jensen 2001). For our model, we will use a discrete probability distribution and assume uniform influence of the parents.

The last step in computing the posterior distribution of variables given evidence is called *Probabilistic Inference* (Jensen 2001). The posterior probability gives sufficient statistics for detection of salient spatial objects, that is, the posterior probability sufficiently explains the likelihood of each component of the saliency vector to be a salient property, considering the objects in the current scene, knowledge of the observer, and the current context.

2.4 Summary

The aim of this chapter was to review related work and to propose a framework for the assessment of salience of spatial objects tailored to the requirements of human navigation. To achieve this goal we conceptualized our understanding of salience, investigated what factors influence the prominence of spatial objects, and proposed a computational framework that combines the different factors in order to determine the object's salience. We introduced the concept of the Saliency Vector, which accounts for the trilateral relationship between observer, observed object, and environment in terms of Perceptual, Cognitive, and Contextual Salience. Further, we investigated the role of attention in the assessment of saliency and used the theories of location-based attention and object-based attention, together with the context of the scene to identify and classify the low-level components (bottom-up and top-down) that modulate salience. Finally, we examined the interdependencies among the components and suggested using a Bayesian network to integrate them into a single computational model.

Chapter 3
Quantifying Saliency

This chapter investigates the translation of the content of a spatial scene into a form that can be used to assess the salience of spatial objects contained therein, and provides a detailed description of the saliency assessment process. The translation mechanism and the resulting representation of the spatial scene are key to the assessment process, as the mechanism extracts those features from the spatial scene that will be used to define the likelihood of low-level features to contribute to the object's saliency vector. We model the translation mechanism and the assessment process in terms of a computational model. Note that in the context of this work, we treat computation as a general term for information processing that can be represented mathematically, or in a more narrow meaning, a process following a well defined model that is understood and can be expressed formally.

The reminder of this chapter is organized as follows: Section 3.1 provides a general overview of the computational strategy, including a description of the involved entities, the data model, and input data. Section 3.2 introduces the computational model, which is a formal implementation of the conceptual framework proposed in Chapter 2. Finally, Section 3.3 concludes the chapter with a summary of the proposed computational model.

3.1 Overview of the Computational Strategy

The act of seeing starts when the lens of the eye focuses an image of the outside world onto a light-sensitive membrane in the back of the eye, called the retina (Howard and Rogers 2002). The retina serves as a transducer for the conversion of patterns of light into neuronal signals. The lens of the eye focuses light on the photoreceptive cells of the retina, which detect the photons of light and respond by producing neural impulses. These signals are processed in a hierarchical fashion by different parts of the brain in order to assimilate information from the environment that helps guide our actions. This process forms the base for our computational model.

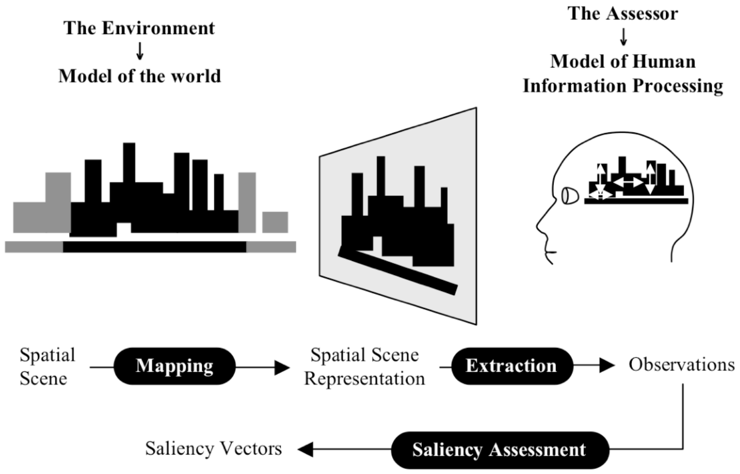


Figure 14 Overview of the computational strategy.

Because the act of seeing is continuous in nature, and therefore hard to discretise, we abstract the process to include the entities and functions depicted in Figure 14. In addition, we limit the act to the processing of single snapshots of the environment. The snapshot of the spatial environment, which we will call Spatial Scene, is mapped onto the retina where it constitutes a representation of the spatial scene, or a Spatial Scene Representation. From this representation we extract a set of object-specific parameters, collectively referred to as Observations, which are subsequently used for assessing the object's saliency. In the following sections we describe the underlying data models for environment, assessor and assessment process. Note that formalizing the conceptual framework in terms of a computational model, in essence, is an abstraction process that approximates reality by formal mechanisms. While some of the mechanism that we propose in this computational model are rather crude approximations of reality, they are all replaceable by more sophisticated methods, without changing the general idea of the model.

3.1.1 The Environment

The model of the environment is an integral part of the computational model. We will treat the model of the environment as a representation of the structure and the properties of the real world. In particular, the model of the environment specifies the types of spatial objects that are part of the environment along with their properties, geometric structure and position with respect to a reference frame, and the spatial configuration, including topological and geometric relations among the objects.

3.1.1.1 Environmental Model

The environmental model is tailored to navigation in cities, or urban environments, such as those investigated by Lynch (1960). Such environments are typically made up of a set of feature types. Lynch argues that five distinct elements, namely *districts*, *edges*, *paths*, *nodes*, and *landmarks*, define the image of the city. The elements are characterized as follows:

- **Districts:**
  Medium-to-large sections of the city, that are conceived of having two-dimensional extent, which the observer mentally enters inside-of (business district, etc.).
- **Edges:**
  Linear elements of urban space that are not used or considered paths by the observer, such as boundaries between two phases (shores, railroad cuts, etc.).
- **Paths:**
  Channels along which navigators customarily, occasionally, or potentially move (streets, walkways, canals, railroads, etc.).
- **Nodes:**
  Strategic points in a city into which navigators can enter, and which are the intensive focus of traveling (junctions, railway stations, etc.).
- **Landmarks:**
  Points of reference, where the navigator does not enter, but rather uses it as reference from an external point of view (building, sign, mountain, store, monument, etc.).

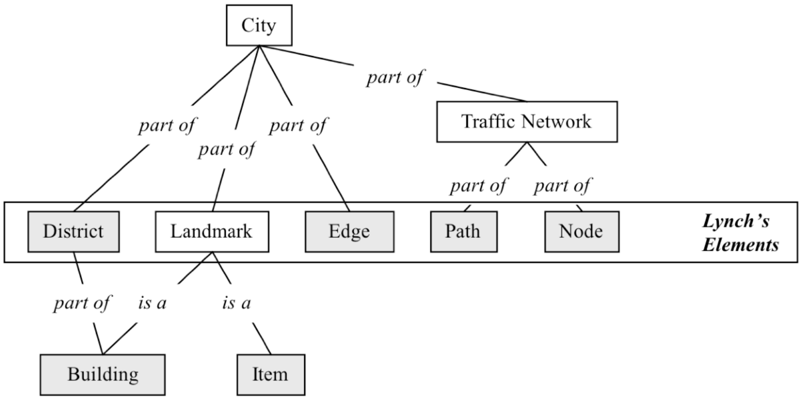


Figure 15 Our model of the environment includes six distinct elements, namely districts, edges, paths, nodes, buildings, and items.

Lynch's elements build the base for the model of the environment. One basic assumption of our conceptualization of saliency, however, is that all elements of urban space are potential landmarks, which results in a terminological conflict with Lynch's vocabulary. Therefore we split Lynch's landmark element in two separate concepts, namely *building* and *item*. In the context of urban navigation, these two concepts capture the semantics of Lynchs's landmark in a consistent way. As a result of these considerations, we redefine Lynch's urban environment as a collection of spatial objects that includes *districts*, *edges*, *nodes*, *paths*, *buildings*, and *items* (cf. Figure 15).

3.1.1.2 Specification of Spatial Objects

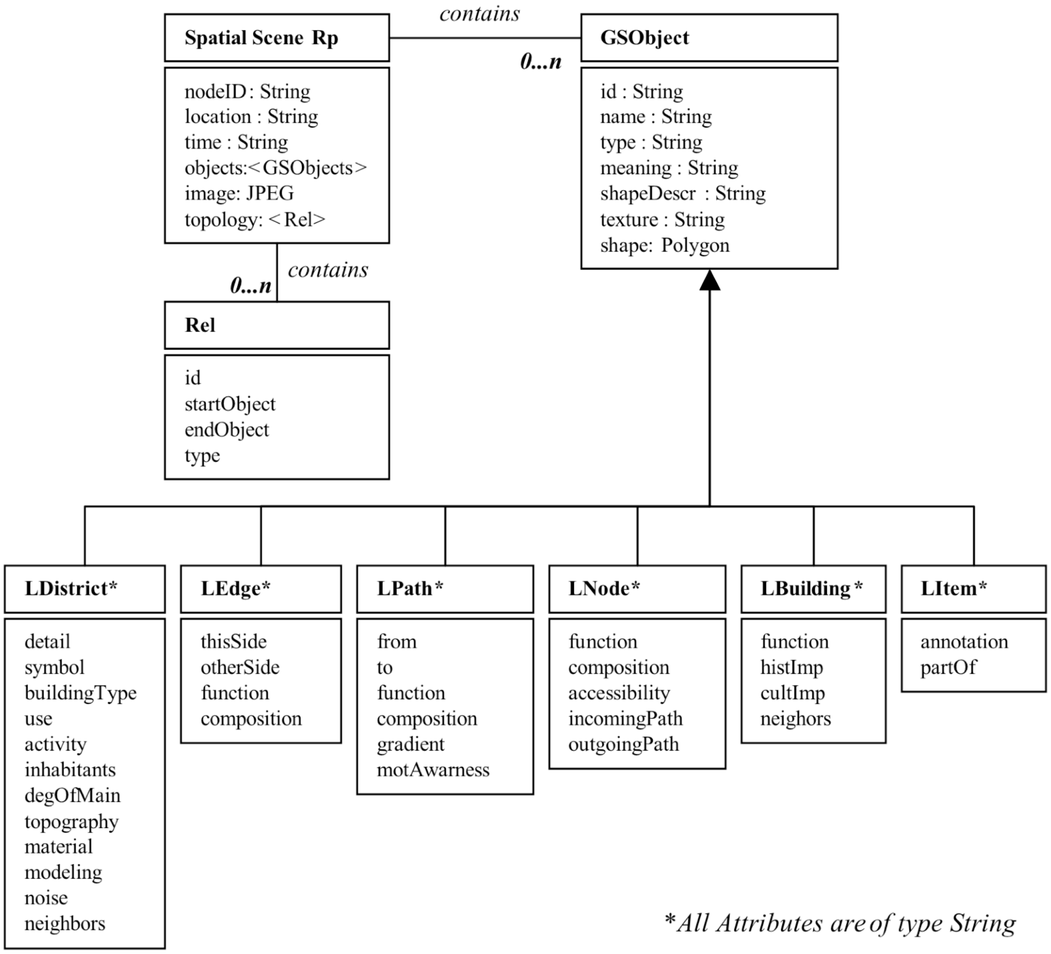


Figure 16 This class diagram summarizes the types of classes and attributes that constitute a representation of a spatial scene.

The conceptualization of salience is based on the situated nature of navigation and is expressed as the tri-lateral relationship between navigator, observed object, and environment. This conceptualization implies that the spatial scene contains observations to a subset of spatial objects, that is, observation to those objects that are perceived from the navigator’s current position. It also implies that the content of the scene changes as a function of the navigator’s position, that is, if the position of the navigator changes, a new scene of the environment is generated and examined. Every scene is an instance of the class diagram depicted in Figure 16, which also summarizes the classes and their attributes.

3.1.1.3 Relations between Spatial Objects

The layout of spatial objects in the scene is characterized in terms of a set of binary relations between pairs of objects. The binary relation between spatial objects conveys important information about the salience of objects. Such information includes topological relationships (occlusion, adjacent, adjoin, etc.), but also metric information (distance, direction, etc.). Due to the computational complexity of assessing the topology of objects in spatial scenes, however, we restrict the set of topological relations for this work to adjacent and disjoint.

3.1.2 The Assessor

The role of the assessor is to sense the environment, extract relevant information, and to assess the saliency of spatial objects contained therein. Sensing the environment, in this context, is equivalent to the instantiation of a representation of the spatial scene, and corresponds to the input data. The next step is to extract a set of object-specific observations from this representation. This step simulates a series of cognitive processes and involves memory and previous knowledge. The final task of the assessor is to derive the salience of spatial objects based on the set of observations. The following sections provide a detailed description of the steps involved in the extraction process, the model of memory, and the assessment process.

3.1.2.1 Information Processing

The process of extracting information from the sensed input (i.e., spatial scene representation) consists of two distinct steps: 1) a pre-attentive processing step, where the input data is prepared such that object-based statements are possible, and 2) an attentive processing step that estimates the relative importance of spatial objects based on former experience and knowledge stored in long term memory. The two steps prepare the raw input data for the assessment process.

3.1.2.2 Long-term Memory

As the navigator moves along a path, a route map is created in long-term memory. The route map consists of an ordered list of nodes, one for each scene that the navigator comes across while traveling, a set of observations, and a set of mental objects (Figure 17). Each node is associated with a set of observations to spatial objects, i.e., a vector of measurements that characterizes the spatial objects from the given point of view. The set of observations to the same spatial object constitutes a mental spatial object, which is used to reason about object-specific prior knowledge and experience.

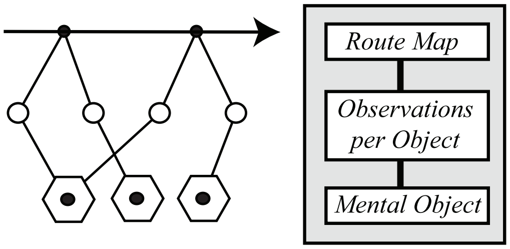


Figure 17 The model of long-term memory used in our model. The route map consists of the nodes visited along some route, along with the observations made at each node, and the set of mental spatial objects created by those observations.

The route map is updated every time a new scene is sensed. Updating the route map consists of adding the new node and the set of observations to the route map, as well as updating the set of mental objects. Updating the set of mental objects consists of figuring out whether the observations at the current node include spatial objects not observed previously, and adding new objects, if necessary.

3.1.2.3 The Assessment Process

The assessment of the saliency of spatial objects is based on the principle of causality. This principle is closely related to human reasoning, as humans are not indifferent to causal relations (Tversky and Kahneman 1977). We understand causality as the relation between a cause and the effect it produces, whereby the observations of a spatial object (or a subset thereof) form the cause that produces a more or less strong effect in terms of salience. This definition implies that we are uncertain about the degree of causation of single components, as the degree may vary from case to case. Therefore, we resort to Bayesian probability in the form of a Bayesian Network, which is a formalism that allows reasoning about beliefs under conditions of uncertainty, to model the assessment process.

We will use the Bayesian network to calculate the probability that an object is salient given a set of observations. The set of nodes of the Bayesian network consists of a set of low-level components (e.g., length, height, color, etc.), a set of auxiliary components, and a set of high-level components. The set of low-level components represents the probability of each component to be salient within the set of spatial objects, which is supported by evidence derived from observations. The auxiliary components model the influence of the low-level components, and the high-level components represent the components of saliency.

The Bayesian network provides the base for the salience assessment process. The assessment process is of predictive nature and uses Bayes' formula to compute the conditional posterior probability that the hypothesis (e.g., object A is salient) is true, given what was observed (e.g., probability of color to be salient). The assessment process produces a ranking of the spatial objects in terms of perceptual, cognitive, and contextual salience.

3.2 Computational Model

This section gives a detailed description of the computational model underlying the assessment of landmark saliency for navigation. We will presume that a route plan is defined in advance, which provides the sequence in which the scenes are processed. Generally speaking, we wish to solve the following problem:

Given:

- A raster image I that represents the spatial scene at the current position of the navigator ϕ;
- A set of raw measurements for the properties of spatial objects SO, which are visible from the same position ϕ, whereby all instantiations of spatial objects have a class and a set of attributes (cf. class diagram in Figure 16);
- A set of binary spatial relations REL between spatial objects as perceived from position ϕ; and
- The context in which the measurements were made, i.e., the modality of travel γ and the path ψ that identifies the route segment for continuation of the journey.

Find:

- The components of the saliency vector, i.e., the probabilities of perceptual, cognitive, and contextual salience for every spatial object in the scene.

Note that we use the term measurements to refer to the raw observations of spatial objects. We use the following algorithm to solve this problem:

```
Start Saliency Assessment
     Sense Environment at position $\phi$
     Perform Pre-attentive Processing on $I$
     FOR each Object in Spatial Scene Representation
          Perform Attentive Processing
     End FOR
     Update Long-term Memory
     FOR each Spatial Object
          Assess Dissimilarity among Objects
     End FOR
     Compute Prior Probability Distributions
     Perform Probabilistic Inference
End Saliency Assessment
```

A more detailed version of the algorithm is illustrated in the following diagram:

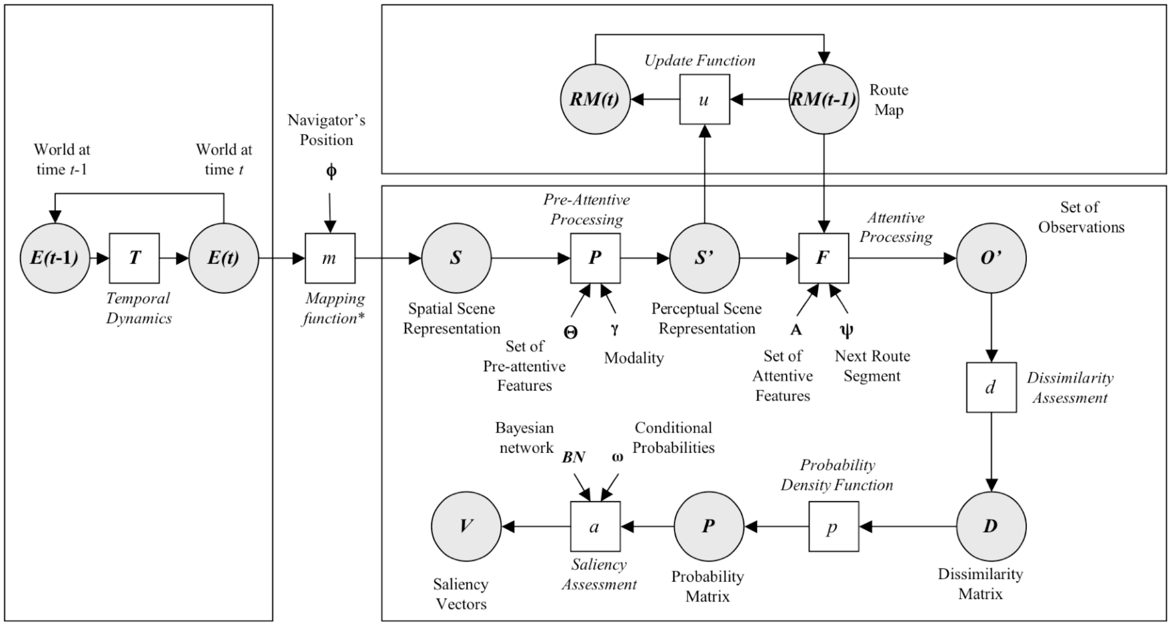


Figure 18 This diagrammatic representation of the assessment process illustrates the steps involved and the sequence in which they are executed.

Note that the diagrammatic representation of the assessment process in Figure 18 includes both, a model of the world and a model of human information processing. In our computational model, however, we consider only the translation of the content of the spatial scene representation into a form that can be used for assessing the salience, along with the saliency assessment process. We do not consider the model of the world and its temporal dynamics nor do we formally define the mapping function. Instead, we will use a raster image as snapshot of the environment, and replace the mapping function by digitizing the outlines of spatial objects contained in the image. Furthermore, we simulate basic cognitive functions, such as text processing, by human input. Nevertheless, the representation of the spatial scene contains only features that are either present in the environment, or that will be an integral part of models of the real world (e.g., 3D-city models).

3.2.1 Quantification of Scene Content

The quantification of the scene content is divided into the two steps of pre-attentive and attentive processing, each consisting of a set of functions applied on the respective input data. Note that these steps are by no means exact and comprehensive simulations of the corresponding stages of human information processing, but rather employed to describe what features are extracted when in our computational model. The model abstracts the processes such that in the pre-attentive processing stage a subset of pre-attentive features is extracted from the raw observation data set (bottom-up), which is then refined and modulated in the attentive processing stage by data stored in long-term memory (top-down).

3.2.1.1 Pre-attentive Processing

"Pre-attentive processing of visual information is performed automatically on the entire visual field detecting basic features of objects in the display. Such basic features include colors, closure, line ends, contrast, tilt, curvature and size. These simple features are extracted from the visual display in the pre-attentive system and later joined in the focused attention system into coherent objects. Pre-attentive processing is done quickly, effortlessly and in parallel without any attention being focused on the display."

(Treisman 1985; Treisman 1986)

Conversion of Scene Representation into Perceptual Representation

In our model, the pre-attentive processing stage takes the raw observation data, which we will refer to as the Spatial Scene Representation *S*, as input and extracts a set of pre-attentive object features. The result of this extraction is a Perceptual Scene Representation *S*' that is ready for further processing. Formally, the Representation of a Spatial Scene *S* is defined as

$$S = \left(I, SO, REL\right) \qquad (1)$$

where *I* is a 360° RGB-image of the spatial scene, *SO* is the set of spatial objects in the scene, and *REL* is a subset of the binary topological relations among the spatial objects in the scene. Each instantiation of a spatial object $so \in SO$ has a class $c \in C$ (cf. data model) and is characterized by a subset of string attributes $od \in OD$ that serve as object descriptors (cf. class diagram). Furthermore, the geometric outline of each instantiation of a spatial object in the visual field is defined by a polygon, which we will refer to as *shape* of the spatial object. As a result, the instantiation of a spatial object is formally defined as

$$so = \left(c, od, shape\right) \tag{2}$$

Set of Pre-attentive Features Θ

The set of pre-attentive feature parameters Θ that is extracted from the raw input data is defined as

$$\Theta = \left\{\theta_i\right\}_{i=1}^{n} \tag{3}$$

where θ is a specific pre-attentive feature and n denotes the number of pre-attentive features. The set of pre-attentive features includes geometric and non-geometric visual properties, as well as properties that define the role of the object with respect to the configuration of the scene and the influence of modality. The following list shows the set of pre-attentive features Θ used in this model, together with references to research that showed they were pre-attentive:

- Length (Treisman and Gormican 1988)
- Width (Sagi and Julész 1985; Treisman and Gormican 1988)
- Size (Treisman and Gelade 1980)
- Color (Nagy and Sanchez 1990; D'Zmura 1991; Kawai, Uchikawa et al. 1995; Bauer, Jolicoeur et al. 1996)
- Intensity (Beck, Prazdny et al. 1983; Treisman and Gormican 1988)
- Orientation (Julész and Bergen 1983; Wolfe, Friedman-Hill et al. 1992)
- Number (Julész and Bergen 1983)
- Position (Julész and Bergen 1983; Treisman and Gormican 1988)

The set of pre-attentive features is completed with two features for shape refinement and modality. The reason for including these two features in the set of pre-attentive features is of practical nature, as they will be used in the attentive processing stage, but are not stored in long-term memory.

Set of Pre-attentive Processing Functions

Given a Spatial Scene Representation *S,* a set of pre-attentive feature parameters Θ, and the modality γ (e.g., walking, driving), a set of pre-attentive feature extraction functions *P* is defined as

$$P : S \rightarrow S' \tag{4}$$

which extract a real-valued feature set $o \in O$ that characterizes the pre-attentively processed observations of spatial objects. Pre-attentive processing integrates information contained in the image *I*, the binary spatial relations *REL*, and the instances of spatial objects $so \in SO$ into a set of coherent observations of spatial objects. We term the resulting set of observations Perceptual Scene Representation *S*' and define it as

$$S' = \left(sd, O\right) \tag{5}$$

where *sd* denotes the description of the scene, and *O* the set of spatial object properties for that scene. The object properties of a spatial object are defined as

$$O = \left\{\left(od_i, pop_i\right)\right\}_{i=1}^{n} \tag{6}$$

where *od* denotes the class-specific subset of objects descriptors (cf. class diagram Figure 16), *pop* the set of pre-attentive object properties extracted from *S*, and *n* the total number of spatial objects in the scene. The difference between *S* and *S*' is that the set *S* consisting of image data, a set of spatial objects, and a set of binary spatial relations has been converted in a coherent set *S*' of object descriptors *od* and a set of pre-attentive object properties *pop*. In the following sections, we define in detail the set of pre-attentive processing functions *P*, which include functions for extracting geometric object properties, non-geometric visual properties, scene-related properties, and the influence of modality.

Extraction of Geometric Object Properties

The geometric properties considered in our model include the pre-attentive features of length, width, and size. In natural scenes, the perception of these properties is dependent on the distance between observer and observed object, whereby the ratio of perceived object size to perceived distance is constant. This is known as the *size-distance invariance hypothesis* (Howard and Rogers 2002). Consequently, for a given image size, the perceived object size is proportional to the perceived distance. We will employ this hypothesis in order to assess length, width, and size of spatial objects.

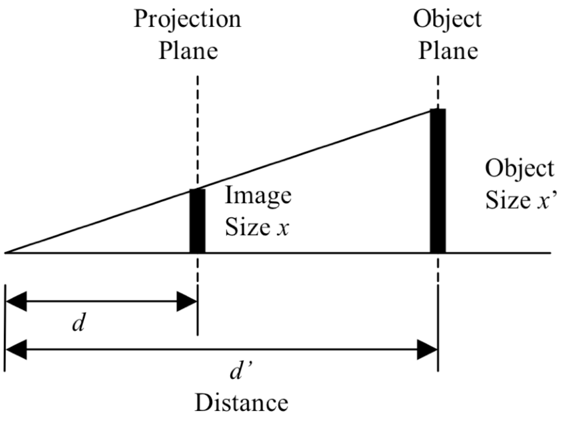


Figure 19 The picture illustrates the size-distance invariance hypothesis.

We use an image as input data, which implies that the spatial objects are projected onto the same plane with the same focal distance. This allows treating the distance to the projected object as constant (i.e., *d* is set to 1), and therefore, using the following equation for estimating the perceived dimensions of spatial objects (proof is left to the reader)

$$x' = d' \cdot x \tag{7}$$

Extraction of perceived distance, or depth, from images of natural scenes is a challenging task. The ability of humans to perceive depth is a function of the arrangement of objects in the perceptual environment, the capacities of the eyes, and the interpretive processes of the brain. There are two types of visual depth perception cues; 1) monocular cues, which are those requiring one eye, and 2) binocular cues, which require two eyes. Given our input data (i.e., panoramic image and object's geometry), we will focus our attention on static monocular cues in order to estimate distances to spatial objects.

The range of static monocular cues includes height in the visual field, relative size, shading, converging lines, saturation of colors, occlusion, and texture gradient (Howard and Rogers 2002). For our purpose, however, we will concentrate on the height of objects in the visual field for distance estimation. This simple heuristic assumes that objects located higher in the visual field are farther away, which works well for outdoor environments, including urban scenes, but not for indoor environment. Formally, we define the distance to spatial objects as

$$D_{so_i} = \begin{cases} \mathbf{elevation}\left(MBR_{so_i}\right), & \text{if } c_{so\,i} \in \{LEdge, LBuilding, LItem\} \\ \mathbf{elevation}\left(CENT_{so_i}\right), & \text{if } c_{so_i} \in \{LDistrict, LPath, LNode\} \end{cases} \tag{8}$$

where so_i is the spatial object under scrutiny, *MBR* is the Minimum Bounding Rectangle of the shape of object so_i, *CENT* is the centroid of the shape of the spatial object so_i, and *c* is the class of the spatial object. We define the centroid *c*=(*x,y)* as

$$CENT_{so_i} = \mathbf{centroid(}shape_{so_i}\mathbf{)} \tag{9}$$

where *so* is the observed object, centroid the extracting function, and *shape* the object's geometric outline.

The reason for using different approaches for distance estimation of the spatial object is the different perception of the spatial extent in the panoramic picture. One group of spatial objects (i.e., districts, buildings, and items) is in general perceived as horizontally extended, while the other group (i.e., edges, buildings, and items) is perceived as vertically extended. The individual distance estimation accounts for this peculiarity.

Width and Height

Based on the distance estimation, we can now compute the perceived geometric properties of the spatial object. Length and width are derived from the Minimum Bounding Rectangle MBR that is defined by the shape of the spatial object.

$$H_{o_i} = D_{so_i} \cdot \mathbf{height}\left(MBR_{so_i}\right) \tag{10}$$

$$W_{o_i} = D_{so_i} \cdot \mathbf{width}\left(MBR_{so_i}\right) \tag{11}$$

Size

The size of observed spatial objects is a two-dimensional property and therefore changes exponentially with distance.

$$A_{o_i} = D^2_{so_i} \cdot \mathbf{area}\left(shape_{so_i}\right) \tag{12}$$

Extraction of Refined Shape Properties

Lynch (1960) identifies shape description as one of the properties that enhances memorability. So far, we have only considered width and length as shape descriptors of spatial objects. Attentive processing refines this characterization by calculating the elongation and the compactness of the spatial object's shape.

Elongation

We derive the elongation of the spatial object from the width and length of the Minimum Bounding Rectangle.

$$EL_{o_i} = \frac{\mathbf{width}\left(MBR_{so_i}\right)}{\mathbf{length}\left(MBR_{so_i}\right)} \tag{13}$$

Compactness

The compactness measure of a spatial object in the scene is derived from its shape and is a numerical quantity representing the degree to which the shape is compact. We will use the circularity ratio as measure of compactness. The circularity ratio measure is the ratio of the area of the shape to the area of a circle having the same perimeter, that is, to the most compact shape with the same perimeter. For a circle the ratio is one, while for an infinitely long and narrow shape, it is zero. Formally, compactness is defined as

$$CP_{o_i} = \frac{4 \cdot \pi \cdot \mathbf{area}\left(shape_{o_i}\right)}{\mathbf{perimeter}\left(shape_{o_i}\right)^2} \tag{14}$$

Extraction of Non-geometric Object Properties

Color, Intensity, and Orientation

We base the extraction of color, intensity, and orientation on Itti and Koch's (1998) implementation of the saliency-based model of bottom-up attention by Koch and Ullman (1985). The system computes saliency-based bottom-up locations of attention from an input image. It has been verified in human psychophysical experiments (Itti 2005; Peters, Iyer et al. 2005), and it has been applied to object recognition (Miau and Itti 2001; Walther, Itti et al. 2002) and robot navigation (Chung, Hirata et al. 2002). Unlike Itti and Koch's model, however, we do not compute a single saliency map from maps for color, intensity, and orientation contrast, but use the single feature (or conspicuity) maps to derive object-specific values for color, intensity, and orientation. In the following sections, we give an overview of the process used to create the maps. For a detailed discussion of the process, however, we refer to Itti and Koch (1998).

