Cogn Process (2008) 9:249–267
DOI 10.1007/s10339-007-0199-2

REVIEW

On the assessment of landmark salience for human navigation

David Caduff · Sabine Timpf

Received: 17 April 2007 / Revised: 3 August 2007 / Accepted: 9 October 2007 / Published online: 13 November 2007

Abstract In this paper, 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. Finally, we discuss the implications, consider restrictions, and explore the scope of the framework.

D. Caduff (✉)
Geographic Information Visualization and Analysis (GIVA),
Department of Geography, University of Zurich – Irchel,
Winterthurerstr. 190, 8057 Zurich, Switzerland
e-mail: caduff@geo.unizh.ch

S. Timpf
Department for Computer Science VI, University of Wuerzburg,
Am Hubland, 97074 Wuerzburg, Germany
e-mail: timpf@ki.informatik.uni-wuerzburg.de

Keywords Navigation · Landmark · Salience · Attention · Information processing

Introduction

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 systems, traffic information systems, etc.) or spatially related information systems (route planners, tourist information systems, location based services, etc.).

Problem statement and motivation

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 (Golledge 1991; Lynch 1960; Montello 1997; Montello and Freundschuh 2005; Siegel and White 1975). Several studies investigated the role of landmarks (Allen 1997; Werner et al. 1997; Fontaine and Denis 1999; Lovelace et al. 1999; Lee et al. 2002; Steck et al. 2003) and affirmed the importance of those as essential part of the production and communication of route instructions (Denis et al. 1999; Tom and Denis 2004; Daniel and Denis 2004; Weissensteiner and Winter, 2004; Newman et al. 2007). Despite this evidence, only few attempts exist to enhance route instructions with landmark knowledge (Nothegger 2003; Nothegger et al. 2004; Raubal and Winter 2002; Winter 2003; Winter et al. 2004) or to incorporate landmarks in route generation algorithms (Caduff and Timpf 2005a, b; Rüetschi 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 (Couclelis et al. 1995; Presson and Montello 1988).

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 paper, we review literature on the assessment of landmark salience, whereby we focus on the use of landmarks for human navigation, and propose a framework for the assessment of the importance of potential landmarks.

Approach

Gaerling et al. (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 trilateral relationship between observer, observed feature, and environment. Accordingly, the central assumption of our approach is that the trilateral relationship between observer, referenced spatial feature, and physical environment 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, however, differs from 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 (Raubal and Winter 2002; Sorrows and Hirtle 1999; 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 (Appleyard 1969; Lynch 1960). 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). 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. This 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 (Marcel and Dobel 2005; Parkhurst and Niebur 2003). 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 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 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 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 indicate 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.

Paper overview

These three aspects, perception, cognition, and context are fundamental aspects of our framework for the assessment of the salience of spatial features. The remainder of this paper is organized as follows. After a review of related work, we conceptualize salience of spatial features for navigation and describe them in detail. Subsequently, we discuss the implications, restrictions, and the scope of our framework, and finally, we conclude with a summary and present current and future work.

Related work

Landmarks are present throughout history as reference points for navigation and play an important role in the development of spatial knowledge and for solving spatial reasoning problems. Siegel and White (1975) introduced a three-phased theory of acquisition of spatial knowledge, which assumes that landmarks are the linking points between *Route* and *Survey Knowledge*, and hence, form the foundation of cognitive maps (Downs and Stea 1977; Tolman 1948). 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. The nature of landmarks has been investigated from various points of view, such as their use as spatial points of reference (Couclelis et al. 1995; Presson and Montello 1988), or their function in the communication of route directions (Denis et al. 1999; Golledge 1991), 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.

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 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 spaces in a comprehensive way, but no formalization is proposed.

An alternative characterization of landmarks and their properties was proposed by Burnett (2000), who suggest permanence, visibility, location in relation to a decision point, uniqueness, and brevity as the main aspects of 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.

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.), and hence, a series of different approaches exits.

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 and can be viewed respectively as 2D and 3D structures, are not considered.