Overview of the Extraction Process

The RGB input image of the spatial scene is low-pass filtered and sub-sampled in order to create a Gaussian pyramid. The pyramid has a depth of $\sigma = 9$ scales that provide horizontal and vertical image reduction factors ranging from 1:1 to 1:256. For each level of the pyramid, an intensity image *Int* is calculated, resulting in a Gaussian pyramid of intensity images. The intensity image is obtained as

$$Int = \frac{r+g+b}{3} \tag{15}$$

In the next step, hue of the *r*, *g*, and *b* input channels is decoupled from intensity by normalization with *Int*, whereby normalization is only applied at locations where $Int > 1/10$ of the maximum over the entire image while other location are set to zero. Analogous to the Gaussian pyramid for *Int*, the normalized channels are used to create four Gaussian pyramids for red *R*, green *G*, blue *B*, and yellow *Y*. The four channels are computed as

$$R = \frac{r-(g+b)}{2}, \quad G = \frac{g-(r+b)}{2}, \quad B = \frac{b-(r+g)}{2}, \quad Y = r+g-2\cdot(|r-g|+b) \tag{16}$$

Orientation information is obtained from *Int* using oriented Gabor pyramids, that is, for each $Int(\sigma)$ a Gabor pyramid $Ori(\sigma,\varphi)$ is calculated for orientations $\varphi \in \{0°,45°,90°,135°\}$.

Next, the pyramids for intensity, color, and orientation are combined across scales into a set of feature maps for each feature. Across scale combination simulates the visual receptive fields and renders the system sensitive to local spatial contrast. Center surround operations are implemented as difference between a fine and a coarse scale for a given feature. The feature maps for *Int* are obtained as

$$Int(c,s) = |Int(c) \ominus Int(s)| \tag{17}$$

where *c* denotes the center, *s* the surround, and ⊖ the across-scale subtraction. Similarly, center-surround differences across the normalized color channels are computed as

$$RG(c,s) = |(R(c)-G(c)) \ominus (G(s)-R(s))| \tag{18}$$

$$BY(c,s) = |(B(c)-Y(c)) \ominus (Y(s)-B(s))| \tag{19}$$

and finally, the orientation feature maps are obtained as

$$Ori(c,s,\varphi) = |Ori(c,\varphi) \ominus Ori(s,\varphi)| \tag{20}$$

The last step in the creation of the conspicuity maps for intensity, color, and orientation consists of summing up the feature maps across scales. Across scale summation consists of normalization of the feature maps, the reduction of each map to a common scale and pixel-by-pixel addition

$$\overline{Int} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N\left(Int(c,s)\right) \tag{21}$$

$$\overline{C} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \left[N\left(RG(c,s)\right) + N\left(BY(c,s)\right)\right] \tag{22}$$

$$\overline{Ori} = \sum_{\varphi \in \{0°,45°,90°,135°\}} N\left(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4}\right) N\left(Ori(c,\sigma,\varphi)\right) \tag{23}$$

where c denotes the center, s the surround, $\oplus$ the across-scale summation operation, and N the normalization operation. These three conspicuity maps establish the base for extracting object-based measures for color, intensity, and orientation for attracting attention

Extraction of Object-based Color, Intensity, and Orientation

In order to derive an object-specific value for color, intensity, and orientation, we combine the object's geometry with the conspicuity maps for each feature. Specifically, we use the area-normalized sum of pixel values inside the polygon that defines the spatial object to compare the potential of attracting attention of each spatial object. The motivation for the area-normalized calculation of these feature values if that we need to detach the values from the object's size in order to have independent measures of the object's features. Formally, intensity contrast *IC* is defined as

$$IC_{o_i} = \frac{\sum\limits_{\forall x_{j,k} \in shape_{so_i}} \overline{Int}\left(x_{j,k}\right)}{\mathbf{area}\left(shape_{so_i}\right)} \tag{24}$$

where o_i is the observation to the spatial object and $x_{j,k}$ the pixel in row j and column k. Likewise, we define color contrast *CC* as

$$CC_{o_i} = \frac{\sum\limits_{\forall x_{j,k} \in shape_{so_i}} \overline{C}\left(x_{j,k}\right)}{\mathbf{area}\left(shape_{so_i}\right)} \tag{25}$$

and orientation contrast *OC* as

$$OC_{o_i} = \frac{\sum\limits_{\forall x_{j,k} \in shape_{so_i}} \overline{O}\left(x_{j,k}\right)}{\textbf{area}\left(shape_{so_i}\right)} \qquad (26)$$

Quantification of Scene Configuration

We quantify the scene configuration in terms of measures derived from topological and metric properties of the layout of spatial object in the visual field. These measures capture the role of objects in the configuration of the scene, which is an important indicator for the salience of objects in a global context.

Topological Relations

Within the scope of our computational model, we use the term topological relations to describe the spatial relationships between shapes of spatial objects in the scene. Specifically, we consider two topological relationships, i.e., 1) adjacent, for neighboring shapes of spatial objects, 2) and disjoint for pairs of shapes that are not direct neighbors. We assume that spatial objects that play an important role in the structure of the environment, such as districts or important nodes, have a high number of neighbors, while objects that do not exhibit as strong a role are linked to fewer other objects. Therefore, we define the *Degree of Connectedness DoC,* which expresses the topological connectedness of two spatial objects, as

$$DoC_{o_i} = \sum_{j=1}^{k} rel\left(so_i, so_j\right) \qquad (27)$$

where so_i and so_j identify the observed object and the object it is compared to, k is the total number of spatial relations in S, and $rel \in REL$. The quality of the spatial relation is expressed as

$$rel\left(so_i, so_j\right) \in \left\{disjoin, adjacent\right\} = \begin{cases} 0 & \text{if } disjoint \\ 1 & \text{if } adjacent \end{cases} \qquad (28)$$

The reason for the use of disjoint and adjacent as topological relations between spatial objects is due to practical reasons, rather than conceptual limitations. Future work may extend the current choice of spatial relations in order to account for the spatial scene's depth.

Metric Refinement

In Gestalt Theory, the first law of organization states that elements tend to be grouped together according to their nearness (Wertheimer 1923). Therefore, we assess the proximity of spatial objects in the scene as a metric refinement of the spatial layout, that is, we derive a measure for the spatial distribution of spatial objects form the layout of the scene. Specifically, we use the centroids of the object's shape to compute the distances between objects and to derive two measures for uniqueness of the spatial location of the object within the scene. The Uniqueness of location *UoL* is assessed for the horizontal and vertical dimensions of the visual field and is defined as

$$UoL_{x,o_i} = \sum_{j=1}^{k} \left| \left(CENT_{so_i}(x) - CENT_{so_j}(x) \right) \right| \tag{29}$$

$$UoL_{y,o_i} = \sum_{j=1}^{k} \left| \left(CENT_{so_i}(y) - CENT_{so_j}(y) \right) \right| \tag{30}$$

where *x* and *y* denote the dimension (i.e., horizontal or vertical), o_i identifies the observed object, and *k* is the total number of objects in the scene. These two indicators express the density of spatial objects along two axes.

Degree of Concept Singularity

Another pre-attentive feature is the number of identical objects within a scene. Research has shown that only a small number of codes or concepts can be rapidly perceived, and that the processing time increases with the number of codes. Although concept recognition requires prior knowledge, and hence, is not a purely pre-attentive task (Lakoff 1987), we will use the number of instantiated objects of same class to approximate this phenomenon. The assumption is that the fewer the number of class instances of a specific class within a scene, the higher the probability that these instances will be salient. Formally, we define the Degree of Concept Singularity as

$$DoCS_{o_i} = \frac{\sum \forall so_j \in S \wedge c(so_i) = c(so_j)}{n} \tag{31}$$

where *n* is the total number of objects in the scene, *i* denotes the current object, and *j* the objects to compare with. Note that the Degree of Concept Singularity will be stored in long-term memory and hence, will be available as further refinement of the object's cognitive attributes.

Influence of Modality

The modality of travel (e.g., walking, driving, or riding) influences both, the cognitive load put on the observer, as well as the degree of physical freedom. The allocation of the remaining resources defines the field of view and the center of attention, which in turn determine the prominence of surrounding geographic features. Modality, even though not a pre-attentive property in the classic sense, acts as a filter applied to the whole scene. Therefore, we consider the influence of modality on the salience of spatial objects within the stage of pre-attentive processing.

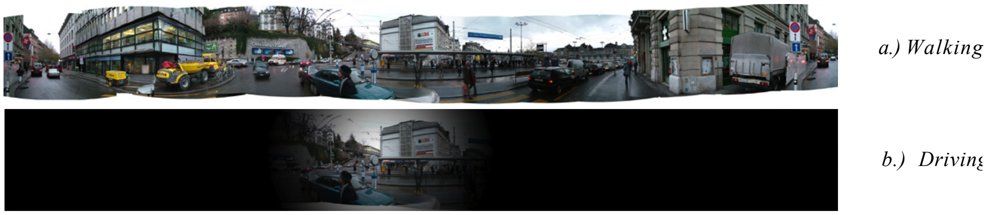


Figure 20 Two fields of view resulting from different modalities: a) Walking requires only few physical and cognitive resources, which results in a field of view that includes the whole scene, while b) driving requires the navigator to focus attention, which results in a narrow field of view.

We model the influence of modality as a function of the field of view for the current modality, and the eccentricity of spatial objects with respect to the center of the field of view. For the purpose of this work, we will focus on walking and driving as modalities. We assume that walking results in a 360° field of view and homogenous allocation of attention across the whole field (Figure 20, *a*). For driving, however, we will assume a field of view of 180° and use the current direction of travel (i.e., next path segment) as center of attention and assume a decline of attention with increasing eccentricity (Crundall, Underwood et al. 1999) (Figure 20, *b*).

Following a route implies a shift of the field of view if the route dictates a turn. In order to model this peculiarity and to account for both, the field of view for incoming and outgoing direction, we use a combined approach. That is, we create two raster images, one based on the incoming direction, and one for the outgoing direction and sum them up. Using this approach, a raster image that represent the Field of View FoV for every modality is created, which is used to assess the influence of modality. Formally, we define the influence of modality IoM as:

$$IoM_{o_i} = \frac{\sum_{\forall x_{j,k} \in shape_{so_i}} FoV_M\left(x_{j,k}\right)}{\mathbf{area}\left(shape_{so_i}\right)} \tag{32}$$

where $x_{i,j}$ is a pixel of the raster image representing the field of view and $shape_{so}$ is the outline of the current spatial object.

Results of pre-attentive Processing

In summary, pre-attentive processing of input data using the set of pre-attentive processing functions *P* combines information contained in the raster image *I*, the set of spatial objects, and the set of binary spatial relations into a vector

$$pop = \left(L, W, EL, A, IC, CC, OC, DoC, UoL_x, UoL_y, DoCS, IoM\right) \tag{33}$$

that holds real-valued measures for the following list of pre-attentive features:

- *L, W, A, EL:* Geometric properties (i.e., length, width, elongation, and size),
- *IC, CC, OC*: Non-geometric visual properties (i.e., color, intensity, orientation),
- *DoC*, *UoL*$_x$, *UoL*$_y$, *DoCS:* Scene-related properties (i.e., degree of connectedness, horizontal and vertical uniqueness of location, concept singularity)
- *IoM*: Influence of modality,

in addition to the set of thematic object descriptors *od*.

3.2.1.2 Attentive Processing

In the attentive processing step we compute the features that are required for the similarity assessment of scene objects. Attentive processing consists of a series of stages, namely the refinement of shape parameters, the quantification of cognitive and contextual components, and the update of long-term memory. The following sections describe the functions we use in the process.

Long-term Memory

Attentive processing involves prior experience and knowledge, which is stored in long-term memory. In our computational model, we define long-term memory as a route map, which is basically a collection of observations to spatial objects along the route. Formally, the route map *RM* is defined as

$$RM = \{DP, O, MSO\} \tag{34}$$

where *DP* denotes the ordered (i.e., sequence of visit) list of *Decision Points* (or scenes) traversed during traveling, $o_i \in O = \left(od_i, pop_i\right)$ is the set of observations to spatial objects, and *MSO* is the set of all observed spatial objects along the route. Note that the set of observations to spatial objects in long-term memory is equivalent to the set of pre-attentively processed observations for all visited scenes. The set of observations in memory, however, is not the same as the subset of observations used for the assessment process, as we first need to quantify cognitive and contextual properties before assessing the salience of spatial objects.

Set of Attentive Features

The set of attentive feature parameters **A** that is extracted from the perceptual scene representation is defined as

$$\mathrm{A} = \{\alpha_i\}_{i=1}^{n} \qquad (35)$$

where *n* is the number of attentive features. The set of attentive features includes refinements of the pre-attentive features, as well as cognitive and contextual features. The following list shows the set of attentive features **A** used in this model:

- Object Recognition
- Idiosyncratic Relevance
- Task-based Context

Object recognition, and idiosyncratic relevance are cognitive properties and involve prior experience and knowledge (i.e., observations stored in long-term memory), and finally, the task of navigation as contextual components completes the set of attentive features.

Extraction of Properties for Similarity Assessment

Set of Attentive Processing Functions

Given a Perceptual Scene Representation *S*', the Route Map *RM* from long-term memory, the set of attentive processing parameters **A**, and the next route segment ψ, a set of attentive processing functions *F* is defined as

$$F : S' \to O' \qquad (36)$$

which extract a set of object-specific features *O*' from the perceptual scene representation *S*'. Specifically, attentive processing extracts a real-valued feature vector $aop \in O'$ that characterizes the attentively processed observations of spatial objects and enhances the observations from the perceptual scene representation *S*' with information from long-term memory and for the relevance of spatial objects for the current task. Formally, we define the resulting set of observations *O*' is as

$$O' = \{pop_i, aop_i\}_{i=1}^{n} \tag{37}$$

where *pop* denotes the subset of pre-attentive object properties, *aop* the subset of attentive object properties, and *n* the number of objects in the scene. In the following sections, we define the functions used for attentive processing.

Quantification of Cognitive Properties

The quantification of the cognitive properties involves prior knowledge and experience, and assesses the degree of recognition, which is subdivided in measures of concept recognition and object recognition, and the idiosyncratic importance of objects. Long-term memory is modeled as a route map *RM* consisting of decision points and observation to spatial objects from these decision points, and conveys information beyond the currently visible part of the environment. Prior knowledge is stored in long-term memory in terms of previous observations to spatial objects, while experience is implicitly modeled as a function of the frequency of observations to spatial objects.

Degree of Concept Recognition

The degree of concept recognition assesses the degree to which an instance of a spatial object can be assigned to a class. We derive the degree of concept recognition from the ratio of the number of observed objects features and the total number of object features. Formally, the degree of concept recognition *DoCR* for spatial object o_i is defined as

$$DoCR_{o_i} = \frac{\text{Nr of Observed Object Properties}}{\text{Total Nr of Object Properties}} \tag{38}$$

Degree of Object Recognition

The degree of object recognition estimates to what extent the current observation (i.e., *aop* and *pop*) can be assigned to an observation in long-term memory, and hence, provides a measure how easy the corresponding spatial object can be identified. Formally, we define the Degree of Recognition *DoOR as*

$$DoOR_{o_i} = \begin{cases} 0 & \text{if } mso_i \notin MSO \\ \mathbf{max}(sim(o_i, O_i)) & \text{if } mso_i \in MSO \end{cases} \tag{39}$$

where *DoOR* is a real number, o_i denotes the current observations, MO_i the set of memorized observations for the corresponding mental spatial object $mso_{i,}$, and *MSO* the full set of Mental Spatial Objects in long-term memory.

The similarity between observed object and memorized object is defined by the maximum degree of similarity between the current observation and the memorized observations. We use the Levenshtein distance, which is widely used in information theory, for assessing the similarity between the two observations that are of type string (Levenshtein 1966). The Levenshtein distance between two strings of unequal length is the number of positions for which the corresponding symbols are different, that is, it measures the number of substitutions required to change one string into the other. Note that we do not assess the similarity for string attributes that are not observed (i.e., those attributes that are set as *not applicable* in the class instance). For assessing the similarity of attributes of type real, in contrast, we will use the 1-norm Minkowski distance (Abdi 2007). Formally, we define the similarity function as

$$sim(o_i, mso_i) = \sum_{k=1}^{n} sim(o_i(f_k), mo_i(f_k)) \tag{40}$$

whereby

$$sim(o_i(f_k), mo_i(f_k)) = \begin{cases} \mathbf{levenshtein}(o_i(f_k), mo_i(f_k)) & \text{if } f \in String \\ |o_i(f_k) - mo_i(f_k)| & \text{if } f \in Real \end{cases} \tag{41}$$

and where f_k denotes the feature being compared, o_i the current observation, and mo_i the memorized observations to the same object as observation o_i.

Idiosyncratic Relevance

The level of idiosyncratic relevance depends on the amount of knowledge and personal experience navigators associate with a specific number of observations. In our model, this information is implicitly modeled in long-term memory as the number of observations associated with specific spatial objects. We assume that a large number of observations to the same spatial object increase its idiosyncratic relevance, while a low number of observations equals low idiosyncratic relevance. Formally, we define the level of idiosyncratic relevance *IR* as

$$IR_{oi} = \begin{cases} 0 & \text{if } mso_i \notin MSO \\ \mathbf{count}(mo_i) & \text{if } mso_i \in MSO \end{cases} \tag{42}$$

where *o* denotes the current observation of a spatial object *i*, *mo* the memorized observations of object *i*, mso_i the corresponding mental spatial object, and *MSO* full set of mental spatial objects.

Quantification of Task-based Relevance

Navigation is the combined activity of locomotion and wayfinding, whereby wayfinding is understood as reasoning about the continuation of the journey. Navigators often rely on spatial objects in the environment for spatial reasoning and especially for identification of the correct path (Golledge 1999; Montello 2003). Therefore, we assume that objects located close to the path are of higher value than objects located further away. This approach is, due to the limitations of the input data, a crude abstraction of the structural salience of landmark for route directions, as proposed by Klippel (2005). We base our quantification of task-based influence on this basic assumption. Formally, task-based influence *TbI* is defined as

$$TbI_{o_i} = \mathbf{dist}\left(CENT(x)_{o_i}, CENT(x)_p\right) \tag{43}$$

where *CENT* denotes the centroid and *p* the designated target path. Note that with this approach, task-based influence is only assessed if the current scene is not the last scene, because in the last scene, the destination is reached, and hence, *p* is no longer available.

Long-term Memory Update

The last step in attentive processing is the update of the route map in long-term memory. Updating the route map simulates the learning process during navigation and is imperative for assessing the cognitive components of saliency. In our computational model, we will use the following algorithm for updating the route map:

```
Start Update Route Map
    Add current Scene to List of Decision Points in RM
    FOR each Observation in Scene
        Add Observation to List of Observations in RM
        IF NOT SO(Observation) exists in List of MSO
            Add new MSO to List of MSO in RM
        END IF
    End FOR
End Update LTM
```

The update function is executed after attentive processing, but before the saliency assessment. Also note that because attentive properties are a function of long-term memory, the route map is updated with the set of pre-attentive observation of spatial objects, rather than with the refined set of attentively processed observations. Formally, the update function is defined as

$$u : S' \rightarrow RM \tag{44}$$

$$RM(t+1) = \forall o \in S' \begin{cases} \mathbf{add}\left(dp_i, o_i, o_i\left(MSO_{o_i}\right)\right) & o_i\left(MSO_{o_i}\right) \notin RM(t) \\ \mathbf{add}\left(dp_i, o_i\right) & o_i\left(MSO_{o_i}\right) \in RM(t) \end{cases} \tag{45}$$

where *u* denotes the update function, *S*' the perceptual scene representation, and *RM* the route map in long-term memory.

Result of Attentive Processing

In summary, attentive processing of the perceptual scene representation *S*' using the set of attentive processing functions *F* ensures that all properties for the assessment process are extracted from input data and long-term memory. The result is a vector of attentive properties

$$aop = \left(DoCR, DoOR, IR, TbI\right) \tag{46}$$

that holds real-valued measures for the following list of attentive features:

- *DoCR, DoOR, IR*: Cognitive properties (i.e., concept and object recognition, and idiosyncratic relevance), and
- *TbI:* Task-based influence.

Attentive processing completes the extraction of object properties, as the set of pre-attentive and attentive object properties for the assessment process is now complete.

3.2.2 Saliency Assessment Process

Lynch (1960) describes the key components of landmarks using the attributes of individual importance, singularity, and uniqueness. These properties are closely related to the notion of salience, which denotes relatively distinct, prominent or obvious features compared to other features. The quality of standing out relative to neighboring items implicitly includes an assessment of dissimilarities (or similarities) between objects. Consequently, we treat salience as a function of the individual dissimilarity of spatial objects, rather than as a measure of their absolute values.

In our computational model, we implement the assessment process as a probabilistic model that calculates the probability of spatial objects to be salient in some aspect (i.e., perceptually, cognitively, or contextually) as a function of their individual differences. The quantification of the scene content provides all information required for assessing how dissimilar the spatial objects contained therein are. The scene quantification is used as input for the assessment process, which consists of three separate steps: 1) the dissimilarity assessment, 2) the computation of the probability distributions based on individual dissimilarities, and 3) the saliency assessment process, performed as probabilistic inference in a Bayesian network given the probability distributions. The following sections provide a detailed description of the steps involved in the assessment process.

3.2.2.1 Dissimilarity Assessment

The dissimilarity assessment transforms the set of observations $O' = (\{pop\}, \{aop\})$ into a dissimilarity matrix **D** that builds the base for the assessment of saliency. Formally, given a pair of observation $o_1, o_2 \in O'$ a feature dissimilarity function *d* is defined as

$$d : O' \to \mathbf{D} \tag{47}$$

that extracts a vector of real-valued dissimilarity measures. This vector corresponds to a low-level component, as described in the framework for the saliency assessment (Chapter 2), which implies that the dissimilarity function includes a reduction of the set of properties that defines an observation of a spatial object to the set of low-level components. This reduction is motivated by the assumption that relative differences in the low-level components epitomize the salience of spatial objects. In our model, we achieve this reduction by treating the low-level components as vectors consisting of observed object properties and by using the Euclidean distance between pairs of vectors as dissimilarity function. The following list summarizes the low-level vectors and their contributing object properties:

- Size (*L, W,A*)
- Shape (*EL,CP*)
- Color Contrast (*CC*)
- Intensity Contrast (*IC*)
- Orientation Contrast (*OC*)
- Topology (*DoC*)
- Metric Refinements (UoL_x, UoL_y, NoI)
- Degree of Recognition (*DoCR, DoOR*)
- Idiosyncrasy Relevance (*IR*)
- Task-based Context (*TbI*)
- Modality (*IoM*)

Formally, we define the dissimilarity δ between spatial objects *i* and *j* as

$$\delta_{ij} = \sqrt{\sum_{k=1}^{n} \left(x_{ik} - x_{jk} \right)^2} \tag{48}$$

where *i* and *j* denote the two objects that are being compared, and *k* the features that define the corresponding low-level component. The dissimilarity matrix **D** for one low-level component is defined as

$$\mathbf{D}_i := \left(\mathbf{d}_{k,l}\right)_{n\times n} \tag{49}$$

where *i* denotes the low-level components (size, shape, etc.), *k* and *l* the entries, and *n* the number of objects in the spatial scene. The matrix describes pairwise distinctions between the *n* objects in the scene. It is a square symmetrical *n*×*n* matrix with the *kl*-th element equal to the value calculated by the dissimilarity function between the *k*-th and the *l*-th spatial object. The diagonal elements are equal to zero, that is, the distinction between an object and itself is postulated as zero. Since all object properties are of type real, the same function is applied to assess the dissimilarity of all low-level components, resulting in a set of dissimilarity matrices, one for every low-level component.

From the dissimilarity matrices, we define the total dissimilarity Δ of spatial objects as

$$\Delta_{oi} = \sum_{j=1}^{n}\left(\delta_{ij}\right) \tag{50}$$

where o_i denotes the spatial object *o* in column *i*, *j* the row, and *n* the total number of spatial objects. The total dissimilarity expresses how dissimilar the spatial object is with respect to all other objects in the scene.

3.2.2.2 Specification of the Bayesian Network

We propose to use a Bayesian network as probabilistic model to assess the salience of landmarks. A Bayesian network is a directed acyclic graph (DAG) in which nodes represent random variables, and the absence of arcs represents conditional independence in the following formal sense: A node is independent of its non-descendants given its parents. Informally, we can think of a node as being “caused” by its parents. For instance, location-based attention is directly influenced by color and intensity. Accordingly, in the DAG, the arc that connects the nodes that represent color and intensity indicates that location-based attention is caused by these two factors. In the following sections, we describe and specify the Bayesian network insofar as is required to understand the mechanism that is applied for the assessment process. For a detailed discussion of Bayesian network we refer the interested reader to the vast literature on the topic.

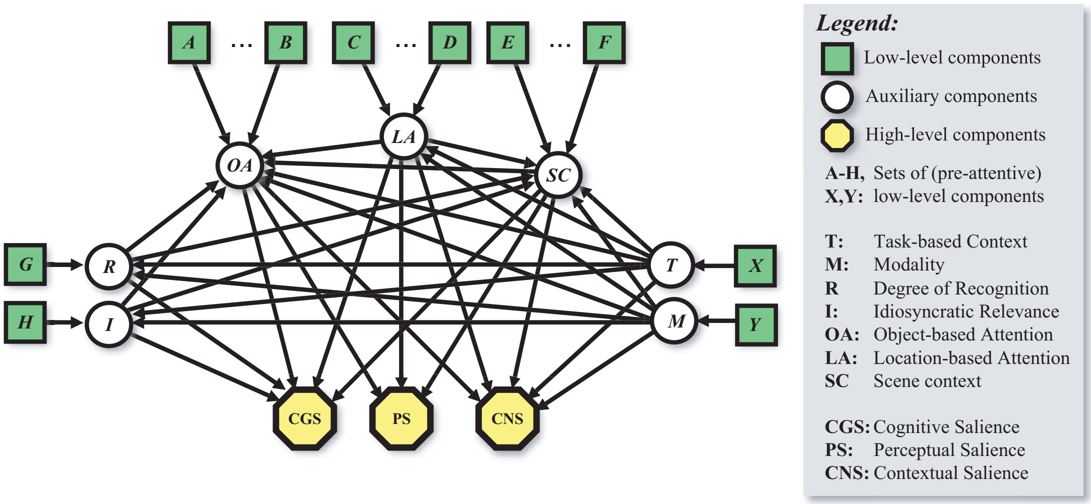


Figure 21 The Bayesian network used in our computational model. The network is defined by a set of nodes that represent the components of salience and their dependencies. The set of root nodes represents the low-level components (i.e., root nodes) and the non-root components the auxiliary and high-level components of salience.

Formally, a Bayesian network is defined as a DAG, *D=(V,E)*, where *V* is a finite set of nodes and *E* is a finite set of directed edges (i.e., arrows) between the nodes. The DAG defines the structure of the Bayesian network (cf. Figure 21), whereby each node $v \in V$ in the graph corresponds to a random variable X_v and the set of variables associated with the graph *D* is defined as $X = (X_v)_{v \in V}$. Every node *v* with parents pa(*v*) is described with a local probability distribution, $p(x_v \mid x_{\mathrm{pa}(v)})$. The set of local probability distributions for all variables (i.e., components) in the network is denoted by *PD*. The Bayesian network *BN* for a set of random variables *X* is then the pair (*D*, *PD*). Note that the possible lack of directed edges in *D* encodes conditional independencies between the random variables *X*. In our computational model, we will use the Bayesian network proposed in our conceptual framework for the assessment process.

3.2.2.3 Specification of Prior and Conditional Probability Distributions

Given the Bayesian network graph *D* and the local probability distributions *PD*, we can factor the joint distribution over all the variables into the product of local terms:

$$P\left(X_1,\ldots,X_m\right) = \prod_i P\left(X_i \middle| pa\left(X_i\right)\right) \qquad (51)$$

where $pa(X_i)$ are the parents of node X_i, and $P(X_i \mid pa(X_i))$ is the conditional distribution of X_i. In order to fully specify the Bayesian network, we need to define the probability density functions for each node, in addition to the structure of the Bayesian network proposed in our conceptual framework and the joint probability function. Specifically, we need to define the prior probabilities for the root nodes and the conditional probabilities of all non-root nodes.

For our computational model, we assume that all nodes are discrete, that is, each component is associated with a hypothesis, which is evaluated as either true or false. For instance, the node representing color is attached to the hypothesis that the color is salient or not salient, depending on the properties of the spatial object. Similarly, the node representing object-based attention is attached to the hypothesis that the spatial object attracts object-based attention given its parents, which may be instantiated as either true or false.

Probability Density Function for Root Nodes

We defined the low-level components to be the nodes for which we collect data, that is, the evidence nodes. This evidence is present in terms of dissimilarity matrices that represent differences between pairs of objects in the scene. In order to use the evidence in the Bayesian network, we need to convert the dissimilarities matrices into probabilities for each node. This is done by means of a discrete probability density function, which may be considered as a smoothed-out version of a histogram representing the total differences among the spatial objects. Formally, we define the probability density function P as

$$P\left(\Delta_{o_i f_j} = true\right) = \frac{\Delta_{o_i f_j}}{\sum_{k=1}^{n} \Delta_{o_k f_j}} \qquad (52)$$

where f denotes the low-level components, o_i the observed spatial object, and n the total number of objects in the scene. Semantically, the value calculated by the function indicates the probability that the total difference of a specific feature of the spatial object is salient with respect to the same feature of the rest of the spatial objects in the scene.

The set of resulting probabilities $P_{f_j} = \left\{P\left(\Delta_{o_i f_j}\right)\right\}$ for a certain feature defines a discrete probability distribution. These distributions are unrestricted discrete distributions and the parameters fulfill $\sum_{o_i \in O} P\left(\Delta_{o_i f_j}\right) = 1$ and $0 \leq P\left(\Delta_{o_i f_j}\right) \leq 1$. Using this approach, we calculate probability values for the n observed spatial objects and m low-level components of each object, and collect the results in terms of a probability matrix. The probability matrix $\mathbf{P}$ is defined as

$$\mathbf{P} := \left(p_{i,j}\right)_{n \times m} \qquad (53)$$

where $p_{i,j}$ is the probability to which feature *j* of spatial object *i* is salient. The probability matrix is used to add evidence to the Bayesian network during the assessment process.

Probability Density Function for Non-root Nodes

Each node X has a conditional probability distribution $p(x_v \mid x_{\mathrm{pa}(v)})$ that quantifies the effect of the parent's nodes on the node. We base the definition of the conditional probabilities on the following assumptions: 1) A node's parents fully explain the outcome, (i.e., given the parents are salient, the probabilities of the node sum up to 1), which implies that we assume that there are no other causes influencing the node's posterior probability, and 2) due to lack of evidence and uncertainty about the individual contribution of the parents, we have no reason to expect or prefer one or the other. Therefore, we apply the Principle of indifference and assume that probability density functions for auxiliary and high-level nodes are even distributions. Formally, we define the local conditional probability *CP* for auxiliary and high-level components as

$$CP\left(x_v \middle| x_{pa(v)}\right) = \frac{1}{n} \sum_{i \in pa(v)} p(x_i) \tag{54}$$

where *x* denotes the node, *n* the total number of parents of node *x*, and *pa*(*v*) the set of parents of *x*.

3.2.2.4 Probabilistic Inference Algorithm

The saliency assessment of spatial objects is performed as Probabilistic Inference in the Bayesian network. In our case, probabilistic inference means computing the posterior probability distribution for a set of query variables (i.e., high-level components), given a set of evidence variables (i.e., low-level components) and the conditional independencies encoded in the structure of the Bayesian network. The probabilistic inference algorithm process is executed as follows:

```
Start Probabilistic Inference
     FOR each probability vector in P
           Update prior probabilities for object i
           Perform belief update
           Extract posterior beliefs for cgs, ps, and cns
           Update list of saliency vectors
     End FOR
End Probabilistic Inference
```

We define the probabilistic inference step as a saliency assessment function, which predicts the probability of the spatial object to be salient. Formally, given the probability matrix **P** for the low-level components of the spatial objects and the Bayesian network *BN* with the conditional probabilities *CP*, we define the saliency assessment function *a* as

$$a : \mathbf{P} \rightarrow V \qquad (55)$$

where **P** denotes the probability matrix containing the probabilities for the low-level components, and *V* is the set of saliency vectors. A Saliency Vector $\mathbf{v} \in V$ is defined as

$$\mathbf{v} = \left(ps_i, cgs_i, cs_i \right)_{i=1}^{m} \qquad (56)$$

where *ps* denotes the Perceptual Salience, *cgs* the Cognitive Salience, *cs* the Contextual Salience, and *m* is the total number of saliency vectors.

Note, that in the assessment process, we do not instantiate any nodes as evidence, but rather update the degree of belief in a hypothesis about the current spatial object with the probabilities calculated for the low-level components. The prior beliefs are updated for every spatial object before updating the Bayesian network and collecting the posterior beliefs. The reason for this approach is that we want to rank the spatial objects according to their probability of being salient, which would not be possible if the results are partitioned in a set of salient and a set of non-salient objects. From a computational point of view, however, the two approaches are equivalent, since the prior probabilities are only adjusted for root nodes, and hence, conditional dependencies are preserved. The final result of the salience assessment process is a vector **v** for every spatial object that indicates the probability of the object to be perceptually, cognitively, or contextually salient.

3.3 Summary

In this chapter, we formalized the conceptual framework defined in the previous chapter in terms of a computational model. We first described the computational strategy used for the integrated assessment of landmark salience. The computational strategy considers aspects related to the environment and the specification of the representation of the environment, along with computational strategies related to human information processing and memory. In the second part, we described and formally defined the computational model in terms of extraction of required information, as well as the saliency assessment process. The computational model is required in order to validate the conceptual framework, which is the topic of the following chapter.

Chapter 4
Evaluation

The framework for the assessment of salience and the computational model formed the base for the implementation of a software prototype, or simply prototype. In this chapter, we will find answers to the following two questions: 1) Did we implement the computational model right, and 2) Did we implement the right computational model? In a software development context, the first question is typically referred to as verification and the latter as validation. Both, verification and validation are review processes, whereby verification evaluates whether the prototype conforms to the specifications, while validation assesses the performance with respect to typical usage. Performance with respect to typical usage, in this context, refers to the performance of the computational model with respect to real-world scenarios, such as, in our case, the assessment of the saliency of spatial objects by human observers.

The remainder of this chapter is organized as follows: Section 4.1 gives a short description of the prototype implementation. Section 4.2 provides a detailed account of the verification strategy, including the method of verification, the set of experiments, and the presentation of the results. In section 4.3, the validation strategy is introduced and the degree of correlation between results generated by the computational model and human subjects are analyzed. Finally, section 4.4 concludes the chapter with a summary of the evaluation.

4.1 Prototype Implementation

The prototype is implemented as a platform for simulation and will serve as test-bed for the evaluation. It was implemented in the JAVA programming language and is comprised of two core modules. The first module implements the model of the environment, and the second module imitates the observer, which we will henceforth refer to as agent. Note that in the context of this work, we define simulation as a software program that imitates the salience assessment process by causing the agent to respond mathematically to changing conditions as though it were the process itself. The basic idea, hence, is that the agent moves from scene to scene, assesses the saliency of the objects in the current scene, and generates a ranking of them. Moving to the next scene implies an update of the list of observed objects by dismissing those that are not observed again, adjusting the salience of objects that are observed again, and assessing objects that are now visible.

The agent is implemented such that it can be initialized with different settings. These settings are defined as a list of arguments and include parameters that define the context and the cognitive abilities of the agent. The context is given by the task in terms of a route plan, and by the modality, which is either walking or driving. The route plan is composed of a pre-defined sequence of scenes and dictates what parts of the environment the agent will visit during the simulation. Changing the modality directly affects salience as walking assumes a field of view of 360°, while setting the modality to driving reduces it to 180°. In order to simulate the influence of memory on the salience of objects, we need to be able to control the cognitive abilities of the agent. We do this by enabling or disabling the agent to store observations in Long Term Memory (LTM), which directly influences the degree of recognition and idiosyncrasy.

There are many scenarios that might be thought of when simulating the assessment of landmark salience for human navigation. The settings for our prototype allow covering a wide range of them, including the assessment of single scenes and whole routes, with or without storing previous observations. Given a scene or a route, we have the choice among the following set of Agent Configurations:

1.) $LTM \wedge Walking$
2.) $LTM \wedge Driving$
3.) $\neg LTM \wedge Walking$
4.) $\neg LTM \wedge Driving$

Note that we used the symbols for logical conjunction ('∧') and logical negation ('¬') to identify a specific configuration. Each configuration affects the resulting saliency differently. In the verification section, we will thoroughly investigate the impact and discuss the properties of the different configurations in detail.

4.2 Verification of Prototype

The goal of the verification process is to assure that the computational model fully satisfies the specifications. We understand the specifications as expected requirements and divide the verification process in two sets of experiments. In the first set, we will verify that the model fulfills general requirements and in the second set of experiments, we will investigate whether the computational model accounts for typical characteristics of landmark saliency.

4.2.1 Method of Verification

The method applied for the verification of the computational model includes the following steps; 1) the description and definition of a test case that allows the verification of a distinct aspect, 2) the formulation of the expected outcome against which the results will be tested, 3) the execution of the experiment and collection of results, and finally, 4) the evaluation of the experiment. For each test case, we define three sets of variables: a set of *independent variables*, which are manipulated during the experiment, a set of *control variables*, which are kept constant throughout the experiment, and a *set of dependent variables*, which we will test to see whether the manipulations on the independent variables had the expected effect. The verification method is applied consecutively in order to verify the impact of the independent variable given by the test configuration on the dependent variables (i.e. perceptual, cognitive, and contextual salience).

4.2.2 General Requirements

In this section, we investigate how individual low-level components contribute to the perceptual, cognitive, and contextual saliency. We will conduct three experiments, which are designed such as to reflect the nature of the low-level components that are being verified. The low-level components are measured either in terms of absolute values (i.e., magnitudes), or in terms of the dissimilarity of an object with respect to the other object. Specifically, *Task*, *Modality*, *Recognition*, *Idiosyncrasy*, *Color*, *Intensity*, and *Orientation Contrast* contribute to saliency with their magnitude, while *Size*, *Shape*, *Topology*, and *Metric* are considered as dissimilarities. Both, magnitude and dissimilarity-based components exhibit certain characteristics, which we will verify by performing a set of experiments on the basis of test cases.

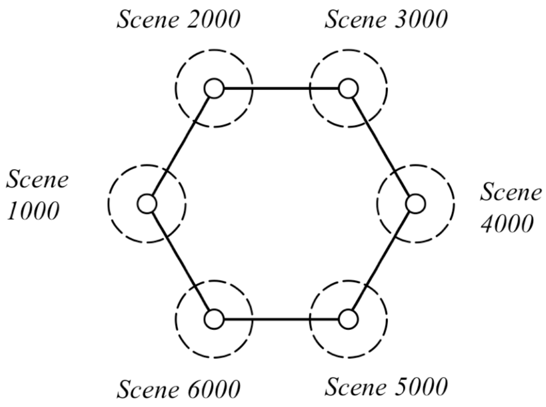


Figure 22 The artificial environment consisting of six spatial scenes that are used to define test casees for the verification of the computational model. The scenes are laid out as a regular hexagon

For the definition of the test cases for general requirements, we will use an artificial environment that consists of six spatial scenes. The spatial scenes are laid out as a regular hexagon (cf. Figure 22), which is implicitly reflected by the interconnections among the scenes. The scenes conform to the data model as defined for the computational model, that is, each scene is defined by a set of objects, the binary relations among the objects, a background image for color, intensity, and orientation contrasts, and the connections to the adjacent scenes. This basic configuration is used to define routes for the test cases.

The test environment provides the means for the definition of test cases, while the implementation of the computational model provides the tool for simulation. In the following sections, we describe the experiments in further detail.

The test cases are defined by the dependent, independent, and control variables. We will use the same route (i.e., from scene 1000 to 6000) for all test cases. Also, for all test cases, the first scene features six spatial objects that are geometrically identical (i.e., rectangles) and that are distributed evenly across the scene. In fact, the only distinction among the six spatial objects of the first scene in terms of low-level components is their class type (i.e., six instances, one instance of each class), and their spatial distance to the direction of connection to the adjacent scenes.

In the following five scenes, however, the independent variables of the low-level components that we wish to investigate are systematically altered, whereby the manipulation is such that the input and expected output is characterized unequivocally. Consequently, the route that defines the test case is defined as a set of systematically altered conditions of the independent variables used to determine the impact of configurations. The results of the simulation using these scenarios provide the evidence for support or rejection of the validity of the expected outcome.

4.2.2.1 Experiment 1: Saliency of Identical Objects

The most general and basic requirement is that objects of equal properties have the same salience. This requirement applies for both, perceptual and cognitive salience, but not for contextual salience. Contextual salience is a function of task and modality, which both emphasize the location of objects, and hence, act as a spatial filter over the presented scene. Therefore, even if objects have the same perceptual and cognitive properties, context filters those objects that are more relevant for the task at hand. Hence, we expect higher values for contextual salience for these objects.

Setup

The data for this experiment consists of six spatial scenes containing six identical objects positioned evenly across the scene. The scenes are linked together as shown in Figure 22. The route is the only independent variable and is defined such that the agent visits each scene exactly once. The choice of walking as mode of traveling reduces the set of influencing factors to a single component, namely task-based influence.

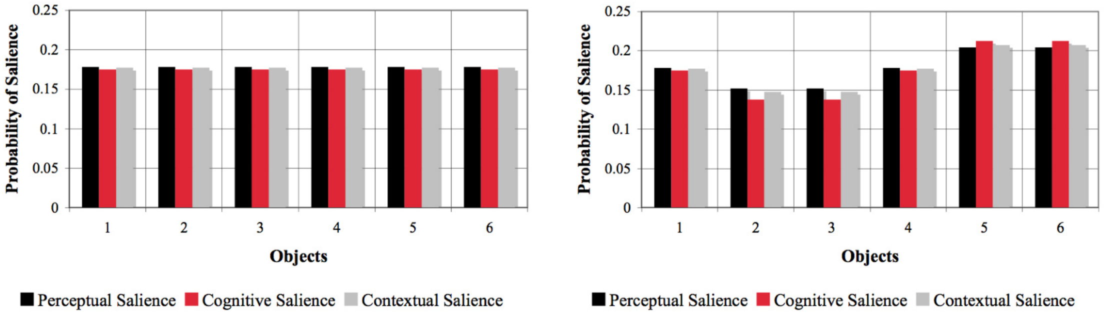
Independent Variable:	*Route* = (1000, 2000, 3000, 4000, 5000, 6000)
Control Variables:	*Agent Configuration* = (*LTM* ∧ *Walking*)
	Magnitude- based variables (equal for all objects)
	Dissimilarity-based variables (equal for all objects)
Observed Variables:	*Perceptual Salience*
	Cognitive Salience
	Contextual Salience

Requirements

- In the absence of contextual influence, all objects must have the same values for perceptual, cognitive, and contextual salience. Given our setup, this should be the case only for the last scene.
- Perceptual and cognitive salience for objects with identical perceptual and cognitive components varies as a linear function of the context. This should be the case for the first five scenes (i.e., 1000 to 5000).

Results

The results show that the requirements are met. For the first five scenes of the route (1000 to 5000), the probability values for perceptual and cognitive salience vary according to the influence of context (i.e., task-based influence). Objects that are located close to the path that leads to the next scene have a higher probability of being salient than objects located further away. The resulting probability for salience for the first five scenes is illustrated in Figure 23 (a). In the last scene, however, task-based influence is absent because the agent arrived at the destination (i.e., scene 6000). Objects are not assessed based on their importance for identification of a target path, and consequently, all object have the same probability of being salient. The probabilities of salience for the last scene are illustrated in Figure 23 (b).



(a) (b)

Figure 23 The diagram shows the salience of identical objects in the presence of context (a), and without contextual influence (b). The magnitudes of the saliency values vary with the number of objects within the scene, and the relative differences between the saliencies are dependent on the structure of the Bayesian network**.**

From the results, we can further see that the probability values for the three types of salience are a function of the number of object contained in the scene. Specifically, given a set of *n* identical objects, the probability that a single object is salient is approximately 1/*n*. Consequently, the more objects in the scene, the lower the probability of saliency. This peculiarity implies that we will be able to rank objects within a single scene, but not across scenes, as the number of objects may vary from scene to scene.

The difference we observe between the probabilities for the different types of salience is a result of the contributing low-level components and the mutual influence of the auxiliary components. For instance, perceptual salience is influenced by object-based attention, location-based attention, and scene configuration, and only indirectly by the contextual and cognitive properties. Cognitive salience, on the other hand, is directly influenced by the same three components, in addition to the degree of recognition and idiosyncrasy influence contextual salience. The same applies for contextual salience.

4.2.2.2 Experiment 2: Influence of Dissimilarities

In this section, we analyze the impact of those low-level components that are derived based on individual dissimilarities. The set of components that are considered with their dissimilarities includes size, shape, topology, and metric of a particular object. Note, that this experiment was conducted for all of the components that are derived from dissimilarities among objects, but for the sake of conciseness is described in detail only for size.

Setup

The data for this experiment consists of a set of six scenes. In the first scene, all objects have identical dimensions (i.e., rectangles). In the second scene, the second object is altered such, that it is larger in size than the other objects (larger rectangle). The third scene is a copy of the second scene, except that the dimensions of the third object are altered the same way as before. This process is applied to all six scenes. In the last scene, therefore, we have a set of objects that have increasing dimensions, whereby the first object is the smallest and the last object the largest.

A peculiarity of this setup is that differences with a positive sign will increase more than differences with negative sign. This is due to the fact that our objects are of rectangular shape. Changing the length and width of the rectangle causes the area to increase to the power of two, and hence, results in a higher probability for size dissimilarity for large objects.

Systematically altering the dimensions of objects across the scenes reduces the number of similar objects in each scene, and at the same time increases the dissimilarities among the objects. The six scenes define the route we will use for the simulation. Note that in order to disambiguate the influence of dissimilarities from other influences, we conducted this experiment without considering cognitive and contextual influences.

Independent Variable: *Route* = (1000, 2000, 3000, 4000, 5000, 6000)
In the first scene, all objects are equal
In the subsequent scenes, we systematically replace one of the objects with a larger object
In the last scene, all objects are different in size

Control Variables: *Agent Configuration* = (¬*LTM* ∧ *Walking*)
(Cognitive and Contextual influences disabled)
Magnitude- based variables (equal for all objects)
All difference-based variables, except *size*

Observed Variables: *Perceptual Salience*

Requirements

- Highly dissimilar objects are more salient than objects with little dissimilarity, and
- For low-level components that are considered with their individual dissimilarities, the probability of being salient increases linearly with increasing difference.

Results

The results of the experiment show that the initial probability of salience is the same for all objects, while in the last scene, where all objects are dissimilar from each other, the objects with the largest differences have the highest probabilities of being salient (Figure 24). This conforms to our expectations.

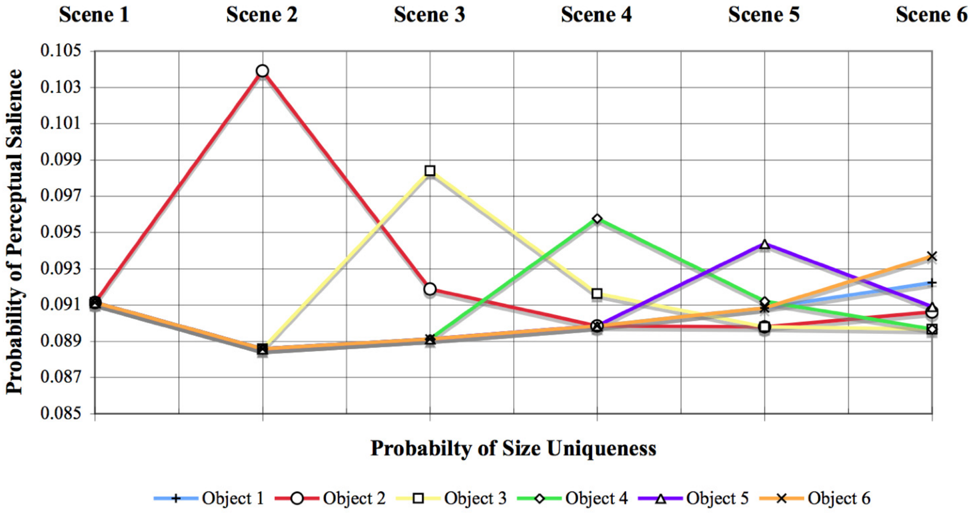


Figure 24 The figure show the change of perceptual salience in the single scenes. Adding objects that are larger than the previous objects influences the probability of the other objects such that large differences result in higher probability values.

In Figure 24 we can also see how replacing an object by a larger object affects the probability values of the objects in the scene. Specifically, we can see how the probability varies with the number of dissimilar objects in combination with the dissimilarities among objects. In the second scene, for instance, only one object is dissimilar, which results in a very high probability for this object of being salient, while in the last scene, where all objects are different, the probabilities are evenly distributed among the objects.

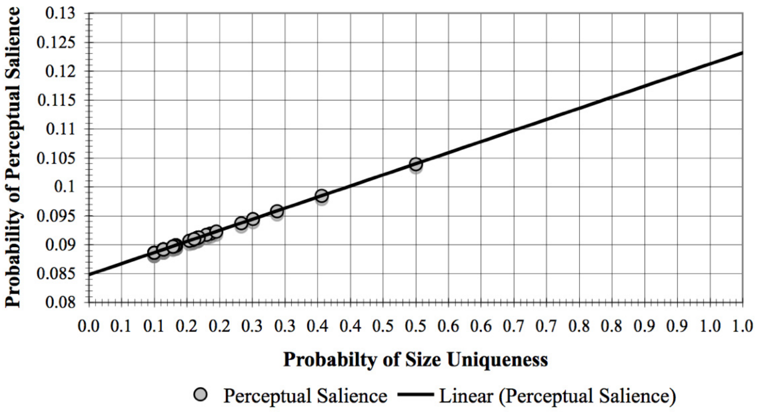


Figure 25 In this figure, we plotted the differences in shape against the resulting probabilities of perceptual salience. The trend line illustrates the positive linear correlation between the two components.

To further examine the influence of dissimilarities on the resulting probability values, we plotted the differences in size against the resulting probability values for perceptual salience (Figure 25). The trend line illustrates the positive linear correlation (correlation coefficient ρ=1) between differences and probabilities.

4.2.2.3 Experiment 3: Influence of Magnitudes

The second type of low-level components comprises those that influence the resulting salience with their absolute value, or magnitude. This set of components includes contrast for color, intensity, and orientation, degree of recognition, and idiosyncrasy. In contrast to the dissimilarity-based components, these components affect the resulting probabilities of salience as a function of their magnitude. Note, that this experiment was conducted for all components that are considered with their magnitude, but for the sake of conciseness is described in detail only for color contrast.

Setup

The setup for testing the influence of components that contribute with their magnitude consists of the route defined by the test environment (i.e., scene 1000 to 6000). In the first scene, all objects have the same properties. In the subsequent scenes, we keep the difference-based components equal for all objects and systematically alter one component that contributes with its absolute value (i.e., color contrast). We alter the component such that it is the highest of the objects in the current scene. Consequently, in the last scene the six objects have increasing values for color contrast, while all other components are equal. The setup is summarized below:

Independent Variable: *Route* = (1000, 2000, 3000, 4000, 5000, 6000)
In the first scene, all objects are equal
In the subsequent scenes, we systematically alter the color contrast of one of the objects in the scene
In the last scene, all objects have a different value for color contrast.

Control Variables: *Agent Configuration* = (¬*LTM* ∧ *Walking*)
(Cognitive and Contextual influences disabled)
Difference- based variables (equal for all objects)
All magnitude-based variables, except *color contrast*

Observed Variables: *Perceptual Salience*

Requirements

- Those components that influence saliency with their magnitude yield higher probabilities of perceptual salience, and
- For, we expect a positive linear correlation between magnitude and resulting saliencies.

Results

The results show that for components that contribute with their absolute value, salience increases linearly with the magnitude of the component. Figure 26 reflects the manipulation of color contrast in every scene. We can see that for the first scene, where all objects are equal, the probabilities of perceptual saliencies are equal as well. In the last scene, however, where the object's color contrast values are different, those objects with higher contrast are also more likely to be perceptually salient.

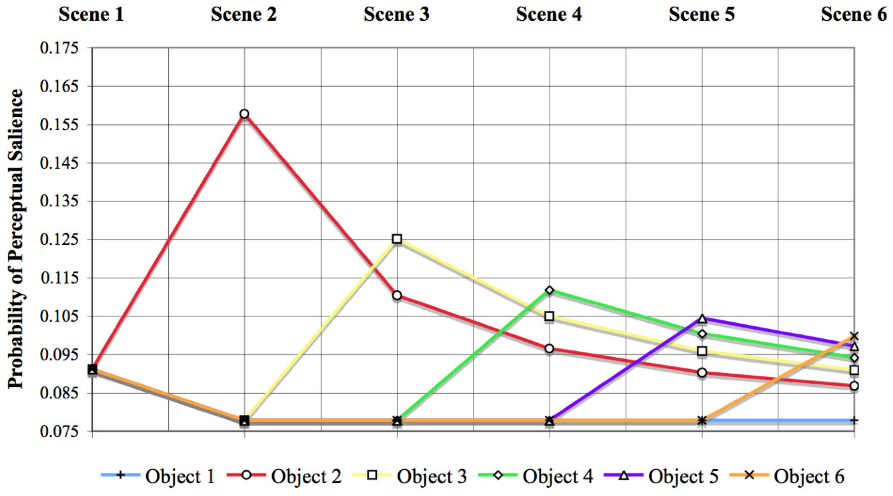


Figure 26 This figure shows how the probability of perceptual salience changes as the color contrast of the objects in the scene changes.

Figure 26 illustrates another interesting difference between dissimilarity- and magnitude-based components in our setup. The probability of perceptual salience for dissimilarity-based components is adjusted with the addition of larger objects. This is not the case for magnitude-based components. The reason therefore is the fact that probabilities for dissimilarity-based components are derived from the average, while the probabilities for magnitude-based components are based on the range of values.

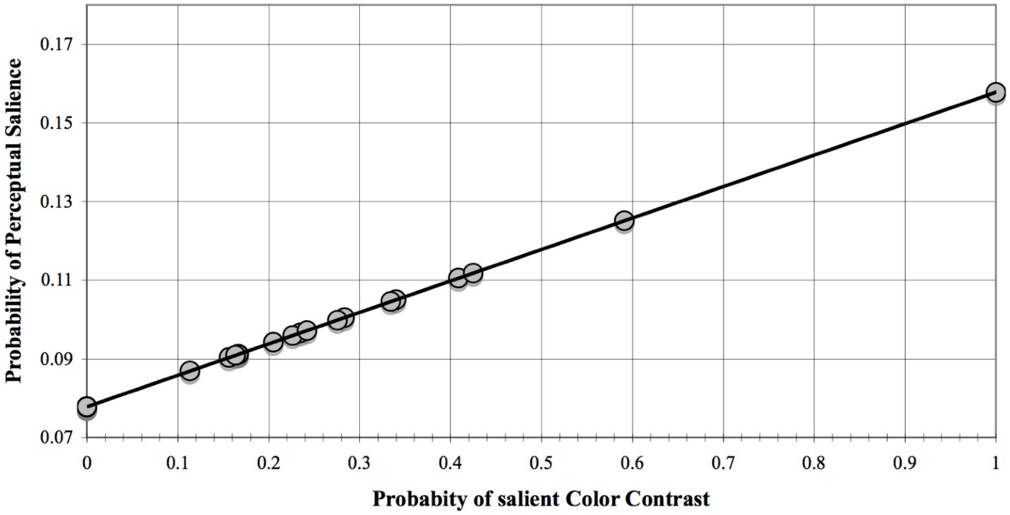


Figure 27 This figure illustrates the linear dependence of probability of perceptual salience from the probability of color contrast.

In Figure 27, we plotted the probability of color uniqueness against the resulting probabilities of perceptual salience for all scenes. The trend line illustrates the linear correlation (correlation coefficient ρ=1) between the probabilities of salient color contrast and the probabilities for perceptual salience. Note that for dissimilarity-based components, where the probabilities for salient size dissimilarities range between 0.2 and 0.5 (cf. Figure 25), the probabilities for magnitude-based components are spread over the whole range of probability (i.e., 0 … 1).

4.2.3 Verification of Integrated Saliency Assessment

The purpose of the first part of the verification was to ensure that the low-level components and the configuration of the agent contribute as expected to the outcome. In the second part of the verification, we investigate the behavior of the computational model for the integrated assessment of salience. Specifically, we want to inspect top-down influence and influence of context in more detail.

4.2.3.1 Test Data Set

For this part of the verification we will use a data set that represents the region around the Main Station of Zurich. This specific region was selected as the test site because it is rich in diversity of object types, features many well-known and recognizable objects, and is open with high visibility. High visibility ensures that objects are observable from multiple points of view, which results in redundant observations for single objects.

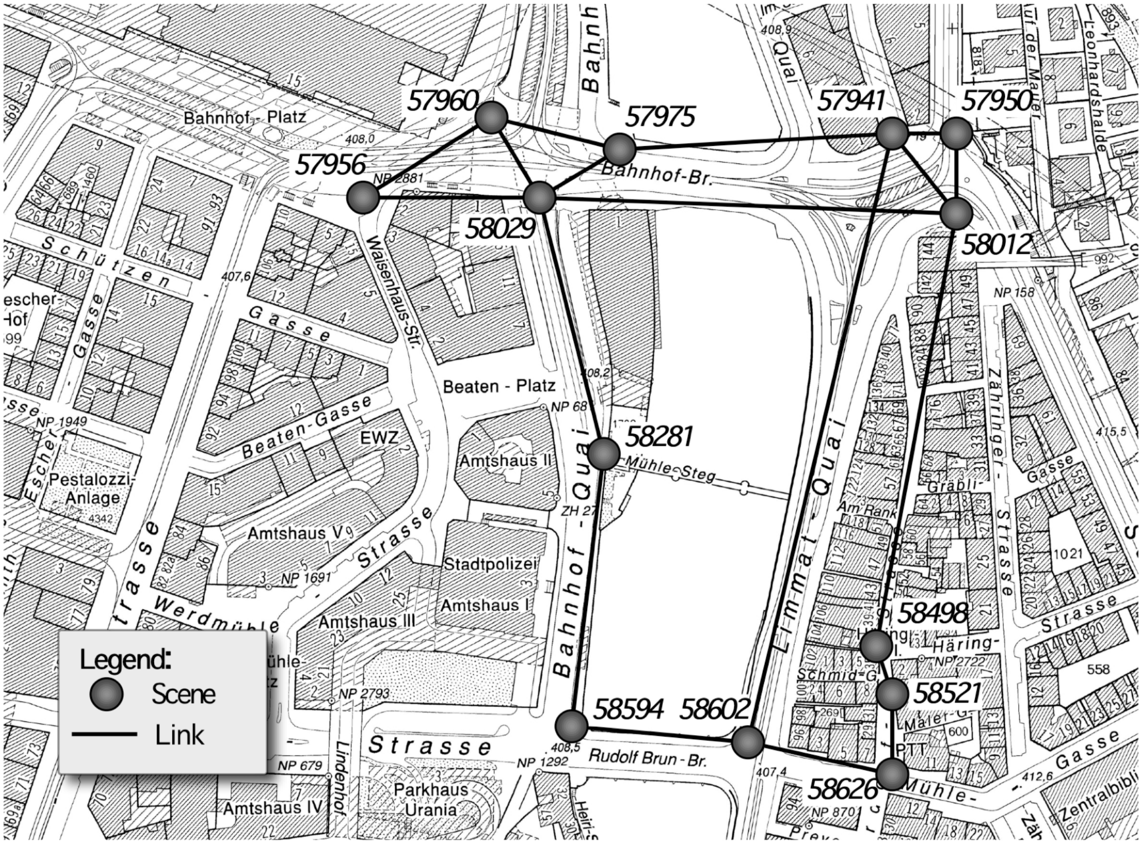


Figure 28 Data set used in the evaluation of the prototype.

The data set conforms to the specification of the computational model and consists of 13 panoramic images, including spatial objects, spatial relations between objects, and links between the scenes (Figure 28). The connections between the scenes are given by the traffic network and define the topology of the environment. The spatial objects in the panoramic images were digitized (i.e., outlines, spatial relations) and attributed corresponding to the specification of the data model. Object attributes are strictly limited to those that are directly visible within the scene (e.g., labeling, type).