Elias (2003a) 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 (2002) salience model. From a computational point of view, the main objective of Elias' (2003b) approach is to automatically extract landmarks from existing data using a data mining approach. 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 (Elias 2003b; Raubal and Winter 2002; Sorrows and Hirtle 1999) 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).

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.

Tezuka and Tanaka (2005) investigated the World Wide Web as 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.

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 et al. 2005). Analogous to Raubal and Winter's (2002) approach, 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.

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 et al. 2002, 2003; Kosmopoulos and Chandrinos 2002), are complemented by more complex approaches involving visual attention and automatic extraction of salient features (Trahanias et al. 1999).

Attention-based models of landmark extraction are typically bottom-up as they extract a set of pre-attentive features (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 question related to the determination of landmark saliency. Typically, however, 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.

Main contribution

The main contribution of this paper is a framework for the 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.

Conceptualization of 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* (Fig. 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

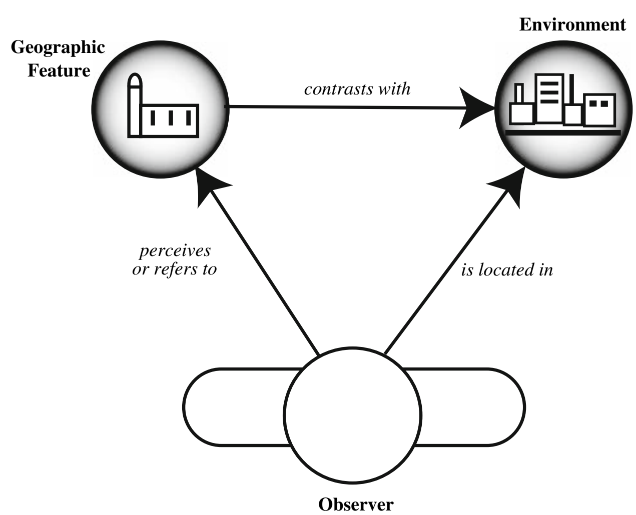

Fig. 1 The trilateral relationship between *Observer*, *Environment*, and *Geographic Feature*. The Observer is located in the environment and perceives or refers to some geographic feature, which contrasts with the environment. This configuration defines the basic assumption of our framework

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. Note that this fact only applies for the acquisition of landmark knowledge during navigation, not for the communication of landmarks or route directions, which may refer to prominent spatial objects that used to exist at specific places. 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 et al. 1995).

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.), and is tightly linked to the limited capacity of our working memory (Miller 1956), for which several explanations, such as the Cognitive Load theory (Sweller 1988), have been proposed.

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 (Eriksen and Yeh 1985; Miller 1956) and theories of human information processing (Gaerling 1999; Newell and Simon 1972). 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 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.

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.,

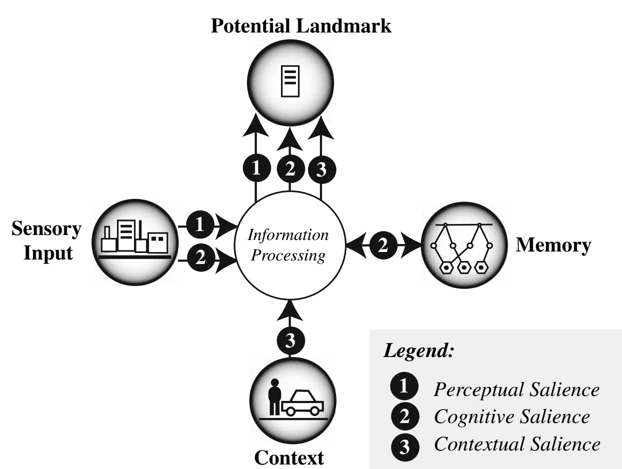

Fig. 2 The three different types of salience that contribute to the overall salience of geographic objects: a part of the sensory input contributes directly to the salience of the landmark (Perceptual Salience). Former experience and memory modulates sensory input in a top-down manner and contributes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

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 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 (Fig. 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 express 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.

Quantifying the components of 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 paper, we refer to spatial features that are potential landmarks as *Spatial Objects*. In the following sections we will discuss the components that define the salience of spatial objects in more detail and describe ways to computationally quantify them.