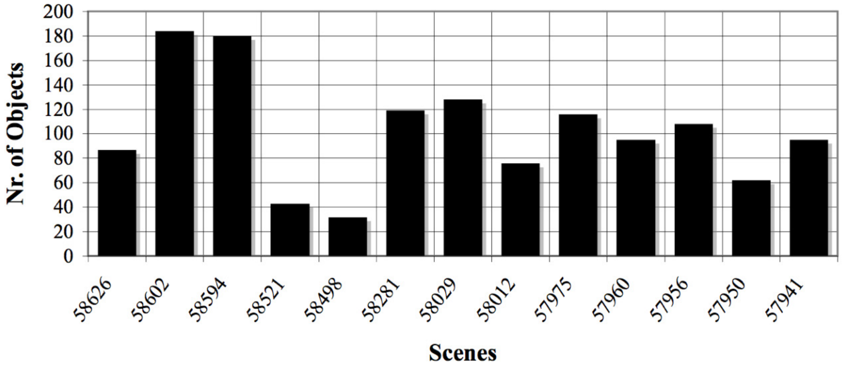


Figure 29 This diagram shows the number of objects contained in each scene. The number of objects (i.e., class instances) per scene corresponds to the number of observations per scene (i.e., one observation for every object in the scene).

The 651 spatial objects are distributed across the scenes as shown in Figure 29. The number of objects is a direct result of the visibility within the scene. Each scene contains in average around 100 objects, whereby the minimum number of objects in a scene is 32 (scene 58498) and the maximum number is 184 (scene 58602). Note that the number of objects in one scene corresponds to the number of observations in that scene, that is, no duplicate observations occur in a single scene. Redundancy in observations is achieved by observations to the same object from different scenes.

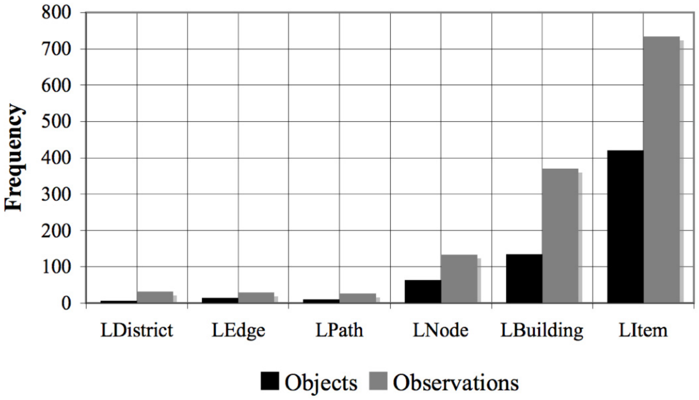


Figure 30 This figure shows the relation between objects and observations for each class, as well as their frequencies.

Figure 30 illustrates the frequency of objects and observations for each class type for the whole environment. We can see that the frequency is directly dependent on the spatial granularity of the objects. The class *LItems*, which includes the dimensionally smallest objects, has the highest frequency, both, in terms of objects and observations. The class *LDistricts*, in contrast, which represents objects that are inherently larger, has the lowest frequency. The average ratio of observations to objects is 2.6, which is a manifest of the good visibility within the test site.

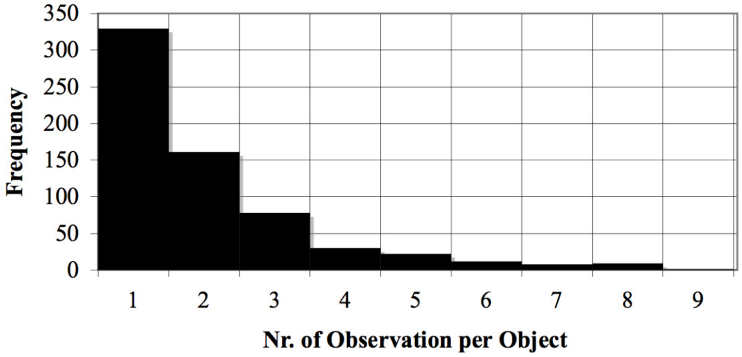


Figure 31 This figure shows the histogram for observations per objects. Most objects are observed only once and nine is the maximum number of observations for a single object.

The 621 objects are described by a total of 1325 observations. The number of observations for single objects ranges from 1 (330 objects) to 9 (1 object) with an average of 14 observations per object. The frequencies are illustrated in Figure 31. The 651 distinct spatial objects are further categorized in 101 different object types. Object types are defined as class attributes and indicate the specific type of a class instance. Among the most frequent object types are objects such as *residential houses*, *traffic signs*, *shops*, etc.

4.2.3.2 Experiment 4: Top-down Influence

In this test, we will investigate the influence of the degree of recognition and idiosyncrasy on the resulting saliency values. Recognition and idiosyncrasy are both the result of learning, and remembering. In real life, the process of learning is based on experience and practice, and may lead to long-term changes in behavior (Richardson, Montello et al. 1999). In our model, learning is implemented as sensing the environment and storing the observations in long-term memory, while change in behavior is understood as the increase of cognitive salience for memorized objects. As such, this change is a direct result of the comparison of perceptual stimuli and observations in long-term memory, and the number of observations for a specific object. It increases linearly with the number of both, degree of recognition and idiosyncratic relevance, and is directly influenced only by context (i.e., task and modality).

Setup

The setup of this experiment consists of a route leading from scene 58012 to 57960. This route was chosen because it offers the best visibility within the environment, which ensures that objects are seen from multiple scenes, and hence, redundant observations contribute to its salience. We keep the route and the environment constant, and run the simulation two times. For the first run, the agent is configured such that cognitive aspects (i.e., degree of recognition and idiosyncratic relevance) are not considered, while for the second run, these components are considered as well. The influence of cognitive aspects is investigated based on the correlation of the resulting rankings (Spearman's ρ) for the two simulation runs.

Independent Variable:	1. Run: *Agent Configuration* = (¬*LTM* ∧ *Walking*) 2. Run: *Agent Configuration* = (*LTM* ∧ *Walking*)
Control Variables:	*Route* = (58012, 57950, 57941, 57975, 57960) *Environment*
Observed Variables:	*Perceptual Salience* *Cognitive Salience* *Contextual Salience*

Requirements

- The probability of cognitive salience increases with the number of observations and the degree of recognition for specific objects, which affect perceptual and contextual salience.
- The correlation of the resulting rankings for this simulation, hence, will vary with the influence of cognitive salience.
- Given the ability of storing observations in memory is the only independent variable, then cognitive salience increases faster than perceptual and contextual salience.

Results

The results show that in the first scene (58012) the rankings correspond exactly to each other. This perfect relationship is due to the fact that no observations are stored in memory yet. For the following scenes (57950 to 57960), we observe a decrease in the rank correlation coefficients. This implies that those objects that have been observed before and are now stored in long-term memory contribute to salience, which results in different rankings and consequently in lower correlation coefficients.

The results also show that the correlation coefficients for cognitive salience decrease faster than the coefficients for perceptual and contextual salience (cf. Figure 32), which is expected since the availability of long-term memory is the only independent variable. Due to the linearity of the model, the probabilities of perceptual and contextual salience follow the trend given by the probability of cognitive salience. Ultimately, this results in the adjustment of the respective rankings, and hence, a reduction of their correlation coefficients. This corresponds to the requirements postulated above.

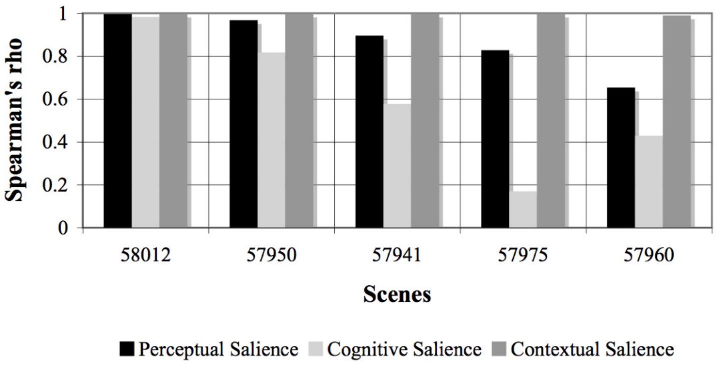
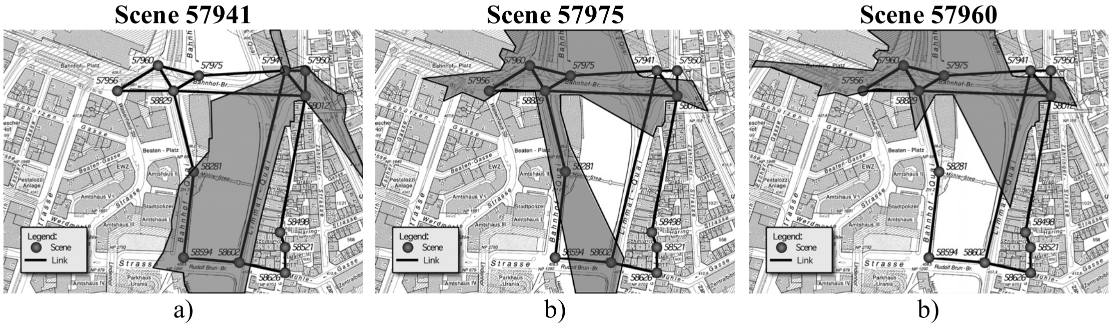


Figure 32 The diagram shows the decrease of Spearman's ρ for perceptual, cognitive, and contextual salience of the test case. The decrease in rank correlation is a result of the influence of storing observed objects in memory.

Another interesting aspect of the influence of degree of recognition and idiosyncrasy is present in the last three scenes visited by the agent. The correlation coefficient for the rankings based on the probability of cognitive salience decreases considerably between scene 57941 and 57975. This peculiarity is due to the higher visibility at node 57975, and to the lower redundancy of observations.



a)

b)

b)

Figure 33 The figures a), b), and c) illustrate the visibility of nodes 57941, 57975, and 57960. 25 objects are seen from both node 57941 and 57975, while 49 are seen from 57975 and 57960.

Figure 33 illustrates the visibility at nodes 57941, 57975, and 57960. The set of visible objects amounts to 95 for scene 57941, 116 for scene 57975, and 95 for scene 57960. The set of objects that are visible from both scene 57941 and 57975 is smaller than the set of objects that is jointly visible from scene 57975 and 57960 (i.e., 25 vs. 49 observations). As a result of reduced visibility and the small redundancy of observations, the correlation coefficient for scene 57975 drops below 0.2, while it increases noticeably for the following scene (i.e., ~0.4 for scene 57960).

4.2.3.3 Experiment 5: Influence of Context

Context directly influences the salience of objects in the environment (Golledge 1999). In our model, the influence of context is determined based on eccentricity of objects in the scene with respect to the target path, and the current modality (i.e., walking or driving). In this experiment, we will investigate the behavior of the model as a result of contextual influence.

Setup

The setup for this experiment consists of a route leading from node 57956 to 58626 for the first run, and from 58626 to 57956 for the second run. The combination of the two runs corresponds to a roundtrip, whereby the agent visits the same node in reversed order on the way back. This setup leads to multiple observations for single objects, but because we are focusing on the influence of context, we do not consider cognitive influences.

In order to analyze the contextual influence in more detail, we focus on a set of eight objects (i.e., objects *A* to *H*) for each scene. These objects are distributed evenly across the horizon in each scene, which facilitates direct comparison between the salience of the object in the first run and the second run. Furthermore, we set the agent's modality to walking. This setting allows for differentiated identification and analysis of task-based influence, rather than the combination of task and modality. The verification of modality-based context is investigated in the next section. The settings for the verification of task-based influence are summarized below:

Independent Variable:	1. Run: *Route* = (57956, 58029, 58281, 58594, 58602, 58626) 2. Run: *Route* = (58626, 58602, 58594, 58281, 58029, 57956)
Control Variables:	*Agent Configuration* = (¬*LTM* ∧ *Walking*) *Environment*
Observed Variables:	*Perceptual Salience* of objects *A* to *H* in all scenes *Contextual Salience* of objects *A* to *H* in all scenes

Requirements

- Task and modality act as a function that selectively weights objects considered important for continuation of the journey.
- The salience of objects is higher if their location is relevant for navigation. It varies with the eccentricity of the objects with respect to the target path.

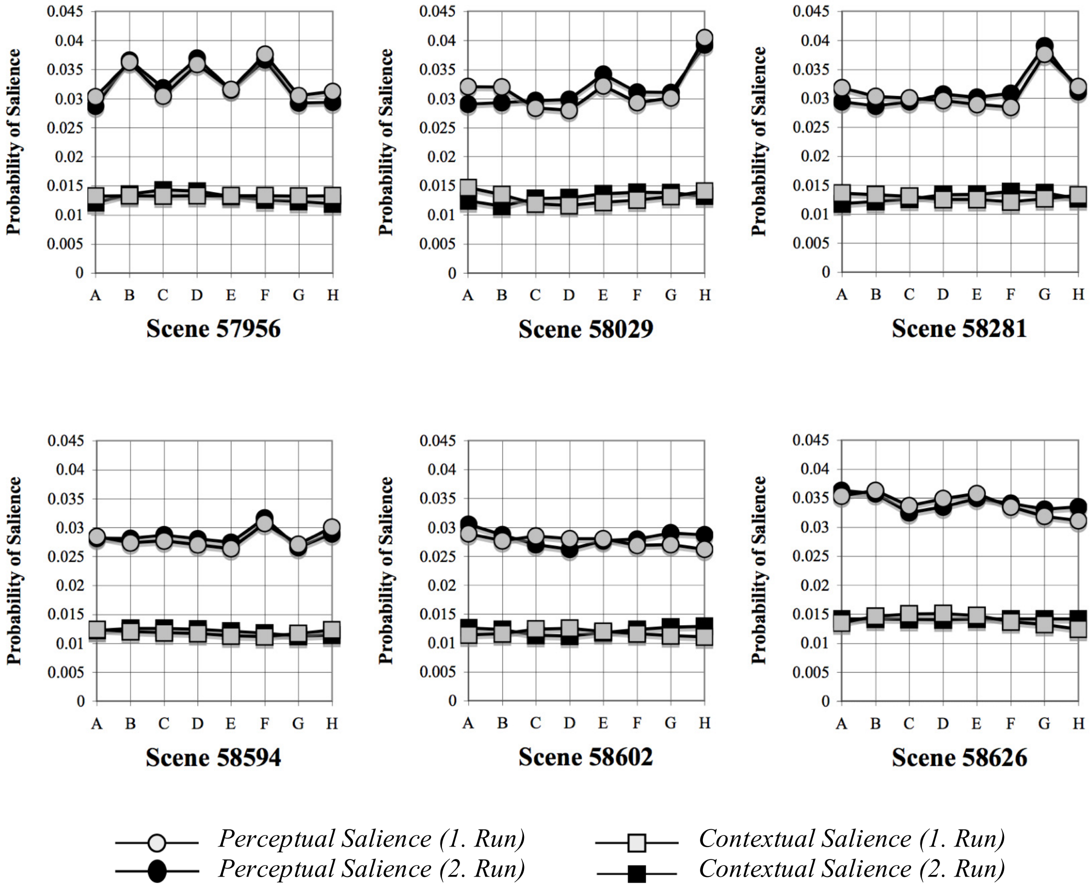


Figure 34 The figures illustrates the results of Experiment 5. The graphs show the Perceptual saliencies for the six scenes for both, the first run and the second run.

The results of this experiment are illustrated in Figure 34 and clearly show the weighting effect of task-based influence. The objects A to H are labeled from left to right with respect to the scene. As a result of this sequential labeling, we observe a continuously changing probability of contextual salience for these objects. This continuous change in probability shows that task-based influence acts as a weighting function for the objects in the scene, which depends on the object's location with respect to the target.

Besides the continuous change of probability, we also observe a mutual dependence between the probabilities for the first and second run. This dependence is due to the fact that objects that are located closest to the target path in the first run are located the furthest away in the second run. With exception of the first and the last scene, this fact is reflected by all probabilities values for contextual salience. The probability of contextual salience for the last scene of the first run (i.e., scene 58626) and the last scene of the second run (i.e., scene 57956) are different, because task-based influence is not considered in the last scene of a route. Consequently, for the last scene of each run, all objects have equal probability of being contextually salient.

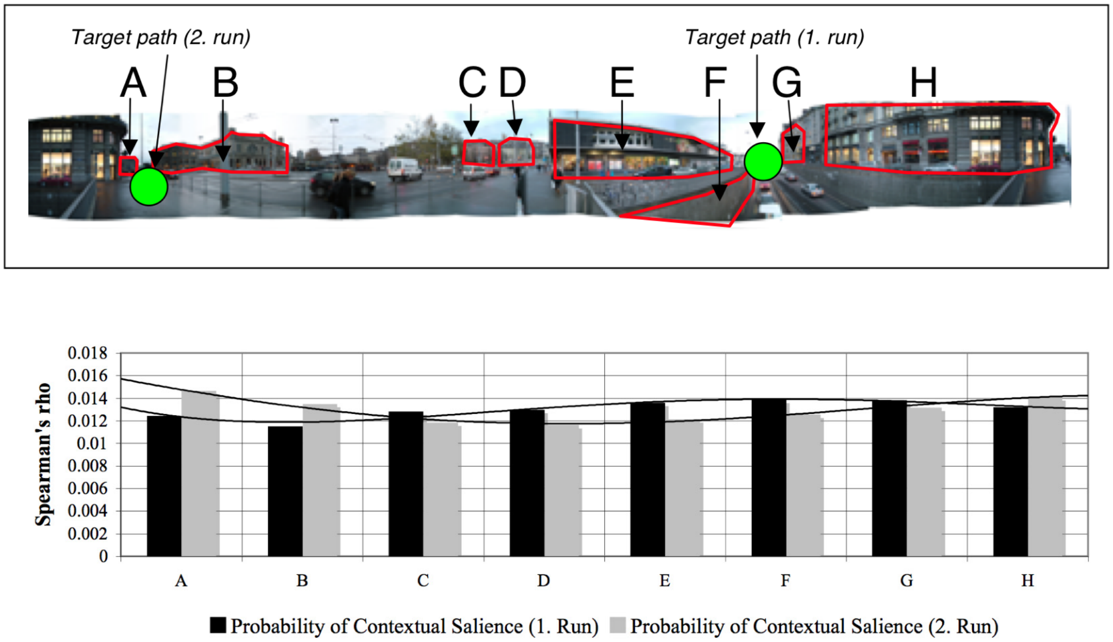


Figure 35 This image shows the spatial setup of scene 58029 and the objects (A to H) along with their probabilities of contextual salience for both runs. The probabilities change according to the eccentricity of the objects with respect to the target path.

In addition to the weighting of the probabilities inflicted by task-based influence, the results also show the influence of context on the probability of perceptual salience. Specifically, we observe a linear dependency of perceptual salience on the influence of context, which causes an adaptation of the resulting rankings to the given context. The probabilities of the objects in scene 58029 (cf. Figure 35), for instance, differ such that the rank correlation coefficient for the rankings for the first and second run is only marginal (ρ=0.349). This finding, together with the aspects discussed above, confirm the requirements tested in this experiment.

4.2.3.4 Experiment 6: Walking vs. Driving

The final experiment for verification of the prototype is concerned with investigating the influence of modality on the salience of objects. The modality of travel influences both, the degree of cognitive resources that can be allocated to the discrimination and assessment of objects in the environment, as well as the field of view. In our computational model, two types of modality are implemented, namely *walking* and *driving*. Walking assumes that navigators will stop at decision points and look around in order to find appropriate objects of reference, which results in an extended field of view of 360°. Driving, in contrast, requires navigators to direct their gaze along the direction of travel, whereby the direction of travel typically coincides with the target path. In this experiment, we will investigate the peculiarities of each modality and compare the results.

Setup

For inspecting the influence of modality, we will perform two simulation runs along the same route (i.e., from scene 58012 to scene 58626). For the first run, we will set the modality to *walking*, and for the second run, to *driving*. As we are interested in the influence of modality only, we will not consider cognitive aspects and disable the agent's ability to store observations in long-term memory. Besides the agent's configuration, we will keep all variables controlled and compare the results of the first and second run (i.e., walking vs. driving).

Independent Variable:	1. Run: *Agent Configuration* = $(\neg \mathit{LTM} \wedge \mathit{Walking})$ 2. Run: *Agent Configuration* = $(\neg \mathit{LTM} \wedge \mathit{Driving})$
Control Variables:	*Route* = (58012, 58029, 58281, 58594, 58602, 58626) *Environment*
Observed Variables:	*Perceptual Salience* of objects *A* to *H* in all scenes *Contextual Salience* of objects *A* to *H* in all scenes

Requirements

- Walking allows for equal consideration of all objects within the scene.
- Driving emphasizes those objects that are within the field of view, that is, the probability of salience of objects that are within the field of view is higher than for those that are not.

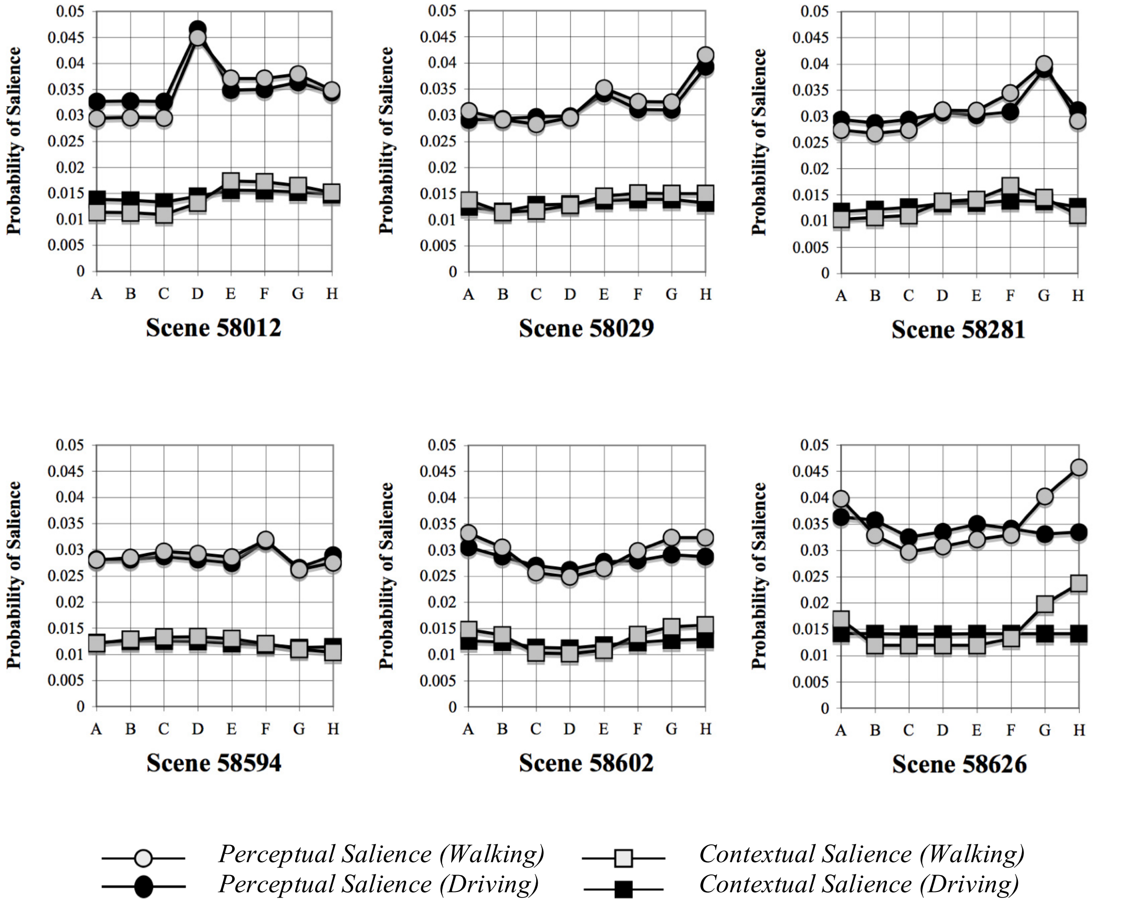


Figure 36 The figures illustrates the results of Experiment 6. The graphs show the resulting probabilities for perceptual and contextual salience of the objects encountered along the way for both modalities, i.e., walking and driving. Overall, we observe that objects in the field of view are assumed to be more important than other objects.

The results of this experiment confirm the requirement the driving emphasizes those objects that are within the field of view. The results of this experiment are illustrated in Figure 36 and show the influence of modality on the resulting probabilities for contextual and perceptual salience. We can see that the probabilities for contextual salience for driving emphasize the trend given by the task-based influence. In other words, if task-based influence increases, the influence of modality increases as well, and if task-based influence decreases, then the influence of modality decreases as well. This behavior is due to the fact that the influence of modality is dependent on the incoming and outgoing fields of view.

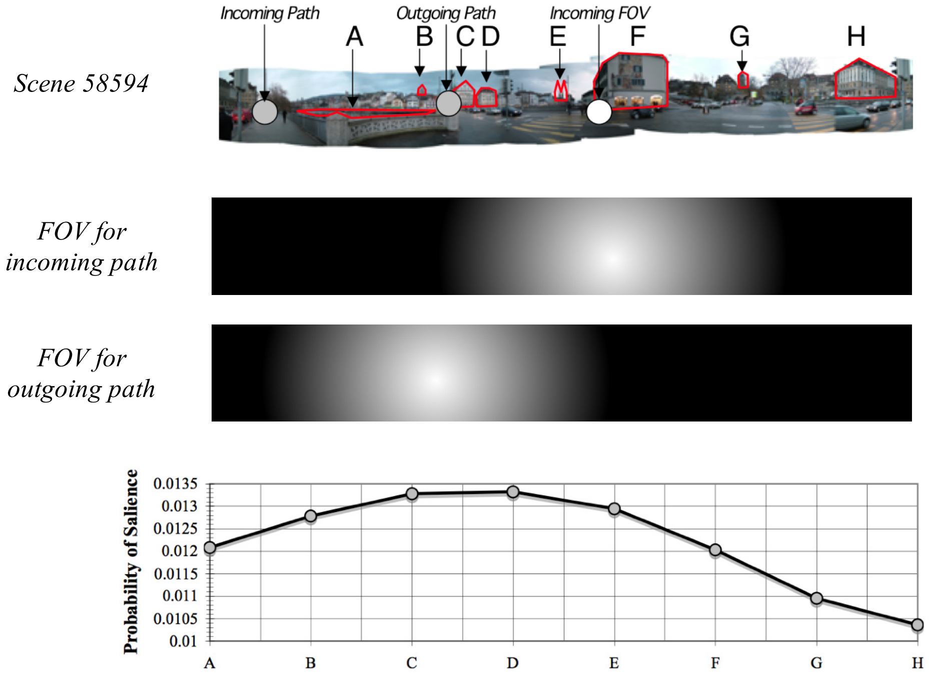


Figure 37 This figure illustrates the influence of modality in terms of the fields of view that are used for its calculation.

The fields of view for incoming and outgoing direction coincide in cases where the incoming and the outgoing path are aligned in a straight line. They vary, however, if the navigator has to go left or right at some decision point. Consider scene 58594 for instance (cf. Figure 37), where the route plan dictates a change of direction towards the target path. This change of direction corresponds to a 90° turn to the agent's left and results in differently oriented fields of view for incoming and outgoing direction. The field of view, hence, differs from the field of view for the incoming path, as shown by the illustrations in Figure 37. For the assessment of the probability of contextual salience, the combination of the two fields of view are combined, which results in the probabilities of contextual salience shown in Figure 37.

4.2.4 Summary of Verification Results

The verification of the prototype confirms the correct implementation of the computational model. In the first part of the verification, we have shown that equal objects yield equal probabilities of saliencies, and we have also shown that the different characteristics of low-level components (i.e., difference or magnitude) contribute correctly to the resulting saliencies. In the second part, we have shown that the integrated saliency assessment considers top-down influences according to the specifications described in the previous chapter. Furthermore, we have also shown that contextual influences (i.e., task and modality) contribute correctly to salience. Consequently, we conclude that the implementation of the computational model performs according to the specifications.

4.3 Validation of Framework

The goal of the validation is to determine how well the rankings produced by the computational model correlate with corresponding rankings from real-world scenarios. Typically, validation is achieved by using a set of benchmark cases, for which a correct diagnosis is known. In our case, we will use the results of an online survey as benchmark cases for the comparison. In the following sections, we will describe the method of validation in further detail, conduct the experiments, and present the results.

4.3.1 Method of Validation

The purpose of the computational model is the production of rankings that reflect the human assessment of landmark salience. In order to evaluate the correlation between results generated by the computational model and human assessment of landmark salience, we will conduct four experiments. Each of the experiments is designed such that a specific configuration of the computational model is validated. The configurations are based on different scenarios of navigation, including: 1) exploring an unfamiliar environment, 2) wandering around in a familiar environment, 3) executing a route plan in an unfamiliar environment, and 4) executing a route plan in a familiar environment. These scenarios correspond to the different combinations of cognition and context, as defined in our conceptual model. The following table summarizes the combinations and the according scenarios:

	Perception	
	¬ Cognition	**Cognition**
¬ Context	*Scenario 1*	*Scenario 2*
Context	*Scenario 3*	*Scenario 4*

Note that we use the logical negation ('¬') to indicate presence or absence of cognition and context. These four scenarios provide the foundation for the definition and formulation of the four experiments that we will conduct for validating the prototype. For each of the scenarios, we will generate the corresponding ranking using the computational model and the data described in section 4.2.3.1. These rankings will be compared with the results of an online survey. The online survey is set up such that the benchmark data for the four scenarios can be deduced. The correlation of the two rankings is statistically compared in order to measure the performance of the computational model.

4.3.2 Online Survey of Landmark Salience

For the collection of the benchmark data, we conducted a survey in the form of an online questionnaire. There are many differences between running a fully controlled experiment in the lab, and having subjects perform the experiment on their own in an uncontrolled environment. In the former case, information about conditions in which the experiment is performed, personal information, as well as other details may be elicited. Conducting an experiment over the web, in contrast, enables the subjects to stay absolutely anonymous, which is a major a drawback, as controlling and monitoring the subjects' performance in terms of task comprehension and influencing factors is impossible. In addition, motivating the subjects to conduct the experiment and keeping them concentrated throughout the process is not possible.

In our case, transferring the experiment outside the controlled lab into the subjects' own environment results in three types of problems: 1) retrieving personal details relevant to the experiment, 2) identifying prior knowledge and experience of participants, and, 3) documenting important aspects, such as reasoning processes and strategies, which is the most critical aspect. Obviously, these problems cannot be solved completely. Nevertheless, we try to retrieve as many parameters as possible in order to increase reliability and accuracy of the results. In the following section, we describe and explain the setup of the survey, the participants, and the nature of the questionnaire.

4.3.2.1 Setup and Participants

The setup of the survey consisted of a single form, which was divided into three parts. In the first part, participants answered questions related to their background, in the second part, a rating of the visual prominence of urban objects was performed, and finally in the third part, the objects were rated based on their relevance for navigation. The ratings for both parts were performed on the basis of the test data set described in section 4.2.3.1, which includes 13 panoramic images of traffic nodes in the city of Zurich.

In order to approximate the field of view navigators would have in the real world, the panoramic pictures were displayed in a viewer that allowed for panning and zooming. At the highest zoom level, participants would be able to see about half of the panoramic picture (i.e., ~180°), which corresponds more or less to the extent of the human visual field. Zooming into the image would reduce the visible extent of the image, as is the case for focusing the human visual system on single objects. Allowing participants to drag the image to the left or to the right approximated head and body movement.

Although only eight scenes were displayed in the questionnaire (i.e., four for each part), all 13 scenes were used in the experiment. The online form was designed and configured such that every time the page was loaded, the full set of panoramic images was shuffled and a subset of scenes were loaded and displayed. The first four panoramic images were displayed for the first task (i.e., visual importance), and the second four images for the second task (i.e., task-relevance of objects). Picking the first eight from the array of 13 images ensured that all images in the questionnaire were unique. Dynamically selecting the images was necessary in order to avoid systematic influences due to the sequence of the panoramas and for getting about the same rating frequency for all 13 scenes. Furthermore, a mechanism for filtering missing or invalid answers was implemented, which ensured that participants completed the questionnaire before submitting their answers.

The invitation for participation was sent out to different groups of people, including staff and students of several academic institutions, professionals with various backgrounds, etc. The main goal of the survey was to collect data for a first validation of the computational model, not to collect the data for its fine-tuning. Therefore, no information related to cultural or social background of participants was collected.

4.3.2.2 Questionnaire

The questionnaire used for the online survey was available in both German and English. It consisted of three parts with a total of 13 questions. The first part included five questions related to the participant's background. The background information included *gender*, *age*, *familiarity* with the site of the test data set (i.e., region around the Train Main Station in Zurich), as well as the *preferred means of transportation* in urban environments. The main purpose of the background questions was to separate participants who have prior knowledge of the test site from participants without prior knowledge. This separation is necessary in order to investigate scenarios 1 and 3 (i.e., presence and absence of cognitive influences).

In the second part, participants were asked to rate the visual importance of a set of panoramic images. Specifically, participants were presented with four panoramic images, in which eight objects of the urban environment were identified in a legend below the image. The participants were ask to rate the visual importance of the objects on a scale from 1 (=least salient) to 99

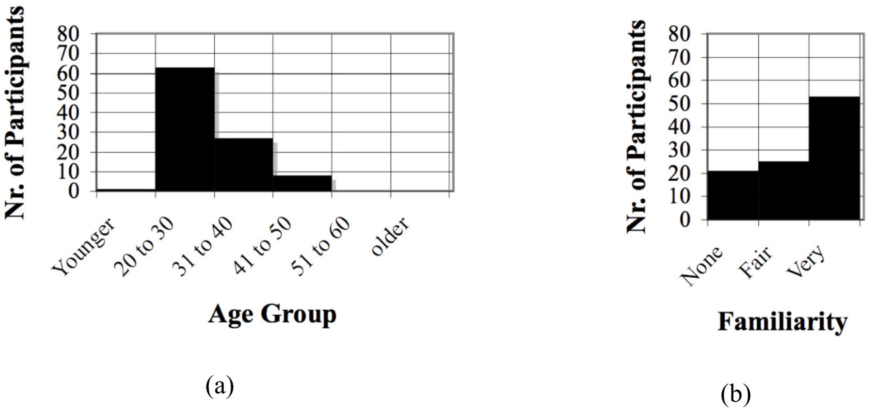
(=most salient). The demarcated objects were chosen randomly, but are distributed evenly across the scene and vary in size, shape, and location.

For the last section, participants were asked to imagine that they are traveling along a predefined route and have to continue their journey in the direction of a target path. The purpose of this section was to rate the objects based on their relevance for describing the target path, whereby a green arrow was used to identify the target path. As in the previous section of the survey, eight objects were demarcated in the legend and participants were asked to rate the importance of these objects for the continuation of their journey. The rating scale was the same as in the previous experiment (i.e., 1 for most important, and 99 for least important).

4.3.2.3 Results

Participants

A total of 99 participants filled out the questionnaire. From these 99, five were eliminated due to erroneous input (e.g., same ratings for all objects, reversed ratings due to use of wrong rating scale). Of the remaining 94 participants, 33 were female and 61 male. 79 participants chose to fill out the form in German and 15 in English. Most participants were between 20 and 30 years of age (cf. Figure 38). From the 94 participants, 21 indicated that they had no prior knowledge of the test region, 20 indicated that they are fairly familiar with it, and the majority (53 participants) pointed out that they are very familiar with the test region.



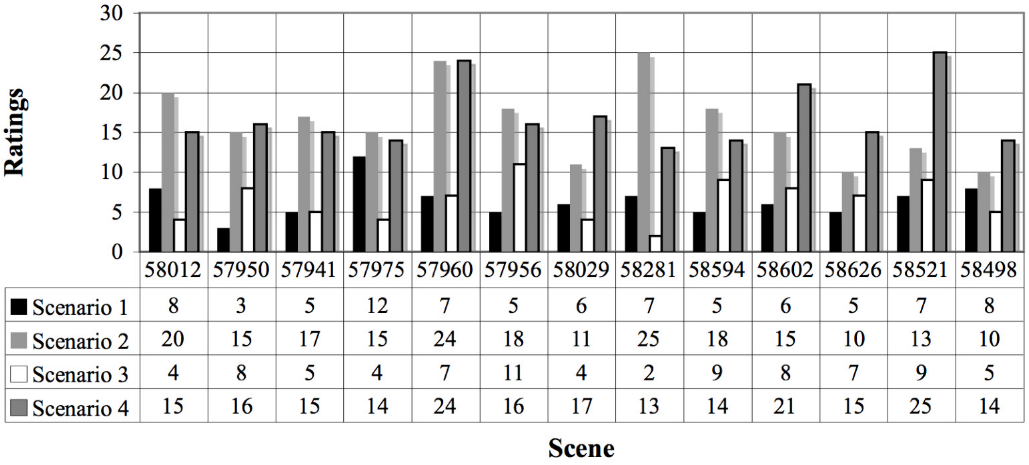
(a)

(b)

Figure 38 The histograms shows the distributions of the participants with respect to (a) age groups, and (b) familiarity with the test region. The majority of participants were between 20 and 30 years of age and familiar with the environment.

Summary of Ratings

The 13 scenes were rated 632 times. For the validation process, we divided the ratings in two separate sets, which reflect the degree of prior knowledge of the participants. The first set contains ratings from participants without any prior knowledge of the test region, while the second set is comprised of ratings from participants who are very familiar with the region. In order to ensure clarity of results, the set of ratings from participants with fair knowledge was not considered for validating the computational model. Figure 39 illustrates the frequency of ratings with respect to scene and scenario. Note that the figure summarizes only those ratings that are used for validating the computational model (i.e., now prior knowledge and very familiar, but not fair familiarity).



	58012	57950	57941	57975	57960	57956	58029	58281	58594	58602	58626	58521	58498
■ Scenario 1	8	3	5	12	7	5	6	7	5	6	5	7	8
▨ Scenario 2	20	15	17	15	24	18	11	25	18	15	10	13	10
□ Scenario 3	4	8	5	4	7	11	4	2	9	8	7	9	5
▩ Scenario 4	15	16	15	14	24	16	17	13	14	21	15	25	14

Figure 39 The diagram shows the frequency of ratings for each scene for all four scenarios. The lowest number of ratings was two for scenarios 1 and 3, which involve no prior knowledge of the test site.

The number of ratings for scenarios 1 and 3 is relatively low (i.e., between 2 and 12) compared to scenarios 2 and 4, where the number of ratings per scene varies between 10 and 25. This is due to the low number of participants without prior knowledge of the test site (21 participants). Because of the lack of redundancy, scenes with less than 5 ratings were not used in the validation process. This includes scene 57950 for scenario 1, as well as scenes 58012, 57975, 58029, and 58281 for scenario 3. For scenarios involving prior knowledge, however, all 13 scenes were used.

4.3.3 Validation

The validation of the computational model is organized in four parts. Each part corresponds to one of the four scenarios described in section 4.3.1 and is investigated individually. The following sections describe the procedure and the results for the different scenarios. Note that because missing benchmark data for assessing the agent's performance for driving, we restrict the modality for the four scenarios to walking.

4.3.3.1 Procedure

For each scenario, we will compute the probability of salience for the objects, establish the according rankings for the scene, and compare these ranking with the corresponding rankings from the survey. The data from the survey serve as benchmark and the quality of the results is measured in terms of Spearman's rank correlation coefficient, which we will use to test if there is indeed a correlation, and to test the strength of the potential correlation.

In order to increase the explanatory power of the validation, we will run the simulation using the full set of scenes. The degree of correlation, however, is computed only for scenes with five or more ratings. For every scene, we computationally derive the probability of salience for the eight objects (i.e., objects A to H) and convert the results into a ranking. This approach produces a total of 13 rankings that can be used for comparison.

The validation procedure is tailored such that general conclusions about the computational model can be deduced. Specifically, we want to test the behavior of the model, not the correctness of single variables. Hence, the hypothesis we want to test is the null-hypothesis: *There is no significant relationship between the probabilities of salience generated by the computational model and the ratings from the survey.* Spearman's rank correlation test is a popular method for investigating possible linear associations in populations underlying sample sets.

We use the following method for aggregating the individual ratings of single objects into a single ranking. The set of ratings for a specific object defines a discrete distribution from which we extract the *median*. The set of medians for all objects in a scene defines a ranking, which will be used as benchmark to test the ranking generated by the computational model. The reasons for using the median instead of the *mode* or *mean* are manifold. The main reason, however, is the fact that we are dealing with subjective ratings on a scale from 1 to 99. Individual ratings are based on subjective ranges of importance for the objects in the scene, which would result in biased rankings, and hence, excludes using the mean for aggregation

The median was favored over the mode because it is unique and always computable. The mode, in contrast, is not unique (i.e., multiple modes are possible within a single data set), and not present in data sets that contain no repeated elements. Besides these advantages, mean and median have the same characteristics with respect to extreme values. This procedure for extracting rankings is applied for each scene and scenario, resulting in a total of 47 rankings (i.e., 4 rankings for each scene minus the five scenes with less than five ratings).

4.3.3.2 Scenario 1: Exploring an unfamiliar Environment

Scenario one assumes that navigators have no prior knowledge of the environment, and that they are just exploring the area without any specific goal or task. For the simulation of scenario 1, hence, we disable the agent's ability to store observations in long-term memory. In addition, we also disable the contextual influences (i.e., task-based influence and modality). The route used for running the simulation, includes all 13 scenes, starting at 58012 and leading to 58498. For the assessment of the correlation coefficients, however, only the 12 scenes with at least five ratings are considered. In summary, we used the following configuration for the simulation of scenario 1:

Route:	(58012, 57950, 57941, 57975, 57960, 57956, 58029, 58281, 58594, 58602, 58626, 58521, 58948)
Agent Configuration:	($\neg LTM \wedge Walking$) (Cognitive and Contextual influences disabled)
Observed Variables:	*Perceptual Salience* of objects A to H in all scenes

The probabilities of perceptual salience resulting from the simulation were translated into rankings and tested for correlation with respect to the rankings derived from the survey data. The ranking for the survey was derived from part one of the survey (i.e., visual salience of objects) and on the basis of ratings from participants without prior knowledge of the test site. The resulting ranking corresponds to the configuration of the agent as stated above.

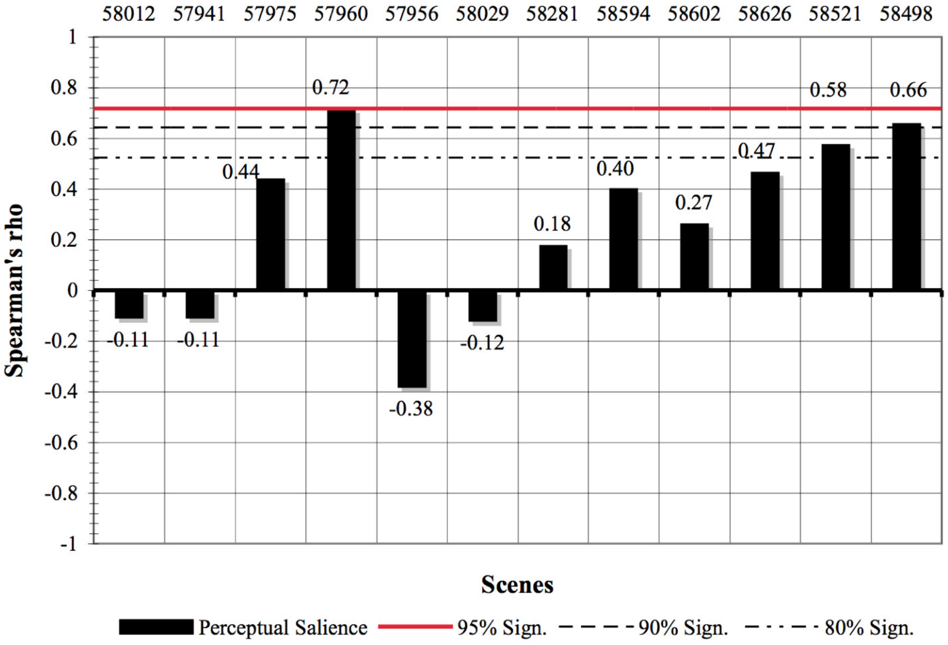


Figure 40 The figure shows Spearman's Correlation Coefficients for the 12 scenes of scenario one. The coefficients are mainly positive, but only one is within the 95% significance level.

Spearman's correlation coefficients for the 12 scenes are summarized in Figure 40. The figure shows that although eight of twelve correlation coefficients are positive, only one correlation coefficient for the rankings of scene 57960 is statistically significant at the 95% level. The correlation coefficient for scene 58498 is significant at the 90% level and at the coefficient for scene 58521 at the 80% level. Note that all correlation coefficients are tested for significance based on the number of pairs *n*=8 of the two rankings, which results in a significance level of 0.717 for the 2-tailed 95% threshold. The diagram also shows the significance levels for the 90% threshold (0.643) and for the 80% threshold (0.524). While the 90% and 80% significance levels do not indicate a very reliable correlation coefficient, they still provide valuable information about the general tendency of the model's performance.

4.3.3.3 Scenario 2: Wandering around in a familiar Environment

This scenario simulates the case where a person familiar with the environment wanders around and observes the objects in the environment. In this case, the assessment of objects' salience is influenced by knowledge and former experience of the observer. The according setup for the agent is such that observations are stored in memory and used to assess the degree of recognition and idiosyncrasy, but that no contextual influence is considered. This simulation produces probability values for both perceptual and cognitive salience. The setup of the agent is summarized below:

Route: (58012, 57950, 57941, 57975, 57960, 57956,
58029, 58281, 58594, 58602, 58626, 58521, 58948)
Agent Configuration: (*LTM* ∧ *Walking*)
(Contextual influences disabled)
Observed Variables: *Perceptual Salience* of objects *A* to *H* in all scenes
Cognitive Salience of objects *A* to *H* in all scenes

The simulation results for scenario two include two rankings, one for perceptual salience and one for cognitive salience. These two rankings are compared to the ranking derived from the survey. The ranking from the survey was established on the basis of part one of the questionnaire (i.e., visual prominence) and the ratings from participants very familiar with the environment. Comparing both rankings from the simulation with the same ranking from the survey may convey additional information about weighting properties of the assessment process (cf. Chapter 6). The correlation coefficients are summarized in Figure 41.

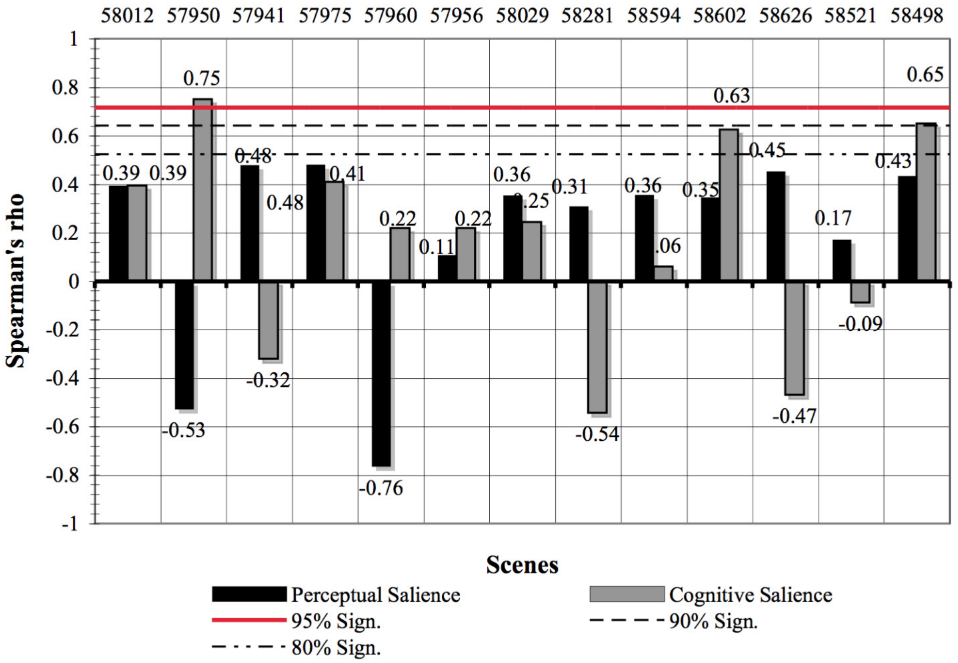


Figure 41 The figure shows Spearman's Correlation Coefficients for scenario 2.

Figure 41 shows that for the correlation coefficients for perceptual salience, 11 of 13 coefficients are positive. While most of the correlation coefficients are between 0.3 and 0.5 (9 of

13), none of them are statistically significant at the upper bound. For scene 57960, we observe a significant negative correlation at the 95% level and for scene 57950 at the 80% level. For the correlation coefficients of the ranking based on cognitive salience and the survey, 10 of 13 coefficients are positive. As in scenario 1, however, only one of the correlation coefficients is statistically significant at the 95% level. Two correlation coefficients (i.e., for scene 58602 and 58498) are inside or just outside the 90% significance level. For scene 58281, we observe a negative correlation that is statistically significant at the 80% level.

4.3.3.4 Scenario 3: Route following in an unfamiliar Environment

In scenario 3, we simulate the case of a person following a pre-defined route plan, but without being familiar with the environment. In this case, we need to consider contextual influence, but not cognitive components. Therefore, we disable the agent's ability to store observations and enable it to process contextual information. The agent's configuration is summarized below:

Route:	(58012, 57950, 57941, 57975, 57960, 57956, 58029, 58281, 58594, 58602, 58626, 58521, 58948)
Agent Configuration:	$(\neg \mathit{LTM} \wedge \mathit{Walking})$
Observed Variables:	*Perceptual Salience* of objects *A* to *H* in all scenes *Contextual Salience* of objects *A* to *H* in all scenes

As in the previous scenario, we compare the ranking from the survey with the two rankings from the simulation. The benchmark ranking for this scenario was established from those ratings of part two of the survey, where participants indicated that they had no prior knowledge of the test region. Note that as in the previous scenario, the comparison between the three rankings (i.e., benchmark, ranking for perceptual salience, and ranking for contextual salience) is supposed to provide insight into task-based weighting of objects (cf. Chapter 6). The resulting correlation coefficients for this scenario are illustrated in Figure 42.

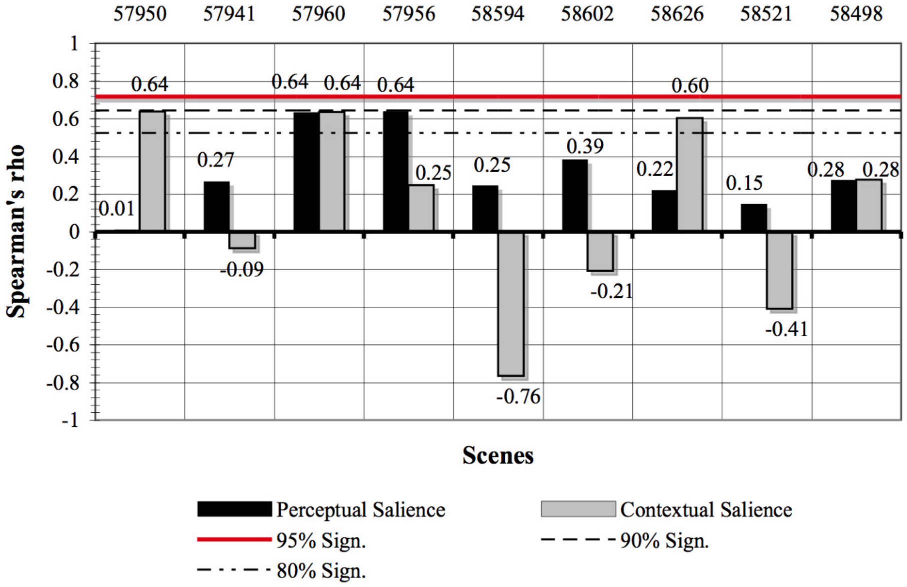


Figure 42 The figure shows Spearman's Correlation Coefficients for scenario 3. The bars indicate the strength of both, the correlation for survey results and perceptual salience, and survey and cognitive salience.

For scenario 3, none of the nine correlations are statistically significant at the upper 95% level. For the rankings based on perceptual salience and the survey, two correlations are significant at the upper 90% level (i.e., scenes 57960 and 57956), while the remaining correlations (7 of 9) are all positive, but none statistically significant. For the correlation coefficients for contextual salience and the rankings from the survey, 5 of 9 are positively correlated. Two correlations are statistically significant at the upper 90% level (i.e., scenes 57950 and 57960) and one at the lower 95% level (i.e., scene 58594). The correlation for scene 58626 is statistically significant at the 80% level, while the remaining coefficients are all lower and less reliable.

4.3.3.5 Scenario 4: Route following in familiar Environment

In the last scenario, we simulate the case of a person following a route in a familiar environment. Therefore we enable the agent's long-term memory and consider contextual influence as well. The agent's configuration is summarized below:

Route:	(58012, 57950, 57941, 57975, 57960, 57956, 58029, 58281, 58594, 58602, 58626, 58521, 58948)
Agent Configuration:	(*LTM* ∧ *Walking*)
Observed Variables:	*Perceptual Salience* of objects *A* to *H* in all scenes *Cognitive Salience* of objects *A* to *H* in all scenes *Contextual Salience* of objects *A* to *H* in all scenes

The results from the simulation are compared to the ranking derived from the survey. For establishing the ranking from the survey, only ratings from the second part of the questionnaire (i.e., task-dependent relevance) where participants indicated to be very familiar with the environment were considered. The correlation coefficients were calculated for all three rankings from the simulation in order to investigate weighting effects between perception, cognition, and context. Figure 43 summarizes the correlation coefficients for scenario 4.

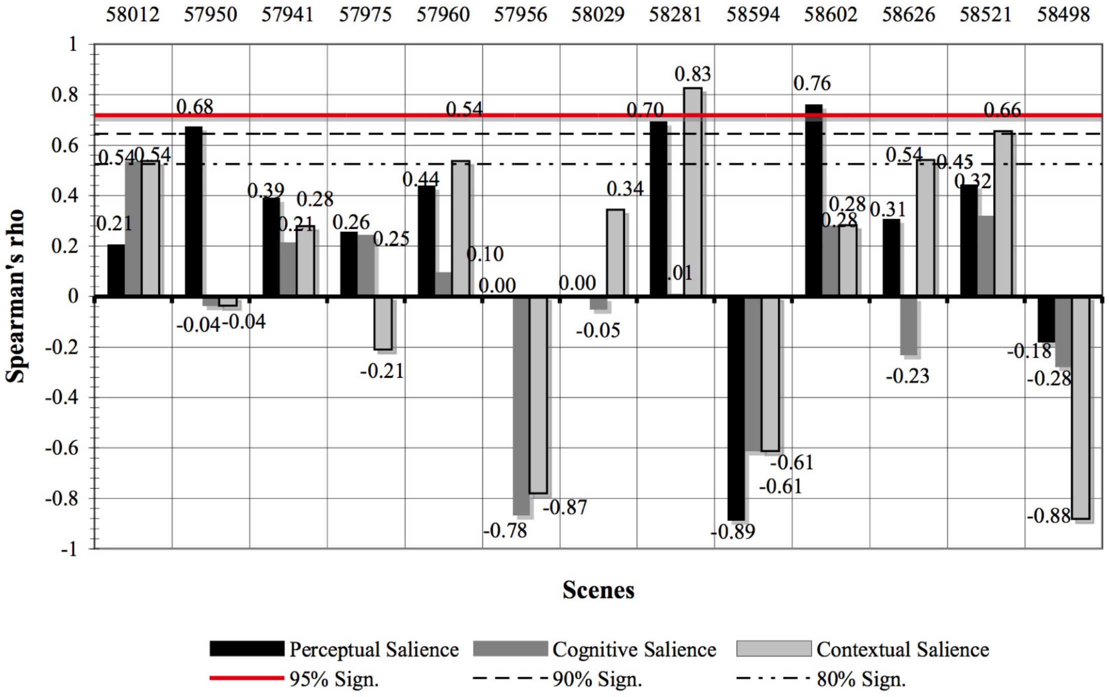


Figure 43 The figure shows Spearman's Correlation Coefficients for scenario 4.

For scenario 4, we see that only one of 13 correlation coefficients for the ranking from the survey and the ranking based on probability of perceptual salience is statistically significant at the 95% level (i.e., scene 58602). Two correlation coefficients are significant at the 80% level (scene 57950 and scene 58281). Two of the 13 correlations coefficients are negative, whereby one is statistically significant at the lower 95% level. For the coefficient for correlation of cognitive salience and the benchmark, none of the correlations are significant at the upper 95% and 90% level, while one is significant at the lower 95% level and one at the lower 90% level (i.e., scenes 57956 and 58594). One coefficient is significant at the upper 80% level, while the other coefficients indicate a lower and less reliable correlation.

The correlation coefficients for the rankings based on the survey and probability of contextual salience have similar characteristics as the coefficients for the correlation based on the ranking for perceptual salience. One correlation is significant at the upper 95% level (scene 58281), one at the upper 90% level (scene 58521), and three at the 80% level (i.e., scenes 58012, 57960, and 58626). Two correlations are significant at the lower 95% level, (i.e., scenes 57956 and 58594), and one at the lower 90% level. The remaining correlations are predominantly positive (3 of 5), but their coefficients are lower and less reliable.

4.3.4 Summary of Validation

The validation of the prototype consisted of a comparison between rankings of saliency generated by simulating four different navigation scenarios and their counterparts derived from ratings of saliency collected by means of an online-survey. A major observation at this point is that the ratings from the survey vary highly among participants. Nevertheless, the following comparison in terms of a non-parametric correlation test (i.e., Spearman) yielded mainly positive correlations. Even though only a few of these coefficients are statistically significant, this suggests that the assessment of salience based on the trilateral relationship between observer, environment, and observed object produces rankings of salience that approximate the rankings produced by humans. However, we also found low and negative correlations for all four scenarios, which we will analyze and discuss in detail in the following chapter.

Summary of Evaluationr.

4.4 Summary of Evaluation

In this chapter we evaluated the prototype implementation of the computational model. The evaluation was divided into two separate parts, one part dealing with the verification of the implementation, and the second concerned with the validation of the framework. Dividing the evaluation into verification and validation is a classic approach and has the goal to provide answers to two main questions. The first question is related to the correct implementation of the model and the second is concerned with the correctness of the model itself, that is, how well does the model represent reality.

In order to verify the implementation of the computational model, we first investigated a set of general properties and aspects on their correctness. After successful verification of these requirements, we thoroughly analyzed the integrated assessment of salience on the basis of test cases. For that purpose, we compiled a data set that represents a part of the region around the

Main Train Station in Zurich and defined a set of test cases to be used in the verification process. The verification showed that the results generated by the prototype implementation complied with the specification. Finally, we concluded the verification process with a summary and short discussion of the results.

The second part of the evaluation was concerned with the validation of the computational model. It consisted of comparing results generated on the basis of simulation scenarios with the results from an online survey. The online survey was designed using the same data set as for the second part of the verification and provided the benchmark against which the model was tested. The tests were performed for the four scenarios and showed a wide range of correlation coefficients with a majority of positive correlations, which is a first, but weak, indication of the correctness of the model. The validation results were presented in terms of correlation coefficient for the four scenarios. In the next chapter, we critically discuss and interpret the results presented in this chapter.

Chapter 5

Discussion and Conclusions

In the previous chapter, we evaluated the framework and the computational model of salience assessment in terms of its correctness with respect to the specified properties, and its performance compared to human judgment. In this chapter, we will critically analyze and discuss performance and limitations of the theoretical framework and computational model. The chapter is organized in three main sections. In the first section, we will discuss limitations of the framework and investigate restrictions of the proposed computational model. Section two describes experiences from the implementation, including issues with data collection, measures of similarity for object properties and scene configuration, and extensibility and adaptability of the model. Section three, finally, deals with the interpretation of the results of the evaluation, which involves a thorough investigation of the relation between scene setup and the model's performance, and a critical review of the validity of the results. The chapter closes with a summary of the discussion and the conclusions.

5.1 Discussion

This work attempts to combine findings from various fields of research into a single framework and computational model. As such, it draws from converging evidence from spatial cognition, neuroscience, information science, computer science, and other related fields. Interdisciplinary research projects are inherently complex and enclose a multitude of challenges. Abstraction is a key aspect in tackling these challenges. Abstraction, however, comes at the cost of clarity and accuracy of the results. In the following section, we will critically discuss the limitations of the framework, the experiences drawn from the implementation, and the interpretation of the results.

5.1.1 Limitations of the Framework

There are a number of assets to the proposed framework, but there are also limitations. First of all, the role of landmarks in spatial orientation or representation of spatial knowledge is often that of a point of reference used to describe spatial cues associated with a location, target object, or behavioral contingency. The use of landmarks in this work is restricted to that of a spatial reference point for human navigation. As a result, the relation of reference point to non-reference point is limited to paths, or to be exact, to the path that the navigator will take next.

The consideration of the different roles landmarks may assume is related to the scope of this work, which we will critically analyze in the first part of this discussion. Two more serious limitations, however, relate to the appearance and emergence of landmarks. Appearance, in this context, is understood as the stimuli and empirical knowledge that characterize landmarks, while emergence relates to the process of complex pattern formation from simpler entities. In the following sections, we describe the limitations of the framework with respect to these two concepts.