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 in people who are not blind 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 (1980) model of *Space-* or *Location-based Attention* and the second is the developing theory of *Object-based Attention* (see Scholl 2001, for a review) (Fig. 3).

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 et al. 1999; Müller and Kleinschmid 2003). Attention may well be object-based in some context, location-based in 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

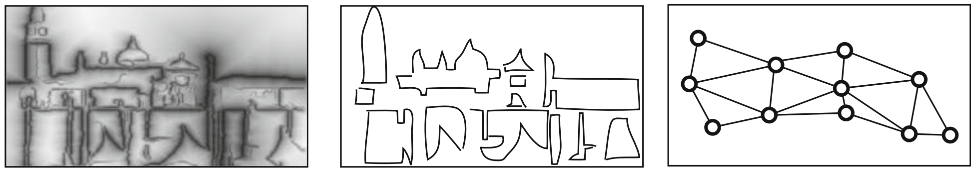

Fig. 3 The three components of Perceptual Salience: **a** location-based attention, **b** object-based attention, and **c** Scene context. Each of the components has its own set of attributes, which contribute to the degree of salience of the object

movement, and hence focus of attention (Aivar et al. 2005; Hayhoe et al. 2000; Shinoda et al. 2001). 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.

A highly successful implementation of location-based attention is Itti and Koch's saliency-based spatial attention model (Itti et al. 1998). A saliency map (cf. Fig. 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.

Object-based attention defines the salience of single objects or groups of objects contained in a scene (Fig. 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

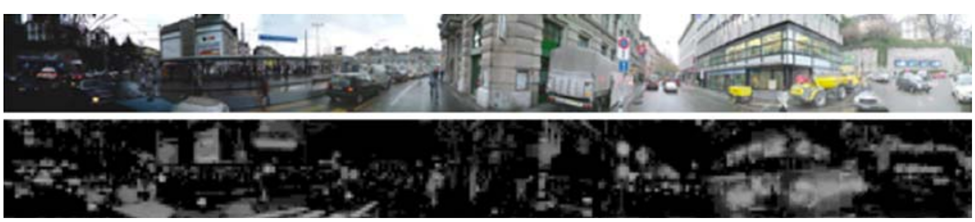

Fig. 4 The picture on top shows a typical urban scene and the picture below shows the corresponding saliency map, as generated by Itti and Koch's saliency-based model of spatial attention. Each salient or conspicuous location in an image or a scene is evaluated with respect to its surrounding

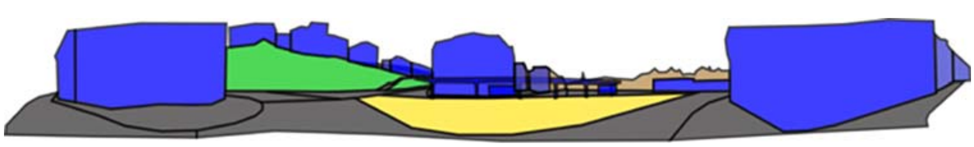

Fig. 5 Object-based attention is influenced by the structure of spatial objects. We base our assessment on the similarity of shape, size, and orientation of objects across the scene

with results from psychology that state that the two types complement, rather than exclude each other (Soto and Blanco 2004).

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 (Fig. 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 to support 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 terms of topology (e.g., 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.

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 et al. 2006). In psychology, the term cognition is 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. 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. 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 well-known objects are preferred over unknown objects (Fig. 7).

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 (Fig. 8). The motivation for this abstraction of the mental representation of navigational space is the incremental nature of route learning (Kuipers 1982; Golledge 1992; Siegel and White 1975). Observations of specific objects are acquired while navigating and stored in long-term memory, from where they are retrieved if necessary.