5.1.1.1 Scope

Using landmarks as points of reference or as pivotal elements in making decisions implies that these objects are salient enough for humans to direct their attention towards them in a specific context. Results from research in human information processing and theories of attention suggest that there are various factors that influence where humans direct their attention. The nature of these factors is exogenous, endogenous, or contextual. Our framework draws from these results as they form the base for the definition of the specific types of salience. The definition of the factors that define the salience of landmarks, however, is tailored to navigation tasks specifically. Hence, there is no claim that the set of components that make up the total salience is comprehensive. It is rather a collection of the most prominent characteristics of landmarks found in literature. The model can be extended to include further components of either type, be it perceptual, cognitive, or contextual.

According to Golledge (1999), the role of landmarks can be characterized as either organizing concept, or as navigational aid. Landmarks emerging as organizing concepts requires a process called cognitive mapping, which leads to a superior structure that is often referred to as the cognitive map (Miller 1956; Kuipers 1982; Golledge 1999) or cognitive collage (Tversky 1993). Within this structure, the role of the landmark changes dramatically, as it is no longer just a navigational aid, but assumes an important role in the organization of the cognitive map.

Although the computational model implements a crude form of previous knowledge (i.e., observations to objects), we do not claim to model such a cognitive map in any sense.

The previous sections describe a comprehensive framework for the assessment of the salience of potential landmarks specifically for wayfinding tasks. The framework is based on the trilateral relationship between observer, environment, and potential landmarks, and accounts for three different types of salience, namely 1) Perceptual Salience, 2) Cognitive Salience, and 3) Contextual Salience. The framework is comprehensive in the sense that it integrates these three types of salience in the context of wayfinding in order to achieve a solid assessment of which objects navigators may refer to as landmarks when standing at specific decision points along a route. The framework treats landmarks as navigational aid, rather than as an organizing concept.

5.1.1.2 Appearance of Landmarks

The proposed framework is limited to visual perception of urban objects in the form of visual observations. Limiting the framework to visual stimuli, however, is a stark abstraction of the human processing of sensory input and does not do justice to reality by any means. It is an abstraction and reduction of both, the range and sensitivity of sensory input, as well as the expressiveness and richness of experiences associated with single objects. Even though the full extent and significance of multi-sensory input for assessing the salience of objects is not fully understood yet, it is clear that visual perception per se does not completely explain salience.

The second major limitation of the framework is in terms of the appearance of landmarks. Appearances, according to Kant (Kant 1968 [1781]), are “empirical objects, and are the objects of empirical knowledge and the objects of experience”. The implication of this statement in the context of landmark salience is that appearance is more than mere perception of objects. It includes the appreciation of a variety of subjective aspects, such as cultural or historical importance, personal significance, or even activities and functions associated with spatial objects. Furthermore, investigations on visual scene understanding revealed that in real-world scenes an object’s semantic plausibility within the context of the scene is coded prior to its fixation and affects that objects saliency as an attentional target (De Graef, Lauwereyns et al. 2000). All these components contribute to the appearance of spatial objects and thus to the quality of landmarks.

The empirical knowledge based on visual perception modeled in the framework is coarse and basic compared to the sophistication and elegance of the system underlying the human assessment of salience. Nevertheless, and despite the limitations discussed above, we want to point out that the framework proposes a first approach to an integrated computational assessment of landmark salience. It was defined with extensibility and adaptability in mind. Future research may enhance and refine the framework in many aspects, as for instance the incorporation of additional sensory input or sources of knowledge, the refinement of memory, as well as the assessment process itself.

5.1.1.3 Emergence of Landmarks

Emergence is the process of complex pattern formation from simpler rules (Ghiselli-Crippa and Munro 1994). In the context of landmark salience, this can be a dynamic process that occurs over time, such as the combination of personal activities and places that may change the semantics of spatial objects, or emergence can happen over disparate size scales, such as the interactions between districts and the buildings that make up the districts, which may both be landmarks. This issue is related to the identification of spatial objects, as interpreted by Spelke (1990). While from some perspective a specific object may perhaps appear as the dominant spatial feature, it will amalgamate with other objects from another perspective.

Formally speaking, emergence refers to a computation or phenomenon at the macro-level that was not hard-coded at the micro-level. This process is, at the core, a perceptual cycle, as described by Neisser (1976), where the significance of geographic objects increases as they are repeatedly used. At the moment of perception our minds grasp and interpret sensory information, and supply us with prepackaged concepts that have specific associations and emotional tones based on past experience. As a result, the concept that is selected and supplied at any moment depends on the flow of stimuli. The flow of stimuli, in turn, is given by sensory input, environment, and context, which are the core components of our approach.

Emergence of landmarks is part of the complexity of the assessment process, but is not implemented in the conceptual framework or in the computational model at this point. Nevertheless, the framework may also be extended to account for the emergence of landmarks. A possible extension of the proposed model, hence, would be the definition of a computational model similar to that proposed by Neisser (1976) that is able to deal with the role of landmarks in the organization of spatial memory and to use this organization to infer new landmarks.

5.1.2 Experiences from the Implementation

In this section, we will discuss experiences made during the implementation of the computational model and the lessons that can be learned from them. Specifically, we will discuss issues related to the collection of data, the assessment of similarity of objects in the scenes, the quantification of the scene configuration, and finally, the usability of the proposed framework and computational model for future research.

5.1.2.1 Data Model and Data Collection

The purpose of data models is to identify and formally organize the data required for performing computational tasks. In our case, the data model can be thought of as a formal representation of the spatial environment that provides the attributes of data element (i.e., spatial objects), as well as the relationship between these elements. As such, the data model supports the understanding of the inherent dynamics of the environment, as well as the interactions between navigator and environment. It is clear that the quality of the data model influences the inferences by the computational model, and hence, special attention should be paid to accuracy and comprehensiveness of the data model.

In our computational model, we use a data model that is based on the seminal work by Lynch (1960). The main reasons for using Lynch's elements was the solid empirical foundation and the thorough description of the attributes of each element. The implementation has shown that the elements are well suited for describing the urban environment, even for the task of assessing the salience of urban objects. Nevertheless, there are three major issues that need to be addressed in this context.

The first issue with the use of Lynch's elements for describing the urban environment is that they represent the highest-possible abstraction of urban environments, which, in some cases is not a desirable property. For instance, consider the case of a prominent item on the face of a building, as is often the case with illuminated advertising. Although there are plenty of such items in urban environments, and some of them are very salient, these items do not exactly correspond to any of Lynch's elements. We worked around this issue by dividing Lynch's landmark into buildings and items. However, for future research related to the integrated assessment of landmarks salience, we suggest the use of a detailed ontology of urban environments.

The second issue with the use of Lynch's elements as data model for assessing landmark salience is related to the emergence of landmarks. Emergence may come about by combining several distinct objects into one object of higher abstraction that represents this group. Consider the case of a set of shops, for instance, that after exploring all of them emerge as a shopping district. Because of the coarse nature of Lynch's elements, such *composite* landmarks are not supported by data models based on those elements.

The third issue with the data model is the correct determination of geometric object properties. Consider the estimation of the distance of object from the observer's position. In the proposed computational model, depth estimation is based on a single heuristic, namely the elevation of the object with respect to the lower bound of the panoramic image. This approach is error prone and considering the importance of distance estimation for both, the weighting of the visible surface of the objects, as well as its importance in the description of the scene configuration, will have to be investigated thoroughly. Future research will have to use data sets or methods of extraction that allow for more accurate estimates of scene depth and distances of objects. Possible data sets include 3D models or panoramic images in combination with depth maps, such as proposed by Torralba and Olivia (2002) or Santos (2007).

Another important aspect with respect to the data used for the assessment of landmark saliency is the collection of that data. For this work, we digitized a portion of the city of Zurich from 360° panoramic images according to a specified data model. Digitizing as the mapping function between reality and representation, however, abstracts a series of processes, including text recognition, concept recognition, and the collection of attributes of spatial objects. Consequently, it leaves abundant space for misconception and subjective interpretation of perceptual properties of objects within the scene. These deficiencies essentially disqualify digitizing as a suitable method of data collection. One of the main aspects of data collection is that it has to be collected according to consistent criteria for the full data set. There are many potential sources of suitable data, including 3D city models or data sets based on remote sensing techniques.

5.1.2.2 Modeling Human Information Processing

The proposed human information processing cycle abstracts the ease of encoding and memorizing single objects (e.g., typical objects are hard to remember while untypical objects are easy to remember) (Anderson 2003). Furthermore, selective attention controls information processing so that sensory input is perceived or remembered better in one situation than another (Schneider and Shiffrin 1977; Shiffrin and Schneider 1977). Incorporating such aspects in the framework would require extensive knowledge of the spatial scene and a mechanism for object and concept identification. Even though the current framework lacks such a mechanism, it may be integrated without affecting the general structure of the proposed information processing model.

Another aspect not considered in this framework is the influence of additional sensory input on allocation of attention. Our model is based on visual sensory input and theories of visual attention as we consider vision the most important sensory input for the discrimination of salient features for navigation. These theories do not consider cross-modal sensory influence, although research has shown that auditory objects can affect visual processing, and as a result, influence the allocation of attention (Turatto, Mazza et al. 2005), Future work will have to assess to what degree cross-modal factors influence visual processing and the results will have to be incorporated in the framework accordingly.

5.1.2.3 Identification and Similarity of Scene Objects

Salience assessment implies a judgment about similarity or dissimilarity of perceived objects. Dissimilarity, or similarity, assessments require a process of establishing the semantic proximity of two entities. Humans perform this task based on knowledge and acquired reasoning strategies, typically resulting in subjective judgments. One aspect of cognition that relies on sophisticated similarity assessment techniques is the *identification* of spatial objects. Identification refers to the process of recognizing a real-world entity and establishing its identity. This task is natural to human navigators, but far from trivial for machines.

In the proposed computational model, we simplify the process of recognizing and establishing the identity of spatial objects by using a unique identifier for every object in the environment. Based on this identifier, memorized observations to the same spatial objects are identified, and subsequently, the similarity between current and stored observations is assessed. The similarity judgment is implemented as the similarity between strings (i.e., Levenshtein distance) that represent attributes of a class instance. It is applied for the degree of recognition of individual objects only, but not for comparing the similarity between different objects.

How critical this limitation is for the assessment of landmark salience remains to be analyzed. Nevertheless, future research work will have to consider this issue in order to refine the computational assessment process. Such refinement may come in any form, but a promising approach is the use of sophisticated semantic similarity assessment techniques from the field of spatial information science. These techniques build on research from psychology and use the structure of the underlying ontology to assess the similarity judgment. Implementing these methods would complement the proposed model by enabling comparisons among objects in the scene, rather than observations to the same object.

5.1.2.4 Topology and Metric of Urban Scenes

Studies of the use of spatial relations in spatial cognition showed that topology accounts for a significant portion of the geometric properties. Metric, in contrast, assumes a complementary role as it quantitatively refines the qualitative properties established by topological relations. Clark (1973) suggested a tight correspondence between perceptual space (i.e., the space humans use to perceive things around them) and linguistic space, which is used by language to represent about the perceived space.

Subsequent research drew from Clark's findings for identifying and characterizing natural-language descriptors of spatial scenes. Talmy's (1983) seminal paper, "How Language Structures Space," establishes the link between spatial configurations and the use of natural-language predicates. This evidently strong link between spatial relations and natural language suggests that accurate description of spatial relations is crucial for reasoning about spatial scenes. Consequently, topology and metric also play an important role in assessing the salience of objects contained in the scene.

In our computational model, topology is abstracted to two spatial relations, namely adjacent or disjoint. This abstraction is a simple approximation of typical scenes of urban environments. It is unable to describe other topological relations that may be present in spatial scenes, such as *partial occlusion*, *full occlusion*, *in front of* etc. Accounting for the full set of topological relations, however, could lead to a more refined saliency assessment process, and potentially produce more reliable and accurate predictions of landmark salience.

The second aspect that is of importance for accurately describing urban scenes is the metric refinement of the spatial relations among objects. The proposed computational model extracts only two metric properties of scene description, namely the horizontal and vertical Uniqueness of Location. Due to the lack of topological relations, it was impossible to extract further metric properties. Such properties include, but are not restricted to, estimates of alongness (e.g., the distance that a street follows a river), or how much of a building is occluded by other buildings. Being able to determine the metric configuration of a spatial scene at a finer level of detail may enable to capture notions of salience, such as local or distant landmark, or to adjust the weights for the assessment of the cognitive salience of objects (e.g., museum partly occluded by inconspicuous building).

5.1.2.5 Extensibility and Adaptability

The framework was designed with adaptability and flexibility in mind. Particularly, we tailored the assessment of salience to the requirements of landmark-based route instructions. Automatically generating route instructions that are not based solely on (geo-)metric properties of the underlying network requires an evaluation of the available spatial features in the surrounding environment. This evaluation is necessary for finding suitable objects for referencing the next section of the route, as proposed by Klippel and Winter (2005), or to reassure navigators that they are still on track (Denis, Pazzaglia et al. 1999). The presented framework supports this evaluation as it allows modeling what navigators will be able to perceive when approaching points of decision along the way. It may also be extended to include random positions along the way, as required for long route segments, where reassurance that navigators are still on track is typically required.

One of the primary advantages of the computational model is that the Bayesian network architecture can easily be extended by adding new low-level components as root nodes, without the need to re-specify the connections between the nodes in the network. For example, adding additional cognitive reasoning strategies, such as historical or cultural importance of objects only requires adding the appropriate low-level component to the model and specifying its impact on the corresponding auxiliary components. The linking between the auxiliary components and the high-level components, however, remains unaffected.

In addition to straightforward extensibility, the model was also built with adaptability in mind. Consider the case of findings from attention research that suggest that object-based attention contributes stronger to perceptual salience than location-based attention. In this case, only the conditional probabilities of the nodes in the network would have to be adapted accordingly, but the links between the nodes would still remain the same. Adapting the Bayesian network to new research findings complements extensibility and provides a flexible and easy to use test bed for further research in the quantitative assessment of landmark salience.

5.1.3 Interpretation of Results

The computational model supplies the platform for assessing the feasibility and investigating the correctness of the conceptual framework. Evaluating the computational model requires assessing the model's consistency with human ratings of the saliency of urban objects. Specifically, we are interested in analyzing whether the inferential mechanism of the model is consistent with the assessment people actually use. In the following sections, we will discuss and interpret the results of the validation in the light of the limitations discussed in the previous sections and the characteristics of the ratings of the survey. The interpretation of the results is organized according to the scenarios validated in the previous chapter.

5.1.3.1 Scenario 1 – No Prior Knowledge, No Task

The validation of scenario one was the starting point of the validation. The aim was to investigate the degree of correspondence between rankings of objects based on ratings from participants without any knowledge of the region and the results from the simulation. The non-parametric rank correlation test revealed that the correspondence of the two rankings is statistically significant for scene 57960 at the 95 % level and for scene 58498 at the 90% level, while the other correlations are not significant. 66% of all correlation coefficients are positive, which is a weak, but nevertheless confirming indicator that the computational model replicates human assessment of landmark salience.

One explanation for the low correlation coefficients can be inferred from a closer look at the rankings from the survey. For example, a total of eight participants rated scene 58012 resulting in eight rankings for this scene. If we assess the agreement among participants by means of a non-parametric correlation test on the rankings, we observe a low correspondence between the single rankings. In fact, only one of the eight correlations is statistically significant at the upper 95% level. The coefficients are evenly spread across 80% of the range (i.e., lowest correlation is -0.8 and highest correlation is 0.8), which points out the difficulty participants had in assessing the visual prominence of objects in the scene.

The wide range of correlation coefficients may lead to the conclusion that for scenes with highly correlated rankings for simulation and survey, the ratings from participants are more consistent. This is only partially the case. Scene 57960, for instance, has the highest correlation coefficient for scenario 1, but the individual coefficients among the single rankings for each participants shows the same range and standard deviation as for scene 58012. If we consider the distributions of the correlation coefficients, however, we observe a majority of, although not statistically significant, positively correlated coefficients for scene 57960. For scene 58012, this is not the case. Hence, we can conclude that there is a certain degree of interaction between the performance of the participants and the resulting correlation coefficients for survey and simulation.

Another interesting aspect that explains the low number of highly correlated rankings is the fact that participants of the survey rated many objects to be equally prominent. Equal ratings for objects result in equal ranks for these objects. Equal rankings, however, are highly unlikely to be produced by the computational model. Even though the probabilities calculated by the model differ in the last digit only, the objects are still assigned different ranks. As a result, rankings from the survey may contain many objects with the same rank, while rankings from the simulation are unlikely to contain any duplicate rankings. This disproportion in the rankings inevitably leads to inaccurate correlation coefficients.

5.1.3.2 Scenario 2 – Prior Knowledge, No Task

This section discusses the validation results of scenario 2, where we investigated the correspondence of the rankings based on ratings from participants with knowledge of the test site and the rankings generated by the computational model. The results of the correlation test reveal that a majority of the correlation coefficients are positive (20 of 26), which indicates a general correspondence between the proposed computational model and the human assessment of landmark salience.

Of the 26 correlation coefficients, however, only two coefficients are statistically significant within the 90% threshold (i.e., the coefficient for scene 57950 at the 95% level and the coefficient for scene 58498 at the 90% level). A plausible interpretation for this rather low rate of highly correlated rankings is that it is due to the set of objects that was selected for validating the computational model. When the set contains objects in a scene that are clearly perceptually distinct, then the correlation coefficient tends to be higher. For scenes with a set of perceptually similar objects, in contrast, the correlation tends to be low. Scenes 57941, 58626, and 57975, for instance, contain rather distinct objects, resulting in higher correlation coefficients than for scenes 57956, 58594, and 58521, where the objects are not as clearly distinguishable.

This interpretation indicates the difficulty associated with estimating the salience of perceptually similar objects. In such cases, people tend to rely on the semantics of the objects, rather than their perceptual properties. Although the proposed computational model implements semantics only on the basis of multiple observations to single objects (i.e., degree of recognition and number of observations), it replicates this behavior for scenes 58602 and 58498. In both cases, the objects are perceptually similar, but only a subset of them has been observed before and is stored in memory. The correlation coefficients are manifest of the influence of prior knowledge, as the coefficients for the rankings based on cognitive salience and the ranking from the survey are stronger than the coefficients for rankings based on perceptual salience and the survey.

Another interesting fact is that the majority of coefficients based on perceptual salience are higher than their counterparts based on cognitive salience (7 of 13). This is an expected result for scenes with a low overlap in the sets of observed objects (e.g., scenes 58029, 58281, and 58626). In other cases, however, it might hint at a weighting problem due to the structure or conditional distributions of the Bayesian network. For example, the set of objects in scene 57941 is exactly the same as in the previous scene with the exception of a single object. Therefore, we would expect a high positive correlation for the coefficient based on cognitive salience. This, however, is not the case. A plausible reason for this behavior is that the perceptual properties of objects in this scene are rather distinct, which results in high probabilities for perceptual salience. The influence of the cognitive salience leads to a distortion of the ranking, and hence a low correlation coefficient. If this peculiarity is due to a weighting problem or to the limited cognitive abilities of the model, however, remains to be investigated.

Scene 57950 has a statistically significant correlation for cognitive salience (0.75), but at the same time a relative high negative correlation for perceptual salience (-0.53). The same pattern applies for scene 57960. A possible interpretation for this discrepancy is that the influence of cognition on perception is not consistent, but varies as a function of the strength of the cognitive components. Consider scene 57950, for instance, which represents a portion of the Central, a well-known node within the city of Zurich. The set of objects that was rated by survey participants contains six highly recognizable objects, including among others a perceptually highly salient advertising for Lindt chocolate, the bridge of the Polybahn that is used to get from the Central to the ETH or University, the building with a Starbucks restaurant and Kiosk, and the Central tram station.

According to the survey, the bridge of the Polybahn is the most important visual feature in this scene, and not the highly colorful and large chocolate advertising. This is not surprising as the majority of ratings for this scene came from participants between 20 and 30 years of age. Considering the fact that the survey was sent out predominantly to members of the university, these participants are presumably students that use the Polybahn on a regular basis, and hence, consider it to be the most important feature in this scene. Such personal preferences exert a strong influence on perception, and in turn on the assessment of visual prominence. Personal preferences as a result of activities or experiences, however, are not captured by the computational model, which explains the low correlation for perceptual salience in this scene. These considerations lead to the conclusion that cognition influences perception at different levels, which needs to be considered accordingly in the computational model.

The ratings from the survey have the same wide variance as in the previous scenario, that is, there is little agreement between the participants on what objects are most salient. We also observe the same tendency as for scenario 1, as scenes with geometrically distinct objects yield better correlation coefficients as scenes with geometrically similar objects. This does not apply, however, for scenes with many semantically similar objects, where the rankings are influenced by personal preferences of participants. As for the previous scenario, the deficiencies of the validation method (i.e., distortions due to duplicate rankings) apply for this scenario as well. Nevertheless, we observe the same general tendency as in scenario 1, as most of the correlation coefficients are positive.

5.1.3.3 Scenario 3 – No Prior Knowledge, Task

The purpose of scenario 3 was to investigate the performance of the computational model with respect to participants who are following a predefined route, but don't have any prior knowledge of the area. For scenario 3, two correlation coefficients for perceptual salience and two coefficients for cognitive salience are statistically significant at the 90% level. As for scenarios 1 and 2, the correlation coefficients are predominantly positive (14 of 18), which supports the claim that the computational model approximates human performance in the assessment of landmark salience.

A striking fact of this scenario is that only four of nine correlation coefficients for contextual salience are higher than their counterparts for perceptual salience. This is surprising because the task of identifying the path of continuation clearly act as a filter for selecting appropriate objects, which is modeled accordingly in the computational model. A possible reason for this unexpected discrepancy might again be given by the structure and the conditional probabilities of the Bayesian network. The conditional probabilities define the influence of parent nodes on the common child node. In the proposed computational model, these influences are assumed to be the same for all parent nodes, which apparently turns out to produce inaccurate results for certain cases.

Consider scenes 58594 and 58521, for instance, where a subset of buildings are apparently larger and more colorful, and hence, perceptually more salient than the rest of the objects. The resulting probability of perceptual salience of these buildings is accordingly higher compared to the other objects. The other objects, however, are all located closer to the follow-up path, and hence, get a higher contextual salience. Because the conditional probabilities of the Bayesian network are such that all parent components are considered to influence the child to the same degree, the influence of perceptual salience is higher than the influence of contextual salience. Such configurations lead to results that do not conform with the results from the survey. A refinement of the conditional probabilities of the Bayesian network may alleviate this problem and confirm this speculation.

If we analyze the ratings from the survey for scenario 3, we observe the same wide variation in ratings from single participants as in the previous two scenarios. Again, we interpret this variation as evidence for the complexity of the task. The choice of a different set of objects with easily distinguishable properties would probably have yielded better results in terms of correspondence between computational model and reality. Future research, however, will have to investigate this claim.

5.1.3.4 Scenario 4 – Prior Knowledge, Task

Scenario 4 investigates the correspondence between the simulation of route following with memorizing observations and ratings from participants of the survey with prior knowledge of the area. It is the most complex of all scenarios, as it involves both, the implications of prior knowledge, as well as the assessment of the importance of objects given a context, in our case urban navigation. The results support the general trend observed with the previous scenarios, that is, 23 of 39 correlation coefficients show a positive correlation, which indicates a weak, but still positive correspondence between model and reality. As in the previous scenarios, however, only few correlations are statistically significant. We have discussed possible reasons for the low number of high correlations in the discussion of the previous scenarios. Most of these reasons, including complexity of task and resulting variation in ratings from participants, validation method, etc, apply for this scenario as well. In the following sections, therefore, we will presume previous findings and focus on the interaction between the different types of salience for this scenario.

For 5 of 13 scenes, the correlation coefficient for perceptual salience is higher than the correlation coefficient for contextual salience. This weighting effect is again due to the biased weighting discussed in scenario 3. The current definition of the Bayesian network implements a smooth filtering (i.e., equal conditional probabilities for all parents of a node), which fails to select contextually salient objects over perceptually salient objects, even if the perceptually salient objects are irrelevant for the identification of the next route segment. This suspicion is supported by the fact that this effect is present for scene 58594 in both scenario 3 and scenario 4.

Another observation for this scenario is that for 11 of 13 scenes, the correlation for contextual salience is higher or equal to the correlation of cognitive salience. Specifically, the correlation coefficient for the rankings based on cognitive salience is particularly low compared to the coefficient for contextual salience for scenes that have no overlap in the set of objects. This result was anticipated, because the effect of context affects the salience of objects in all the scenes. This is not necessarily the case for cognitive influences, due to the fact that cognitive influence is assessed based on prior observations.

For scenes 57956, 58594, and 58498 the rank correlation test yields a high negative correlation. Besides the high variation in ratings from survey participants, which corroborates the complexity of these scenes (i.e., the scenes contain predominantly perceptually similar objects), this suggests a difficulty with the definition of the task and the estimation of the task-based influence on objects. This suspicion is supported by the fact that in all three scenes, the location of the point that identifies the path for continuation of the journey in the test data set is located close to objects that are obviously irrelevant for identifying this path. For example, in scene 57956, the pointer is located next to a building that is very close to the observer. The computational model considers this vicinity as an indicator of strong contextual dependency between pointer and building. In reality, however, the pointer indicates a target that is located far away from the building, which dramatically reduces the usefulness of the building for describing the path. Due to the lack of depth information in the current data set, the computational model is unable to extract the correct spatial distance between pointer and building, which leads to wrong conclusions about the contextual relevance of objects.

A similar problem is present in scene 57975, where we find a positive correlation for rankings based on perception and cognition, but a negative context. In this scene, the pointer that identifies the target is located close to the Limmat, which is the river that is very prominently visible in this scene. Unlike survey participants, who are able to correctly interpret information in the panoramic image of the scene, the computational model misinterprets the importance of the river, and hence, produces results that do not correspond to human judgment. The problem in this scene, however, is less related to depth information than to the semantics of individual objects for navigation. The river may be located close to the target, but its role in the structure of the scene is such that it is of little relevance for identifying the path. This suggests that there is a lack of semantic relevance of objects for describing the structure of the scene, and hence, the next rout segment, which will have to be addressed in future work.

5.1.3.5 General Observations

The most general observation in the interpretation of the results is the general tendency of the correlation coefficients. A large percentage of the total number of correlation coefficients (65 of 95) indicate a positive correlation between human performance and the results generated by the computational model. Despite the fact that most correlations are not statistically significant, this high percentage of positive coefficients still provides evidence for the relative correctness of the model, especially considering the large number of abstractions and heuristics of both data model and computational model.

Another obvious finding from the interpretation of the results is that the weighting of the single components, that is, the conditional probabilities of the Bayesian network, needs to be reconsidered and refined in order to produce more accurate results. The initial assumption that all components contribute equally to salience can clearly be rejected. This conclusion is strongly supported by the second part of the survey (i.e., assessment of salience given a task), where we observe low correlations between the human judgment and results generated by the computational model for scene with strongly differing perceptual properties. We suspect that the reason for these low correlations is due to the set of conditional probabilities and could be refined by adapting them.

In this context, another aspect of the computational model needs to be considered. The three types of salience (i.e., Perceptual, Cognitive, and Contextual) constitute a Saliency Vector that has the favorable property of supporting communication when referring to landmarks. For instance, consider the case of a tourist asking a local for directions to some destination. Typically, the local will adjust the route instructions to the tourist’s knowledge of the environment and refer primarily to prominent perceptual features instead of idiosyncratic objects. Now consider the case of the local explaining the route to another local. In this case the instructions do not only refer to perceptually salient features, but may also include references to features that both relate with subjective cultural values or personal experience. The difference in the two sets of route instructions is basically a result of the weighting of the components of the saliency vector. Our approach supports individual weighting of the single components, and hence, the production of individualized route instructions.

Even though the design of the Bayesian network is such that the saliency vector can be extracted, the overall structure of the net needs to be reconsidered. The calculation of the three types of salience as proposed in the computational model merely differs in the mutual influence of the low-level components, whereby the mutual influence is given by the structure of the

Bayesian network. Future research will have to answer the question whether computing the different types of salience in three separate steps, rather than in a single step yields better results. Instead of implicitly modeling the different types of salience in terms of the structure of the Bayesian network, the agent's configuration could be used to determine the type of salience that is computed. This approach would considerably alleviate the weighting problem and at the same time unravel the interaction between the single components in the structure of the Bayesian network.

Although the data from the survey described in the previous chapter and the simulation of the assessment process afford the most direct comparison between human and computer performance on salience assessment tasks, the comparison is not completely fair. As seen in the discussion for scenarios 1 to 4, there are several criteria that must be kept in mind when making such comparisons and assessing the degree of correspondence between results from the survey and the computational model. The main criterion is related to the method of comparison, for which we applied a non-parametric rank correlation test. The method should be able to efficiently and accurately assess the degree of correlation between the two sets of rankings, which is not always the case (same rank for multiple objects, etc.). In addition, using the Median for comparison may lead to loss of information due to aggregation of the single ratings, which is not a desirable property. A thorough investigation of possible alternatives, along with a refinement of the computational model will provide further evidence about the correctness of the computational model.

5.1.4 Validity of Results

Validity is informally defined as the degree to which a study supports the intended conclusion drawn from the results (Cronbach and Meehl 1952) and is typically subdivided in internal and external validity. Internal validity is an estimate of how much measurements are based on clean experimental techniques, so as to make clear-cut inferences about cause-consequence relations, or in our case, the correspondence of the model's performance to human judgment. The issue of external validity, in contrast, concerns the question to what extent one may safely generalize the conclusions derived from a statistical evaluation to the population outside the confines of the experimental situation. In the following sections, we will discuss the validity of the results derived from the online study in terms of both internal and external validity.

5.1.4.1 Internal Validity

The notion of internal validity refers to the degree of successful elimination of confounding variables within the study itself. In the case of our survey, one major source of confounding arises from patterns in the reasoning strategies of participants for assessing visual prominence of urban objects. It is clear that participants made assumptions that are not explicitly represented in the model when rating perceptually and cognitively similar objects. Consider scene 57956 for instance, where the set of objects includes mostly well-known objects, such as a Mc Donald's restaurant, a Sports Bar, the COOP grocery store, and an Asian Food restaurant. These objects have very similar perceptual properties (color, geometry, etc.), but are very distinct in their semantic relevance for individual participants. The ratings for the objects in this scene show low correspondence, and hence, a high variance, which implicitly reflects the characteristics of the scene, along with individual rating strategies.

Ratings of object in scenes with many well-known objects might be based on former experience with any of these objects, but just as well on personal preferences, or recent activities related to the objects. Understanding and considering these reasoning strategies is mandatory for increasing the validity of the results. Determining these strategies and considering them accordingly in the validation process, however, requires moving the experiment to a controlled environment, the use of sophisticated recording techniques, and pertinent methods of analysis. The set of potential methods, in this context, may include think-aloud techniques, refined questionnaires, or the use of eye-tracking systems.

5.1.4.2 External Validity

External validity is concerned with the aspect of generalizing conclusions to the population outside the experiment. Specifically in our case, we need to reflect on the problem of the *consensus ranking* based on the ratings from single participants and its representative strength. The problem of computing a consensus ranking of the alternatives given the individual ranking preferences of several participants is called the *rank aggregation* problem. Rank aggregation has been studied in many disciplines, most extensively in the context of social choice theory (Borda 1781; Cohen, Schapire et al. 1999). The inherent difficulty in rank aggregation is to design an aggregation method that is both regular and fair, in the sense that it reflects the overall preferences of all participants, and hence, of the outside population.

In this work, the method applied for the validation of the computational model uses the Median over the vector of rankings of objects in the scenes, whereby the rankings are based on the individual ratings from participants of the survey. This method, however, may lead to conflicts with respect to desired properties of rank aggregation methods. For instance, consider the property of the Median of selecting the ratings at the position that divides the distribution of ratings into halves. Using this rating for establishing a consensus ranking may blur or even discard ratings, which are preferred by most participants. There exist a plethora of sophisticated rank aggregation methods that attempt to overcome such limitations, including methods based on *Dictatorship*, *Democracy*, and *Positional Rank Aggregation* (Dwork, Kumar et al. 2001). Each method may perform differently for different tasks. Choosing the appropriate method would certainly increase the validity of the results, but would also require a detailed and thorough investigation into the effects and peculiarities of the single methods given our setup. This investigation, however, is beyond the scope of this thesis.

5.2 Conclusions

This work is a further step toward developing a computational model for the integrated assessment of landmark saliency for human navigation. Relying on established psychological findings about the nature and peculiarities of human assessment of landmark salience, this thesis developed a flexible and adaptable framework for assessing the salience of urban objects, along with a formalization of the framework in terms of a computational model. In this section, we will conclude the research questions, evaluate the research work, and describe its contributions.

5.2.1 Answering the Research Questions

The experiences from the implementation of the computational model and the results form the comparison with human ratings are additional proof that the assessment of landmark salience is a highly complex and challenging task. Nevertheless, during the implementation of the computational model and its evaluation, we gathered the evidence required to conclude the research questions.

Question 1*: What are the fundamental components of salience?*

In the conceptual framework, we proposed that the answer to this question be based on the trilateral relationship between observer, environment, and observed objects. This trilateral relation essentially formed the base for further investigation into the identification of contributing components. The result of this investigation revealed that research in psychology, spatial cognition, and other related fields, identified perception, cognition, and context as the three major aspects that play a distinct role in the assessment of landmark salience. These finding led to the idea of defining these aspects as sets of low-level components, whereby each low-level component captures a specific aspects of the trilateral relationship between observer, environment, and observed object.

For urban navigation, we proposed that the set of low-level components includes size, shape, color, intensity, orientation, scene topology and metric for perception, degree of recognition and familiarity for cognition, and task-based relevance and modality for context. The validation of the computational model supports the choice of low-level components, but also provides the evidence that the proposed set is not sufficient for accurate predictions of salience for scenes containing perceptually similar, but semantically distinct objects. In particular, the findings suggest that personal significance of objects and activities associated with these object need to be considered in the assessment as well.

Question 2: *How do the individual components of salience influence each other?*

In order to determine the mutual influence of the low-level components and their effect on the resulting salience, we resorted to findings from attention research. On the basis of theories of attention, we introduced a set of auxiliary components, which have the purpose to model the relationship between the low-level components. The auxiliary components represent the different types of attention (i.e., location-based and object-based attention), the global scene context, recognition and idiosyncrasy of objects, as well as the task and the cognitive resources that can be allocated. Each auxiliary component corresponds to a subsystem of the human information processing cycle that integrates a subset of low-level components on one side, and provides the input to the high-level components, namely perceptual, cognitive, and contextual salience on the other side.

For the implementation of the computational model based on the conceptual framework, we proposed to use a Bayesian network. Bayesian networks are mechanisms that follow rigid mathematical rules and at the same time are very flexible and adaptable. The proposed structure of the Bayesian network integrates all three types of components, and implicitly models their mutual influence. Initially, we assumed that the strength of the influence is equal for all components. The validation of the prototype, however, provides strong evidence that this is not the case, which leaves the research question partly unanswered. Nevertheless, the general tendency of the model motivates further research on the basis of the proposed conceptual framework and computational model.

Question 3: *Is a computational model for integrated saliency assessment feasible?*

Although the proposed computational model heavily abstracts human sensory input and merely approximates the information processing cycle, it has shown that an integrated saliency assessment is possible. The results point out several issues with the model, but they also indicate that the general tendency of the computational model is correct. Considering the multitude of sensory input involved, the diversity of strategies that can be applied, and the plethora of personal and cultural preferences that contribute to the complexity of the assessment process, a perfect replication of human performance is highly unlikely. Maybe the best answer to this question is that future research will have to focus on approximating human performance, rather than fully replicate it.

In addition to the complexity of the task, there are several questions related to the computational modeling of cognitive processes that need to be answered. Such questions include, but are not restricted to, the level of generalization exhibited by humans and computer systems, the handling of real-world complexity, scalability, flexible learning strategies, and computational performance.

Question 4: *How well does the computational model replicate saliency rankings by human subjects?*

One of the obvious shortcomings of the model is the difficulty in fully replicating the rankings of salient objects provided by participants of the online survey. The analysis as to why this is the case is challenging, given that people's rating results often vary, and sometimes, the ratings show no consensus among people at all. In addition, there does not seem to be a "clean" method of interpreting the results of the validation. The comparison of the rankings from the simulation and the survey by using the Median for aggregating the rankings does not appear to be a very good measure of the quality of the computational model. Future research will have to consider different methods for establishing reliable benchmark data.

Besides the shortcoming of the validation method and the fact that the general tendency of the results motivates further research, we can also conclude that further research is necessary in order to refine the influence of single components on the resulting salience. The model will have to be more adaptable to unnatural stimuli and include additional, especially semantic aspects, in order to further refine the prediction of landmark salience.

5.2.2 Evaluating the Research Work

In this section we will evaluate the research work in that we approve or reject the hypothesis statements (HS_1 and HS_2) put forward in the introduction. The evaluation of the computational model and the discussion in the previous sections provide the evidence required to make an assessment of whether the hypotheses are correct or not. Remember that the initial hypothesis of this thesis was formulated as:

> *"If salience of urban objects is a result of the trilateral relationship between observer, environment, and observed object, then a computational model based on this relationship approximates saliency judgments by humans."*

This general hypothesis was refined in terms of contributing components and interaction between these components, and reformulated as hypothesis statements HS_1 and HS_2, which we will evaluate in the following sections. The first hypothesis statement was formulated as:

> ***HS_1:*** *If perceptual, cognitive, and contextual aspects fully explain the trilateral relationship between observer, observed object, and environment then a computational model that integrates these components produces saliency values that approximate saliency judgments by humans.*

The evidence gathered by the comparison of the computational model and the survey supports HS_1. Although only few of the correlation coefficients for the comparison of simulation and survey are statistically significant, the discussion and interpretation of the results revealed additional supporting aspects. For scenes with simple configurations and easily distinguishable perceptual properties, we observe high correlation coefficients, which is a strong indicator for the contribution of perceptual factors. For scenes with high scene complexity and semantically similar objects, however, the comparisons with the results from the survey indicate that cognition plays a major role in the assessment process. Finally, for scenes that contain objects with easily distinguishable perceptual properties we observe fuzzy results if context is considered, which confirms the influence of context. However, the findings also suggest that the proposed set of low-level components is incomplete. In particular, the set of cognitive components needs to be reconsidered and enhanced. The findings that support HS_1 are summarized in the bulleted list below:

- A majority of positive correlations between simulation results and human judgment,
- High correlation coefficients for scenes with easily discernible objects,
- Low correlation coefficients for scenes with semantically similar objects, and finally,
- Low correlation coefficients for scenes with easily discernible objects and contextual influence.

The second hypothesis statement that we formulated in the introduction is concerned with the interaction between the components and reads as follows:

> ***HS_2:** Perceptual, cognitive, and contextual components contribute equally to landmark salience.*

The results from the evaluation provide strong evidence for the rejection of *HS_2*. The first observation that leads to this conclusion is the inaccurate weighting of contextual components. For scenes with relatively distinct objects, the influence of perceptual components is stronger than the influence of contextual components, which leads to low correlation coefficients for the comparison of simulation results and human judgment. This weighting problem is clearly due to the conditional probabilities of the Bayesian network, which are defined such that all components contribute equally to the resulting saliency values.

The second observation that supports the rejection of *HS_2* is the conclusion that scenes with semantically similar objects not only require the modeling of personal preferences and activities related to the objects, but also a sophisticated weighting strategy. As for the first observation, this weighting can be achieved by a refinement of the structure of the Bayesian network and the adjustment of the conditional probabilities. The following list summarized the findings that lead to the rejection of *HS_2*:

- Low influence of context on the salience of perceptually distinct objects, resulting in biased saliency values, and,
- Low influence of cognitive components on the salience of highly recognizable objects.

These outcomes confirm the exploratory nature of the approach and imply that the conceptual framework and the computational model need to be further refined in order to approximate human judgment of landmark salience more accurately.

5.2.3 Scientific Contribution

The scientific contribution of a computational model of salience assessment has a theoretical and a practical dimension. From a theoretical point of view, it improves the understanding of the complexity of the saliency assessment process, while from a practical point of view, it supports the research and development of more sophisticated computational models and will help in further understanding the abilities and limitations of computational models for non-trivial tasks. Theories in the field of spatial development and spatial cognition have placed heavy weight on the construct of landmarks, but so far, no computational model for the integrated assessment of the salience of landmarks existed. The main scientific contribution of this thesis, therefore, was to fill the gap between theory and practice by bridging several scientific disciplines and providing the base for further research.

The thesis proposes an attention-based approach for the identification of contributing components and a framework for the integrated assessment of landmark salience. On the basis of this framework, a computational model was developed and implemented. The computational model of integrated landmark saliency assessment has the potential to produce saliency values that cannot be accomplished by any combination of the individual components of salience, and hence, has the potential to produce more information than models that do not integrate these concepts.

The prototype implementation may be considered a testing engine for hypothetical models of integrated salience assessment that is easily adaptable and extensible. A first set of hypotheses was proposed as part of the conceptual framework and tested within the scope of the thesis. The validation results indicate that reliable assessment of landmark salience requires accurate models of the environment, along with sophisticated models of memory that are able to deal with human preferences, experiences, and activities associated with spatial objects. Future research might use the prototype implementation as test-bed for further investigation and evaluation of refined hypotheses for integrated assessment of landmark salience.

5.3 Summary

We have defined a framework and developed a computational model for the assessment of landmark saliency that formalizes and integrates the components people use in reasoning about important objects for human navigation. Our framework serves as a bridge between findings from spatial cognition research and practical applications. In this chapter, we pointed out limitations of the conceptual framework, discussed the experiences from the implementation of the computational model, and interpreted the results of the evaluation of the computational model. The findings suggest two major aspects that need to be considered in future research. The first relates to a systematic weighting issue of low-level components due to the structure of the Bayesian network, and the second aspect is the indication of the results that the model lacks explanatory power due to the limited number of low-level components, in particular for cognitive components. Finally, the chapter evaluated the hypotheses and concluded with an account of the scientific contribution of the research work.

Chapter 6
Summary and Outlook

On the basis of the vast research on the nature and use of landmarks for human navigation, this thesis developed a conceptual framework for the integrated assessment of the salience of spatial objects in urban environments. The framework essentially describes an ontology of the components that are required for the saliency assessment. The formalization of the conceptual framework in terms of a computational model provides the test-bed for the evaluation of the developed concepts and methods. This chapter reviews objectives, methods, and results of this thesis, and discusses possible future research. The first section summarizes the major topics and gives an overview of the research of this thesis. It is structured according to the research question formulated in the introduction. In the second section, we present the major results before concluding the thesis with a discussion of potential directions for future research.

6.1 Summary

Landmarks are conceivably the most fundamental pieces of spatial information. People use them for a plethora of tasks related to description, understanding of, and reasoning about our physical environment. Landmarks, however, come in many shapes and forms, and estimates about the quality of landmarks are intuitive, qualitative, and subjective. Intuition, qualitative reasoning, and subjective judgment, however, are human assets, but not comprised in the abilities of machines, including information system and other formal mechanism. Assessing the quality of landmarks, consequently, is a highly challenging task, and requires elaborate strategies and sophisticated mechanisms, in addition to accurate representations of the spatial environment.

In this work, we attempt to approximate the assessment process by identifying the contributing components of salience and investigating the mechanisms that underlie human judgment of landmarks. We base this analysis on the assumption that salience is not a property of objects per se, but is the result of the trilateral relation between observer, environment, and observed object. This assumption is at the core of this work and the hypothesis. The premise that this trilateral relationship dictates what objects are considered good landmarks inevitably leads to the conclusion that perception, cognition, and context define the starting point of the investigation into the nature of salience.

Perception, cognition, and context, in our work, are defined as a set of low-level components of salience. We propose to exploit theories of attention and human information processing in order to identify the individual components and the complex mechanisms that glue them together. Visual attention research revealed two different types of attention, which are categorized according to the basic unit of interest. In the case of location-based attention, this basic unit is any location in the visual field that contrast with the surrounding locations, while the basic unit of object-based attention is assumed to be any entity of interest recognized by the observer. The major distinction between the two approaches is postulated to be the degree of involvement of cognitive mechanism. While location-based attention is presumed to be purely bottom-up or perceptual, object-based attention is also affected by top-down or cognitive properties.

In addition to location-based and object-based attention, research has also shown that the allocation of attention is influenced by the configuration of the spatial scene and the context. Drawing from these findings, we propose that the conceptual model includes the low-level components of color, texture orientation, and intensity for location-based attention, object size and shape for object-based attention, topology and metric for the scene configuration, degree of recognition and idiosyncrasy as top-down properties, and finally, task and modality for contextual influence. These 11 low-level components constitute the attributes, from which we will derive the salience of urban objects.

The next step in the development of the conceptual model is the integration of these components into a single mechanism. We propose to complement the low-level components with a set of auxiliary components that defines the interdependence between them. In addition, we propose the definition of three types of high-level components, which correspond to different types of salience, namely perceptual salience, cognitive salience, and contextual salience. The high-level components constitute a vector that represent the overall salience of an object and has the desirable property of supporting the communication of landmarks between persons with different levels of prior knowledge of a specific environment.

The last step in the development of the conceptual model was to identify the structure and define the degree of mutual influence of the components. For this, we proposed the use of a Bayesian network, which is essentially an adaptable and extensible probabilistic mechanism for modeling causality. The structure of the network was again based on findings from attention research. For the mutual influence of the components, however, we had to resort to the initial assumption that each component contributes equally to salience. This assumption defines a hypothesis, which will be challenged by the evaluation process.

On the basis of the conceptual model, we developed and implemented a computational model, which is designed according to the paradigm of agent-based simulation. This approach enables using the computational model as a hypothesis-testing engine, which is beneficial given the complexity of the task. In order to evaluate the computational model, we verified its correctness and validated its performance with respect to a benchmark data set. The benchmark data was collected by means of an online survey and the validation method consisted of a non-parametric rank correlation test (Spearman) between the rankings produced by the computational model and the ratings from the survey. The comparison included four different scenarios, each tailored such as to extract specific information requires for the evaluation of the hypotheses.

6.2 Major Results

The evaluation of the computational model by means of the four scenarios postulated in the previous section produces mainly positive correlations. Even though only a few of these coefficients are statistically significant, this suggests that the assessment of salience based on the trilateral relationship between observer, environment, and observed object produces rankings of salience that approximate the rankings produced by humans. This conclusion leads to the confirmation of hypothesis statement 1, which states that the trilateral relationship between observer, environment, and observed object is fully explained by the integration of perceptual, cognitive, and contextual components.

Although we found support for hypothesis statement 1, however, the findings also show that the set of components proposed in the framework is not sufficient for accurate predictions of salience for complex scenes. Specifically, we found low correlations for scenes that contain semantically distinct, but perceptually similar objects. This low correlation is manifest of the missing abilities of the computational model in terms of cognitive skills, presumably essentially in terms of personal preferences and experience related to specific objects. The importance of cognitive influence is also visible in the results from the survey, where we observe, especially for participants with prior knowledge, a tendency for selecting meaningful objects as most salient.

In addition, the results also show that the interaction between the components of salience in the integrated saliency assessment varies with the content of the scene. That is, cognitive aspects contribute stronger to salience if the objects are perceptually similar. This claim is further supported by the observation that for scenes with perceptually distinct objects, the influence of perceptual components is such that the contextual influence is marginalized, resulting in biased rankings, and hence, low correlation coefficients. These findings lead to the rejection of hypothesis statement 2, which postulated that the low-level components contribute equally to salience.

The experiences from the implementation of the concepts and strategies developed with the conceptual model show that (1) a computational model of integrated saliency assessment for human navigation in urban environments is feasible, (2) the approach based on the trilateral relationship produces reasonably good approximations of salience values for simple scenes, and (3) that the framework needs to be further refined, especially in terms of cognitive capabilities and integration of components, in order to produce better results for complex scene configurations.

The online survey about real-world judgment of salience provided the benchmark data set for the evaluation of the computational model. It also provided evidence for the complexity inherent in the assessment of salience. Specifically, judgments of object's salience showed a high variation for scenes with semantically similar objects, which suggests that there is no consensus among participants on a single rating strategy. Rather, it confirms the results from previous research that assigns a prominent role to cognitive aspects. These observations and findings from the survey are relevant, because they provide the basis for the revision and refinement of the proposed framework and computational model. The refinement of the computational model is related to the major result of the work from a practical point of view. The prototype implementation of the computational model was developed with adaptability and extensibility in mind, which enables to use it as a hypothesis-testing engine for future research.

6.3 Future Research

This research sought to address the quantitative assessment of landmark salience, which is a highly complex task with many challenging questions. For this purpose, we developed a conceptual framework along with a computational model that implements the concepts and ideas of the framework. The primary aim of the computational model was to provide an adaptable and extensible engine for testing hypothesis related to the mechanisms of salience assessment. The results and findings of the evaluation of the framework and the experiences from the implementation have left some questions only partially answered and inspired several new ones.

In the following sections, we point out some of these questions and describe further research that might be conducted to clarify them.

6.3.1 Extension of Conceptual Framework

The findings have shown that the proposed conceptual model has the potential to approximate human judgments about the salience of spatial objects for navigation in urban environments. The findings have also shown, however, that accurately approximating such judgments requires the extraction and modeling of additional factors of salience. In the following sections, therefore, we describe a set of extensions to the conceptual framework that may be considered in future research.

6.3.1.1 Integration of Multi-sensory Input

We base our framework on the initial assumption that appearance of landmarks is strictly visual. While this assumption may apply for a large part of the population, it certainly is not the case for all groups of people. Overcoming this lack of visual abilities implies a shift of strategies for spatial orientation. For the blind, for instance, orientation to the environment occurs when an individual has achieved awareness of self-to-object relationships, and object-to-object relationships (Golledge 1999). Once oriented, the individual is able to locate both nearby and distant objects. An individual gains such orientation by using search strategies to establish relationships, whereby blind persons may use a variety of strategies for searching unfamiliar space.

Incorporating such strategies in the conceptual framework is a necessity if we are to extend the current scope of the conceptual framework. Incorporating such strategies, however, requires the consideration of additional sensory input, such as sound and motion. As pointed out in the conception of the conceptual framework, the idea was to develop the framework with adaptation and flexibility in mind. Therefore, future work that attempts to incorporate additional sensory input is well served with the current framework as base for further research.

6.3.1.2 Considering Cultural and Personal Significance

The focus in this work is on the physical appearance of landmarks, rather than their cultural or personal significance, function, etc. The results of this work show that this approach is not sufficient for fully explaining salience of urban objects. Specifically, the results suggest that, besides the degree of recognition and familiarity, additional semantic components need to be considered. Therefore, a second major extension of the conceptual model relates to the incorporation and integration of these aspects into the assessment process.

Cultural and personal significance of objects is the result of experiences, activities, and facts associated with these objects. The degree of cultural and personal significance varies with characteristics of individuals (e.g., age, preferences, knowledge) and geographic regions (e.g. local, regional, interregional meaning and significance), and consequently, is accordingly difficult to model. A possible extension to the proposed approach is to model cultural and personal importance in the conceptual framework and implement the computational model as a multi-agent simulation. Multi-agent simulations enable to dictate each agent's behavior by a set of individual characteristics (preferences, activities, etc.) and to draws conclusion from the corporate behavior of agent groups.

Another possible approach for enhanced consideration of the cognitive abilities is to define the characteristics of cultural and personal significance of objects on a conceptual level, and to use location-based queries for initializing the agent preferences and prior knowledge. The location-based queries may be restricted to the representation of the urban environment, or extended to include the World Wide Web, similarly to the methods and techniques used by spatialized search engines.

6.3.1.3 Adaptation to additional Types of Space

The context of this work was defined as the determination of landmark salience for human navigation. As a result, the attributes of the conceptual model are tailored specifically to suit the peculiarities and characteristics of urban space. Navigation, however, is an activity that does not only take place in urban space, but includes all possible types of space, including concrete types of space, such as rural space or interior of buildings, moderately abstract spaces, such as map space or mobile space, but also highly abstract types of space, such as the World Wide Web or large data repositories. Although the present framework is tailored specifically to navigation in urban space, many of the aspects of salience described in this work apply also to the other types of space. Therefore, a third possible extension to the conceptual model is the adaptation the additional types of space. In the following section, we will describe two examples of potential extensions of the conceptual model.

The closest extension to urban space is the adaptation of the model to rural space. Urban space and rural space share some commonalities (i.e., physically navigable geographic space), but are also distinct in several aspects, including scale, set of objects, and higher-level concepts (districts vs. mountains, etc.). Extending the proposed conceptual framework to include rural space implies the use of a different data model and appropriate methods for the extraction of the low-level components of salience. The overall structure of the conceptual framework accounts for most of the perceptual, cognitive, and contextual properties encountered during navigation in rural space. Therefore, the extension to rural spaces promises to be a fruitful area for future research.

Another conceptual extension of the framework is related to the navigation of maps in terms of understandability and usability. The design of maps has long been considered an art. Recent research effort in the context of digital maps, however, attempts to demystify the artistic nature of maps and to define formal rules for the computational generation of such maps. The focus of research is on perceptual and cognitive properties of map understanding, which is related to the focus of this work. The trilateral relationship that forms the base of this work would need to be rephrased as the trilateral relationship between observer, map, and observed entity on the map. As with the previous example of future work, however, many of the concepts and methods developed in this work are applicable and potentially beneficial for identifying principles of visualization and map creation.

6.3.2 Extension of Computational Model

The first set of options for future research addressed possible extensions on a conceptual level. In this section, we will describe potential enhancements on the computational level. The experiences from the implementation and the results of the comparison with human judgment have shown that such enhancements are not only desirable, but also necessary in order to increase the quality of the salience estimation. In the following sections, we provide a short overview of possible extensions of the computational model in the sequence of their urgency.

6.3.2.1 Refinement of Assessment Mechanism

Besides the weakness on a conceptual level, the results of the validation have also shown that the computational model needs to be refined in order to produce better predictions of salience. Specifically, we found that the low-level components do not contribute equally to salience, but according to their pertinence for specific aspects. For instance, we found that perceptual components were incorrectly weighted stronger than contextual components. This finding led to the rejection of hypothesis statement 2 and to the question of the calibration of the Bayesian network.

In order to calibrate the Bayesian network, future research will have to further analyze the interaction between the different low-level components. Ideally, the model would need to have the ability to adapt to different contextual and perceptual settings. This adaptation could be achieved by using the proposed model as hypothesis-testing engine, in combination with, for instance, eye-tracking experiments, for a sensitivity analysis.

6.3.2.2 Data Model and Extraction of Object Properties

Another refinement of the computational model is in terms of the applied data model and the extraction of object properties from that model. For the purpose of this work, we used a data model based on Lynch's elements imageability of the city, along with panoramic images and the digitized representations of the objects within these images. Using different data models, as for instance 3D city model or models based on Rapid Mapping techniques, would discard the subjectivity of the current approach and certainly increase the reliability and accuracy of the extracted properties. In order to bridge the current gap, further work is required, especially form an engineering point of view.

6.3.2.3 Offline Assessment of Landmark Salience

A very interesting line of future research is related to the a posteriori assessment of landmark salience. A posteriori, in this context, is understood as the assessment of salience from memory without direct perceptual stimuli, as is the case for route planning, that is, for the selection of appropriate objects for route instructions. As our approach is based on the situated nature of navigation, the current implementation of the computational model does not support this offline version of the assessment process. The implications of this restriction are twofold. The first is of conceptual nature in that the assessment of salience is restricted to single scenes and does not scale to whole environments, while the second is of technical nature, that is, the architecture of the computational model is required to conform the paradigm of agent-based simulation. Overcoming this restriction, consequently, requires additional research work in terms of conceptual and computational solutions.

6.3.2.4 Generation of Route Instructions

The initial idea and motivation for this work was to provide the means for the integration of landmarks in the route generation process. That is, for navigation, travelers need route directions, which are preferably expressed as a sequence of instructions, as for instance, "face towards the tower" and "move along the river". Such instructions typically rely on qualitative references to landmarks, instead of quantitative descriptions. The initial problem in this context is the assessment of landmark salience, for which this thesis provided a potential solution.

The second problem in the automatic generation of route instructions is to develop a method to find routes in a network with the property that they can be described by a simple sequence of instructions. The key problems that need to be solved are (1) how to attribute landmark information to the network and (2) how to find an optimal route. An initial approach to this problem was presented by Rüetschi, et. al. (2006). Combining the findings of this thesis with the approach for the incorporation of landmarks in the route generation process seems to be a promising approach for the realization of the idea that sparkled this research.

Bibliography

Abdi, H. (2007). Distance. Encyclopedia of Measurement and Statistics. N. J. Salkind. Thousand Oaks, CA, Sage.

Aivar, M. P., M. M. Hayhoe, et al. (2005). "Spatial memory and saccadic targeting in a natural task." Journal of Vision 5(3): 177-193.

Allen, G. L. (1997). From Knowledge to Words to Wayfinding: Issues in the Prediction and Comprehension of Route Directions. Spatial Information Theory: A Theoretical Basis for GIS, International Conference COSIT '97, Laurel Highlands, Pennsylvania, USA, Springer.

Anderson, M. L. (2003). "Embodied Cognition: A field guide." Artificial Intelligence 149: 91-130.

Appleyard, D. (1969). "Why Buildings Are Known - Predictive Tool for Architects and Planners." Environment and Behavior 1(2): 131-156.

Bauer, B., P. Jolicoeur, et al. (1996). "Visual search for colour targets that are or are not linearly-separable from distractors." Vision Research(36): 1439–1446.

Beck, J., K. Prazdny, et al. (1983). A theory of textural segmentation. Human and Machine Vision. K. P. J. Beck and A. Rosenfeld. New York, New York, Academic Press: 1-39.

Biederman, I. (1972). "Perceiving Real-World Scenes." Science 177(4043): 77-80.

Borda, J. C. (1781). "Memoire sur les elections au scrutin." Histoire de l'Academie Royale des Sciences.

Burnett, G. (2000). "“Turn right at the Traffic Lights”: The Requirement for Landmarks in Vehicle Navigation Systems." The Journal of Navigation 53(3): 499-510.

Busquets, D., C. Sierra, et al. (2002). "A Multi-Agent Approach to Fuzzy Landmark-Based Navigation." Journal of Multi-Valued Logic and Soft Computing 9: 195-220.

Busquets, D., C. Sierra, et al. (2003). "A Multi-agent Approach to Qualitative Landmark-Based Navigation." Autonomous Robots 15: 129-154.

Caduff, D. and S. Timpf (2005a). The Landmark Spider: Representing Landmark Knowledge for Wayfinding Tasks. AAAI 2005 Spring Symposium, Stanford, CA, AAAI Press.

Caduff, D. and S. Timpf (2005b). The Landmark Spider: Weaving the Landmark Web. STRC'05 - 5th Swiss Transport Research Conference, Monte Verità, Switzerland, ETH.

Chater, N., J. B. Tenenbaum, et al. (2006). "Probabilistic models of cognition: Conceptual foundations." Trends in Cognitive Sciences 10(7): 287-291.

Chung, D., R. Hirata, et al. (2002). A new robotics platform for neuromorphic vision: Beobots. Lecture Notes in Computer Science. Berlin, Germany, Springer. 2525: 558-566.

Clark, H. (1973). Space, Time, Semantics, and the Child. Cognitive Development and the Acquisition of Language. T. Moore. New York, NY, Academic Press: 27-63.

Cohen, W. W., R. E. Schapire, et al. (1999). " Learning to order things." Journal of Artificial Intelligence Research 10: 243-270.

Couclelis, H., R. G. Golledge, et al. (1995). Exploring the anchor-point hypothesis of spatial cognition. Urban Cognition. T. Gärling. London, Academic Press: 37-60.

Cronbach, L. J. and P. E. Meehl (1952). "Construct Validity in Psychological Tests." Psychological Bulletin: 281-302.

Crundall, D., G. Underwood, et al. (1999). "Driving experience and the functional field of view." Perception 28(9): 1075-1087.

D'Zmura, M. (1991). "Color in visual search." Vision Research 31(6): 951–966.

Daniel, M. P. and M. Denis (2004). "The production of route directions: Investigating conditions that favour conciseness in spatial discourse." Applied Cognitive Psychology 18(1): 57-75.

De Graef, P., J. Lauwereyns, et al. (2000). Attentional Orienting and Scene Semantics. Psychological Reports, Laboratory of Experimental Psychology, University of Leuven, Belgium: 22.

Denis, M., F. Pazzaglia, et al. (1999). "Spatial Discourse and Navigation: An Analysis of Route Directions in the City of Venice." Applied Cognitive Science 13(2): 145-174.

Downs, R. M. and D. Stea (1977). Maps in Minds.

Dwork, C., R. Kumar, et al. (2001). Rank aggregation methods for the web. Proceedings of the tenth international conference on World Wide Web, Hong Kong.

Elias, B. (2003). Determination of Landmarks and Reliability Criteria for Landmarks. Fifth Workshop on Progress in Automated Map Generalization, IGN, Paris, ICA Commission on Map Generalization.