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 et al. (2006), who proposes modeling recognition memory using the similarity structure of input, we will use

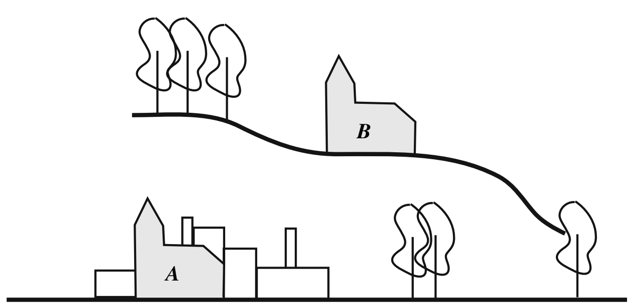

Fig. 6 An example of a spatial scene, where objects *A* and *B* have the same attributes and salience, but the spatial configuration provides additional information about the salience of the object

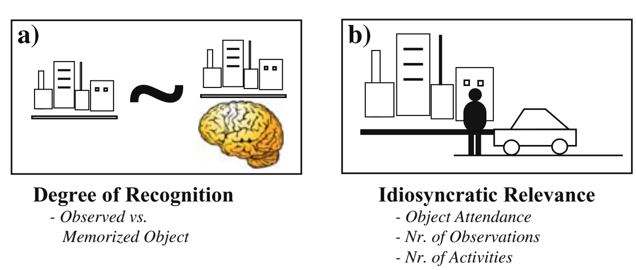

Fig. 7 The two components of Cognitive Salience: **a** the degree of recognition, and **b** the idiosyncratic relevance

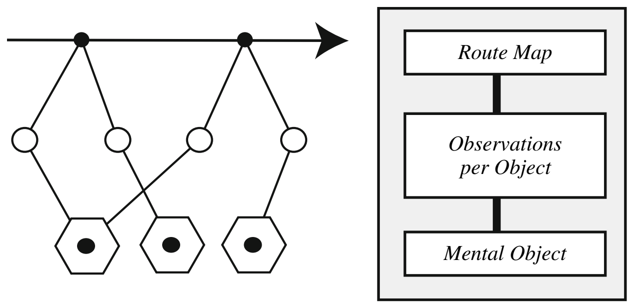

Fig. 8 The structure of the route map that is created when navigating: at each waypoint along the route observations to geographic objects are collected. The sum of observations to a single geographic object constitutes a mental object, which we will use in the assessment of Cognitive Salience

the similarity between observed object features and mental object features for assessing the degree of recognition.

Familiarity, on the other hand, relates to idiosyncratic relevance, which is tightly linked to personal, cultural, or historical significance of objects, and is the result of experiences, activities, and facts associated with these objects (Presson and Montello 1988). The degree of personal and cultural significance varies with characteristics of individuals (age, preferences, knowledge, etc.) and geographic regions (local, regional, interregional meaning and significance, etc.), and consequently, is accordingly difficult to model. 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.

We abstracts idiosyncrasy to the level of repetitive activities and observations associated with particular objects, and assume that idiosyncratic relevance increases with the number of recurrences of a specific object, along with the number of activities associated with this 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. 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.

Contextual Salience

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 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 (Fig. 9), and (2) *Modality-based Context*, which describes the mode of transportation and the amount of resources that need to be allocated (Fig. 10).

A definition of the task that is 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 1999b). 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

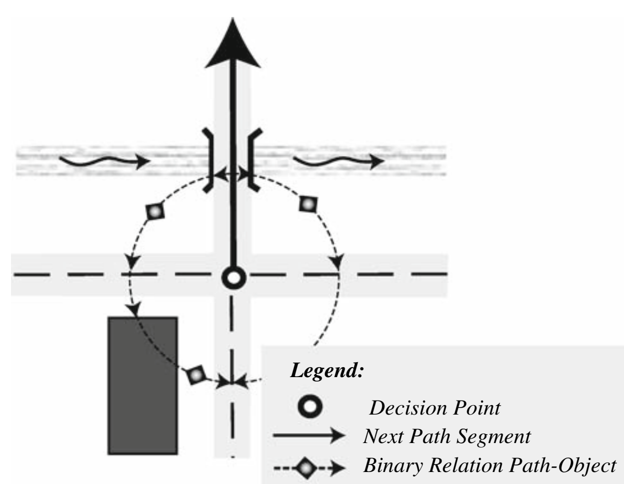

Fig. 9 A spatial scene including four possible paths and three potential landmarks (i.e., a river, a bridge, and a building), as experienced by observers during navigation. The binary relation between path and geographic feature defines how valuable geographic features are when considering a specific path

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.