Elias, B. (2003). Extracting Landmarks with Data Mining Methods. International Conference on Spatial Information Theory, COSIT 2003, Kartause Ittingen, Switzerland, Springer-Verlag.

Eriksen, C. W. and J. D. St James (1986). "Visual attention within and around the field of focal attention: A zoom lens model." Perception & Psychophysics 40(4): 225–240.

Eriksen, C. W. and Y. Y. Yeh (1985). "Allocation of attention in the visual field." Experimental Psychology: Human Perception and Performance 11(5): 583-597.

Escrig, M. T. and F. Toledo (2000). "Autonomous robot navigation using human spatial concepts." International Journal of Intelligent Systems 15(3): 165-196.

Fontaine, S. and M. Denis (1999). The Production of Route Instructions in Underground and Urban Environments. Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science, International Conference COSIT '99, Stade, Germany, Springer.

Funes, M. J., J. Lupianez, et al. (2005). "The Role of Spatial Attention and other Processes on the Magnitude and Time Course of Cueing Effects." Cognitive Processing(6): 98-116.

Gaerling, T. (1999). Human Information Processing in Sequential Spatial Choice. Wayfinding Behavior: Cognitive Mapping and other Spatial Processes. R. G. Golledge. Baltimore, John Hopkins University Press: 81-98.

Gaerling, T., A. Boeoek, et al. (1986). "Spatial orientation and wayfinding in the designed environment: A conceptual analysis and some suggestions for postoccupancy evaluations." Journal of Architectural and Planning Research(3): 55-64.

Galler, I. (2002). Identifikation von Landmarken in 3D-Stadtmodellen. Institut für Kartographie und Geoinformation. Bonn, Rheinische Friedrich-Wilhelms-Universität Bonn: 120.

Ghiselli-Crippa, T. and P. W. Munro (1994). Emergence of global structure from local associations. Advances in Neural Information Processing Systems, San Mateo, CA, Morgan Kaufmann Publishers.

Gigerenzer, G. and D. J. Murray (1987). Cognition as Intuitive Statistics. Hillsdale, NJ, Erlbaum.

Golledge, R. G. (1991). Cognition of Physical and Built Environments. Environment, Cognition and Action: An Integrated Approach. T. Gaerling and G. W. Evans. NY, Oxford University Press: 35-62.

Golledge, R. G. (1992). "Place Recognition and Wayfinding - Making Sense of Space." Geoforum 23(2): 199-214.

Golledge, R. G. (1999). Human wayfinding and cognitive maps. Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes. R. G. Golledge, John Hopkins University Press: 5-45.

Golledge, R. G. (1999). Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes, John Hopkins University Press.

Haken, H. and J. Portugali (2003). "The face of the city is its information." Journal of Environmental Psychology 23(4): 385-408.

Hayhoe, M. M., H. Shinoda, et al. (2000). "Attention in natural environments." Investigative Ophthalmology & Visual Science 41(4): S422-S422.

Henderson, J. M. and A. Hollingworth (1999). "High-level Scene Perception." Annual Review of Psychology 50: 243-271.

Hollands, M. A., A. E. Patla, et al. (2002). ""Look where you're going!" Gaze Behaviour associated with maintaining and changing the direction of locomotion." Experimental Brain Research 143: 221-230.

Howard, I. P. and B. J. Rogers (2002). Seeing in Depth. Toronto, Canada, I. Porteous Publishing.

Itti, L. (2005). "Quantifying the Contribution of Low-Level Saliency to Human Eye Movements in Dynamic Scenes." Visual Cognition 12(6): 1093-1123.

Itti, L., C. Koch, et al. (1998). "A model of saliency-based visual attention for rapid scene analysis." Ieee Transactions on Pattern Analysis and Machine Intelligence 20(11): 1254-1259.

James, W. (1890). The principles of psychology. New York, Henry Holt & Co.

Janzen, G. and M. v. Turennout (2004). "Selective neural representation of objects relevant for navigation." Nature Neuroscience 7: 673-677.

Jensen, F. V. (2001). Bayesian Networks and Decision Graphs, Finn V. Jensen. Bayesian Networks and Decision Graphs. Springer, 2001.

Julész, B. and J. R. Bergen (1983). "Textons, the fundamental elements in preattentive vision and the perception of textures." Bell System Technical Journal 62(6): 1619-1645.

Kant, I. (1968 [1781]). Kritik der reinen Vernunft. Werke in zwölf Bänden. W. Weischede. Mineola, N.Y., Dover Publications. III u. IV.

Kawai, M., K. Uchikawa, et al. (1995). Influence of color category on visual search. Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, Florida.

Kersten, D. (2002). "Object perception: Generative image models and Bayesian inference." Biologically Motivated Computer Vision, Proceedings 2525: 207-218.

Kersten, D., P. Mamassian, et al. (2004). "Object perception as Bayesian inference." Annual Review of Psychology 55: 271-304.

Kersten, D. and a. Yuille (2003). "Bayesian models of object perception." Current Opinion in Neurobiology 13(2): 150-158.

Klippel, A. (2004). "Wayfinding choremes - conceptualizing wayfinding and route direction elements." KI 18(1): 63-64.

Klippel, A., K.-F. Richter, et al. (2005). Wayfinding Choreme Maps. Visual Information and Information Systems, 8th International Conference, VISUAL 2005, Amsterdam, The Netherlands, Springer.

Klippel, A. and S. Winter (2005). Structural Salience of Landmarks for Route Directions. Spatial Information Theory - COSIT05, Ellicottville, NY, USA, Springer-Verlag.

Koch, C. and S. Ullman (1985). "Shifts in Selective Visual-Attention - Towards the Underlying Neural Circuitry." Human Neurobiology 4(4): 219-227.

Kosmopoulos, D. I. and K. V. Chandrinos (2002). Definition and Extraction of Visual Landmarks for Indoor Robot. Methods and Applications of Artificial Intelligence: Second Hellenic Conference on AI, SETN 2002, Berlin / Heidelberg, Springer.

Kosslyn, S. M. (1989). "Understanding Charts and Graphs." Applied Cognitive Psychology 3(3): 185-226.

Kubovy, M., D. J. Cohen, et al. (1999). "Feature integration that routinely occurs without focal attention." Psychonomic Bulletin & Review 6(2): 183-203.

Kuipers, B. J. (1982). "The ``Map in the Head" metaphor." Environment and Behavior 42(2): 202-220.

Lacroix, J. P. W., J. M. J. Murre, et al. (2006). "Modeling Recognition Memory Using the Similarity Structue of Natural Input." Journal of the Cognitive Science Society 30(1): 121-145.

Lakoff, G. (1987). Women, Fire, and Dangerous Things. Chicago, IL, University of Chicago Press.

Lee, P. U., H. Tappe, et al. (2002). Acquisition of landmark knowledge from static and dynamic presentation of route maps. Twenty-fourth Annual Conference of the Cognitive Science Society.

Levenshtein, V. I. (1966). "Binary codes capable of correcting deletions, insertions, and reversals." Soviet Physics Doklady(10): 707–710.

Lewis, D. (1973). "Causality." The Journal of Philosophy(70): 556-567.

Lovelace, K. L., M. Hegarty, et al. (1999). Elements of Good Route Directions in Familiar and Unfamiliar Environments. International Conference COSIT'99, Stade, Germany, Springer Verlag.

Lynch, K. (1960). The Image of the City. Boston, The M.I.T. Press.

Marcel, A. and C. Dobel (2005). "Structured perceptual input imposes an egocentric frame of reference -- pointing, imagery, and spatial self-consciousness." Perception 34(4): 429-451.

May, A. J., T. Ross, et al. (2003a). "Drivers' information requirements when navigating in an urban environment." Journal of Navigation 56(1): 89-100.

May, A. J., T. Ross, et al. (2003b). "Pedestrian Navigation Aids: Information Requirements and Design Principles." Personal Ubiquitous Computing 7: 331-338.

Miau, F. and L. Itti (2001). "A neural model combining attentional orienting to object recognition: Preliminary explorations on the interplay between where and what."

Miller, G. A. (1956). "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information." Psychological Review 63: 81+97.

Montello, D. R. (1997). The Perception and Cognition of Environmental Distance: Direct Sources of Information. Conference on Spatial Information Theory: A theoretical basis for GIS, COSIT'97, Laurel Highlands, Pennsylvania, USA, Springer Verlag.

Montello, D. R. (2003). Navigation. Handbook of visuospatial cognition. P. Shah and A. Miyake. Cambridge, Cambridge University Press: 257-294.

Montello, D. R. and S. Freundschuh (2005). Cognition of Geographic Information. A research agenda for geographic information science. R. B. McMaster and E. L. Usery. Boca Raton, FL, CRC Press: 61-91.

Moulin, B. and D. Kettani (1999). "Route Generation and Description Using the Notion of Object's Influence Area and Spatial Conceptual Map." Spatial Cognition and Computation 1: 227-259.

Müller, N. G. and A. Kleinschmid (2003). "Dynamic Interaction of Object- and Space-Based Attention in Retinotopic Visual Areas." The Journal of Neuroscience 23(30): 9812-9816.

Nagy, A. L. and R. R. Sanchez (1990). "Critical color differences determined with a visual search task." Journal of the Optical Society of America(7): 1209–1217.

Neisser, U. (1976). Cognition and Reality. San Francisco, Freeman.

Newell, A. and H. A. Simon (1972). Human problem-solving, Prentice Hall, Englewood Cliffs.

Newman, E. L., J. B. Caplan, et al. (in press). "Learning your way around town: how virtual taxicab drivers learn to use both layout and landmark information." Cognition.

Nothegger, C. (2003). Automatic Selection of Landmarks. Institute of Geodesy and Geophysics. Vienna, University of Technology: 78.

Nothegger, C., S. Winter, et al. (2004). "Computation of the Salience of Features." Spatial Cognition and Computation 4(2): 113-136.

Olshausen, B. a., C. H. Anderson, et al. (1992). "Computer-Simulation of a Dynamic Routing Model of Visual-Attention." Investigative Ophthalmology & Visual Science 33(4): 1263-1263.

Parkhurst, D. J. and E. Niebur (2003). "Scene content selected by active vision." Spatial Vision 16(2): 125-154.

Peters, R. J., A. Iyer, et al. (2005). Components of Bottom-Up Gaze Allocation in Natural Scenes. Proc. Vision Science Society Annual Meeting (VSS05).

Posner, M. I. (1998). Foundations of Cognitive Science, MIT Press.

Presson, C. C. and D. R. Montello (1988). "Points of Reference in Spatial Cognition - Stalking the Elusive Landmark." British Journal of Developmental Psychology 6: 378-381.

Raubal, M. and S. Winter (2002). Enriching Wayfinding Instructions with Local Landmarks. Geographic Information Science. M. J. Egenhofer and D. M. Mark. Berlin, Springer. 2478: 378-381.

Richardson, A. E., D. R. Montello, et al. (1999). "Spatial knowledge acquisition from maps and from navigation in real and virtual environments." Memory & Cognition 27(4): 741-750.

Ruz, M. and J. Lupianez (2002). "A review of attentional capture: On its automaticity and sensitivity to endogenous control." Psicologica 23: 283-309.

Rüetschi, U.-J., D. Caduff, et al. (2006). Routing by Landmarks. STRC'06 - 6th Swiss Transport Research Conference, Monte Verita, Switzerland, ETH.

Sagi, D. and B. Julész (1985). "Detection versus discrimination of visual orientation." Perception(14): 619–628.

Santos, P. E. (2007). "Reasoning about Depth and Motion form an Observer's Viewpoint." Spatial Cognition and Computation 7(2): 133-178.

Schneider, W. and R. M. Shiffrin (1977). "Controlled and Automatic Human Information-Processing.1. Detection, Search, and Attention." Psychological Review 84(1): 1-66.

Scholl, B. J. (2001). "Objects and attention: the state of the art." Cognition 80(1-2): 1-46.

Scholl, B. J. and P. D. Tremoulet (2000). "Perceptual causality and animacy." Trends in Cognitive Sciences 4(8): 299-309.

Serences, J. T., J. Schwarzbach, et al. (2004). "Control of object-based attention in human cortex." Cerebral Cortex 14(12): 1346-1357.

Shannon, C. E. (1948). "A Mathematical Theory of Communication." The Bell System Technical Journal 27: 379-423, 623-656.

Shiffrin, R. M. and W. Schneider (1977). "Controlled and Automatic Human Information-Processing.2. Perceptual Learning, Automatic Attending, and a General Theory." Psychological Review 84(2): 127-190.

Shinoda, H., M. M. Hayhoe, et al. (2001). "What controls attention in natural environments?" Vision Research 41(25-26): 3535-3545.

Siegel, A. W. and S. H. White (1975). The development of spatial representations of large-scale environments. Advances in child development and behavior. H. W. Reese, Academic Press. 10: 9-55.

Silva, M. M., J. A. Groeger, et al. (2006). "Attention-memory interactions in scene perception." Spatial Vision 19(1): 9-19.

Sorrows, M. E. and S. C. Hirtle (1999). The Nature of Landmarks for Real and Electronic Spaces. International Conference COSIT'99, Stade, Germany, Springer Verlag.

Soto, D. and M. J. Blanco (2004). "Spatial attention and object-based attention: a comparison within a single task." Vision Research 44(1): 69-81.

Spelke, E. S. (1990). "Principles of Object Perception." Cognitive Science 14(1): 29-56.

Staal, M. A. (2004). Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework. Moffett Field, California 94035, National Aeronautics and Space Administration, Ames Research Center: 177.

Steck, S. D., H. F. Mochnatzki, et al. (2003). The Role of Geographic Slant in Virtual Environment Navigation. Spatial Cognition III, LNAI 2685, Tutzing, Bavaria, Germany, Springer.

tevens, Q. (2006). "The shape of urban experience: a reevaluation of Lynch's five elements." Environment and Planning B-Planning & Design 33(6): 803-823.

Talmy, L. (1983). How language structures space. Spatial Orientation: Theory, Research, and Application. J. H. L. Pick and L. P. Acredolo. New York, Plenum Press. Spatial Orientation: Theory, Research, and Application: 225-282.

Tezuka, T. and K. Tanaka (2005). Landmark Extraction: a Web Mining Approach. Spatial Information Theory - COSIT2005, Ellicottville, NY, USA, Springer.

Tolman, E. C. (1948). "Cognitive Maps in Rats and Men." Psychological Review 55(4): 189-208.

Tom, A. and M. Denis (2004). "Language and spatial cognition: Comparing the roles of landmarks and street names in route instructions." Applied Cognitive Psychology 18(9): 1213-1230.

Torralba, A. and A. Olivia (2002). "Depth Estimation from Image Structure." IEEE Transactions on Pattern Analysis and Machine Intelligence 24(9).

Trahanias, P. E., S. Velissaris, et al. (1999). "Visual recognition of workspace landmarks for topological navigation." Autonomous Robots 7(2): 143-158.

Treisman, A. (1985). "Preattentive processing in vision." Computer Vision, Graphics and Image Processing(31): 156–177.

Treisman, A. (1986). "Features and Objects in Visual Processing." Scientific American 255(5): 114-125.

Treisman, A. and S. Gormican (1988). "Feature analysis in early vision: Evidence from search asymmetries." Psychological Review 95: 15-48.

Treisman, A., A. Vieira, et al. (1992). "Automaticity and Preattentive Processing." American Journal of Psychology 105(2): 341-362.

Treisman, A. M. and G. Gelade (1980). "Feature-Integration Theory of Attention." Cognitive Psychology 12(1): 97-136.

Turatto, M., V. Mazza, et al. (2005). "Crossmodal object-based attention: Auditory objects affect visual processing." Cognition 96.

Tversky, A. and D. Kahneman (1977). Causal thinking in judgment under uncertainty. Basic Problems in Methodology and Linguistics. R. Butts and J. Hintekka. Dordrecht, Holland, D. Reidel Publishing Company: 167-190.

Tversky, B. (1993). Cognitive Maps, Cognitive Collages, and Spatial Mental Models. Conference on Spatial Information Theory: COSIT'93, Elba Island, Italy, Springer-Verlag Berlin.

Walther, D., L. Itti, et al. (2002). Attentional selection for object recognition – a gentle way. Lecture Notes in Computer Science. Berlin, Germany, Springer. 2525: 472-479.

Weissensteiner, E. and S. Winter (2004). Landmarks in the Communication of Route Directions. Geographic Information Science. Heidelberg, Springer. 3234: 313-326.

Werner, S., B. Krieg-Brückner, et al. (1997). "Spatial Cognition: The Role of Landmark, Route, and Survey Knowledge in Human and Robot Navigation." Informatik 1997(Informatik aktuell): 41-50.

Wertheimer, M. (1923). "Untersuchungen zur Lehre von der Gestalt II." Psychologische Forschung(4): 301-350.

Williams, L. J. (1988). "Tunnel Vision or General Interference? Cognitive Load and Attentional Bias Are Both Important." The American Journal of Psychology 101(2): 171-191.

Winter, S. (2003). Route adaptive selection of salient features. COSIT'03 - Spatial Information Theory: Foundations of Geographic Information Science, Ittingen, Switzerland, Springer Verlag.

Winter, S., M. Raubal, et al. (2004). Focalizing Measures of Salience for Route Directions. Map-Based Mobile Services - Theories, Methods and Design Implementations. L. Meng, A. Zipf and T. Reichenbacher. Berlin, Springer Geosciences.

Wolfe, J. M. (1994). "Guided Search 2.0 - a Revised Model of Visual-Search." Psychonomic Bulletin & Review 1(2): 202-238.

Wolfe, J. M., S. R. Friedman-Hill, et al. (1992). "The role of categorization in visual search for orientation." Journal of Experimental Psychology: Human Perception & Performance 18(1): 34–49.

Wood, S., R. Cox, et al. (2006). "Attention design: Eight issues to consider." Computers in Human Behavior 22(4): 588-602.

Xu, Y. X. (2006). "Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness." Perception & Psychophysics 68(5): 815-828.

Appendix 1
Online Survey and Panoramic Images

This appendix briefly describes the questionnaire and the data material used for the survey and the evaluation of this work. The questionnaire and the data used in the study, as well as the statistical evaluation of the answers and the results are to be found on the CD-ROM to this dissertation.

The questionnaire was set up as an html-based online form, which participants filed out and submitted to the server, where the data was dissected into pieces that could be used to the evaluating the study. The purpose of the study, hence, was to collect data, which would provide the evidence against which the software prototype was tested. Participants examined panorama pictures of real-world urban environments, compared a set of spatial objects contained therein, and ranked them according to their visual salience or prominence. The pictures in this study were 360-degree panorama pictures taken at decision points (junctions of the traffic network) located close to the Main Station in the city of Zurich. If participants agreed to participate, they were requested to complete a survey consisting of 13 questions, which would take no more than 15-20 minutes, whereby no identifying information was collected as part of the survey.

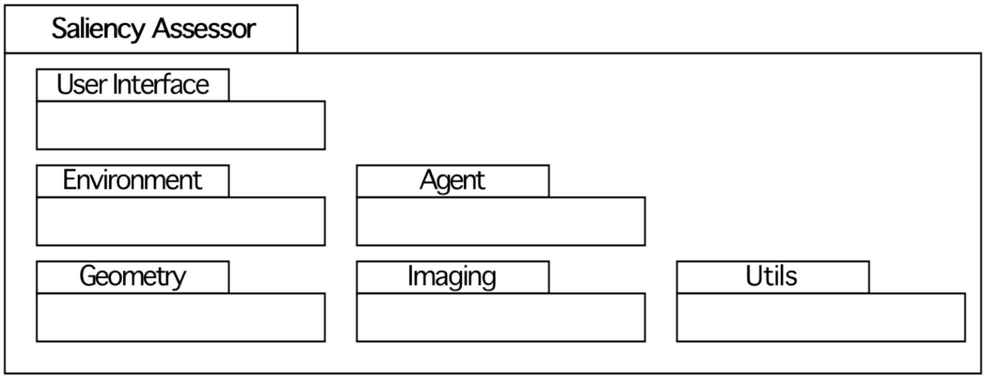
Appendix 2
Prototype Implementation

In this appendix, we briefly describe the prototype implementation of the proposed conceptual model for the assessment of landmark salience (Chapter 2) and the according computational model (Chapter 3). The data model, the methods for data extraction, and the core functionality for the assessment process are formally defined in Chapter 3. This appendix is understood as complementary documentation of the technicalities of the implementation, as well as the workflow for the computation of saliencies.

The appendix is organized as follows. First, we describe the components of the prototype, then we discuss the use of the prototype in terms of implemented functionality and workflow, and finally, we provide a simple walk-trough example for the saliency assessment.

Components of the Prototype

The prototype was implemented as a JAVA application and includes the following packages:



The *Saliency Assessor* Package

This package consists of two classes that glue the single sub-packages into a single application. The main class (Simulator.java) provides the start-point of the application and the second class (AppParams.java) contains the default settings and basic runtime parameters application.

The *User Interface* Package

This package contains the classes that make up the user interface. The prototype is implemented as a command-line application and allows defining simulation scenarios and assessing the salience of spatial objects based on these scenarios. Scenarios can be defined on the base of a configuration file or directly by use of the command provided at the command line.

The *Environment* Package

This package corresponds to the data model as described in Chapter 2.

The *Agent* Package

This package corresponds the implementation of the computational model presented in Chapter 2.

The *Geometry* Package

This package contains classes for dealing with geometric properties of the input data. Specifically, this package provides the functionality required to process vector-based shapes and to extract the according features, such as height, width, area, etc.

The *Imaging* Package

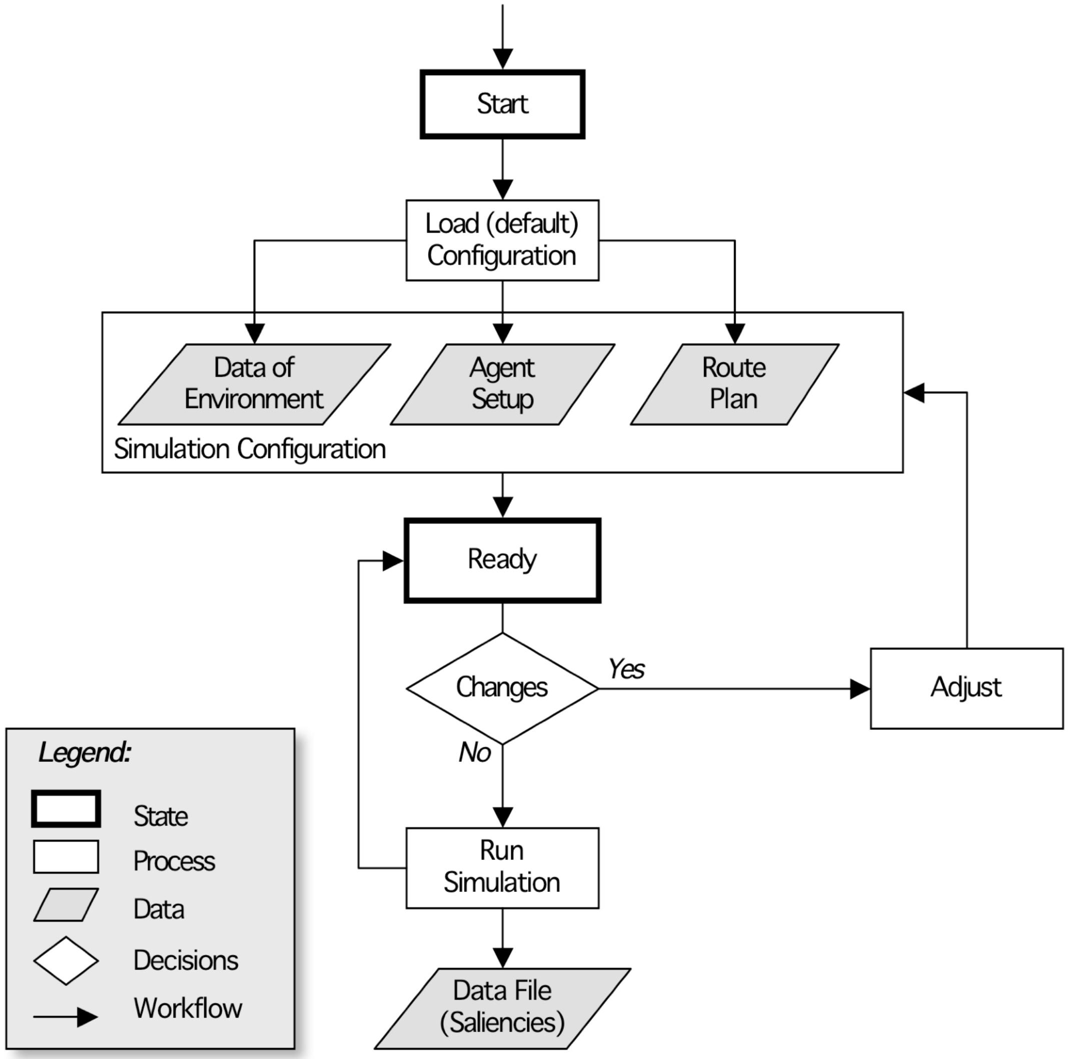
This package provides the functionality required for handling raster data. It implements various filters for the generation of image pyramids, along with methods for extracting color, intensity, and orientation contrast.

The *Utils* Package

This package contains the classes that provide basic I/O-functionality, as well as the methods required for reading and writing Scalable Vector Graphics (SVG)-files.

Using the Prototype

The following flow-chart illustrated the basic workflow of the prototype:



After launching the application, the default configuration file is loaded and the simulation set up accordingly. The configuration consists of the link to the directory that contains the data of the environment and the link to the configuration file that contains the settings for the agent and the definition of the route. This file has the following structure:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
    <ApplicationParameters>
          <SimulationParameters>
                <SimulationName>Scenario 4</SimulationName>
                <EnvDirectory>data/env</EnvDirectory>
                <LogFile>data/log.txt</LogFile>
          </SimulationParameters>
          <RouteParameters>
                <RouteName>Road of Default Setup</RouteName>
                <Route>
                      <Scene>58498</Scene>
                      …
                      <Scene>57941</Scene>
                </Route>
          </RouteParameters>
          <AgentParameters>
                <Modality>walking</Modality>
                <Perception>true</Perception>
                <Cognition>true</Cognition>
                <Context>true</Context>
                <BayesianNetwork>BNetFile</BayesianNetwork>
          </AgentParameters>
    </ApplicationParameters>
```

The structure of the configuration file conforms to the XML 1.0 specifications and consists of four main sections, namely 1) the simulation parameters, 2) the definition of the route, and 3) and the parameters for the agent. The simulation parameters consist of the name of the simulation, the location of the directory that contains the data of the environment, and the location of the log-file where the results will be stored. The route parameters consists of a name for the current route and the scenes that make up this route. A route consists of at least two connected nodes, but has no upper limit on the number of scenes as long as the sequence of scenes corresponds to the network topology of the data in the environment. The parameters for the agent indicate the context of the journey; the configuration of the components of salience that will be assessed in this scenario (i.e., perceptual, cognitive, and/or contextual salience, and the link to the configuration file for the Bayesian network.

When running a simulation, the configuration can be defined before starting the prototype application by adapting the parameters in the default configuration file. The configuration can also be changed and adapted by using the commands provided at command line level. All the parameters of the previously described configuration file can be changed and adapted from the user interface provided at command line.

Walkthrough Example

The following walkthrough example illustrates the workflow for the assessment of landmark salience. For the walkthrough example we will use the same configuration as for the validation of scenario 1 in Chapter 4:

Route:	(58012, 57950, 57941, 57975, 57960, 57956, 58029, 58281, 58594, 58602, 58626, 58521, 58948)
Agent Configuration:	($\neg \mathit{LTM} \wedge \mathit{Walking}$) (Cognitive and Contextual influences disabled)
Observed Variables:	*Perceptual Salience* of objects *A* to *H* in all scenes

The following steps are required to compute the saliency vectors:

1. Define the name of the scenario, e.g. scenario 1
2. Set the location of the data of the environment, e.g., /data/env
3. Set the location of the log-file, e.g., /data/log/scenario1.txt
4. Define the name of the route and the scenes that make up the route
5. Define the modality of travel, i.e., waking
6. Set up the agent, i.e., Perception=TRUE, Cognition=FALSE, Context=FALSE
7. Set the link to the file that contains the specification of the Bayesian network, e.g., BayNet_v0
8. Run the simulation
9. Inspect the results

Steps 1 to 7 may be performed from the command line or, alternatively, by adjusting the parameters of the default configuration file. Running the simulation results in the execution of the computational model described in Chapter 3 and the production of a log-file containing the saliency values for the objects in the environment. Finally, the objects A to H that we wish to investigate can be retrieved from this file for further inspection and processing.

The log-file contains the values for the low-level components for each object and each scene, along with the values for perceptual, cognitive, and contextual salience (i.e., high-level components). The log-file is text-based and can be inspected and analyzed by means of text editors or statistical tools. Note that at this stage no methods for analysis or visualization of the results are implemented.

Curriculum Vitae

Name	Caduff, David
Geboren am	18. Februar 1974
Heimatort	Degen GR, Schweiz

Berufslehre (1990-1994) bei Cavigelli und Partner in Ilanz, GR
Abschluss: Juni 1994 als Eidg. Dipl. Vermessungszeichner

Studium Geodäsie und Geoinformatik (1995-1998) an der Fachhochschule in Muttenz, BL
Abschluss: Dezember 1998 als Dipl. Ing. HTL

Studium Informatik (1999-2000) an der Fachhochschule beider Basel in Muttenz, BL
Abschluss: März 2000 als Dipl. Ing. FH NDS

MS Studium (2000-2002) an der University of Maine at Orono, ME, USA
Abschluss: Dezember 2002 als MS in Spatial Information Science and Engineering

Doktorat (2004-2007) am Geographischen Institut der Universität Zürich,
als Doktorand angestellt: Januar 2004 bis August 2007

Publikationen mit unmittelbarem Bezug zur Dissertation

CADUFF, D. & TIMPF, S. (2005) *The Landmark Spider: Representing Landmark Knowledge for Wayfinding Tasks*. AAAI 2005 Spring Symposium. Stanford, CA, AAAI Press.

CADUFF, D. & TIMPF, S. (2006) *Contextual Salience for Wayfinding: The Influence of Task and Modality on Landmark Salience*, GIScience 2006. Münster, Germay, IfGI Prints, Münster, Germany.

CADUFF, D. & TIMPF, S. (2006) A Framework for the Assessment of Landmark Salience for Wayfinding Tasks (Extended Abstract). ICSC - International Conference on Spatial Cognition, Rome, Italy, Springer Verlag Berlin.

CADUFF, D. & TIMPF, S. (to appear) *On the Assessment of Landmark Salience for Human Navigation*. Cognitive Processing