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 (Staal 2004; May et al. 2003a, b; Lee et al. 2007). 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 (Fig. 10).

The field of view, or visual field, 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. This limited field of view has been termed the *useful visual field* (Ball et al. 1993), and has been shown to be smaller than the peripheral visual field (Roge et al. 2005). Our model accounts for these contextual differences when assessing salience, resulting in a ranking of potential landmarks in a scene that is based on the field of view navigators have when using different modes of transportation.

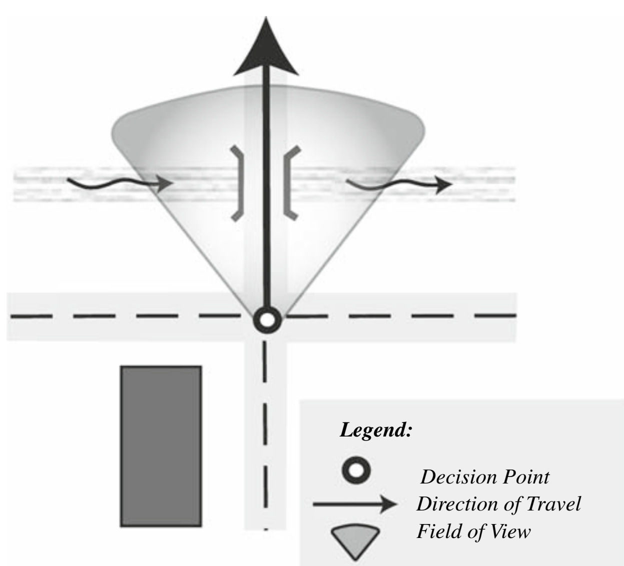

Fig. 10 The modality of travel (i.e., walking, driving, or riding) influences both, the cognitive load put on the observer, as well as the degree of physical freedom. The remaining physical and cognitive resources are allocated accordingly, which influences the focus of attention and field of view and hence, the prominence of surrounding geographic features

Integrated saliency assessment

So far, we have identified three types of *high-level saliency components* (i.e., perceptual, cognitive, and contextual) 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. In order to assess the overall salience of spatial objects, these components need to be integrated in 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 into this process.

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 features 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.

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 (Fig. 11). 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 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 idiosyncratic relevance 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.

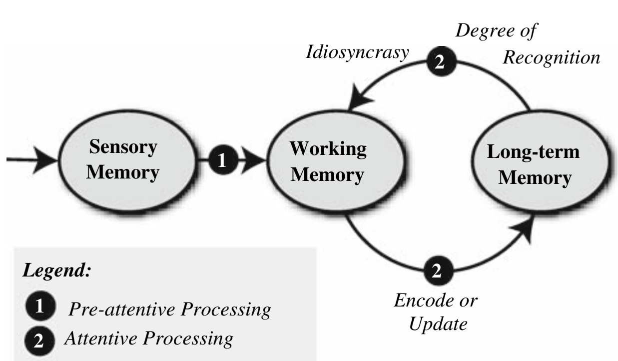

Fig. 11 Model of human information processing: each stage holds a refined representation of the spatial scene. Pre-attentive processing of the data in sensory memory results in a perceptual scene representation in working memory. Objects in the perceptual scene representation are then assessed sequentially for salient features, and finally, objects in long-term memory are updated with new facts

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 that 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 (Chater et al. 2006; Scholl and Tremoulet 2000; Kersten and Yuille 2003; Kersten et al. 2004). 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).

In the context of probabilistic inference, 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).

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 (Fig. 12). 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 et al. 2004; Staal 2004), and finally,
- Scene context influences the allocation of attention to specific objects (De Graef 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 findings from the literature and our own assumptions yield the Bayesian network depicted in Fig. 12. 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 according 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). We propose the use of a discrete probability distribution, in combination with the hypothesis of uniform influence of parent nodes, as initial configuration. The validity of this hypothesis, however, remains to be challenged during the evaluation of the framework, and, when available, revised according to scientific findings.

The last step in computing the posterior distribution of variables given evidence is called *Probabilistic Inference* (Jensen 2001). The posterior probability gives sufficient

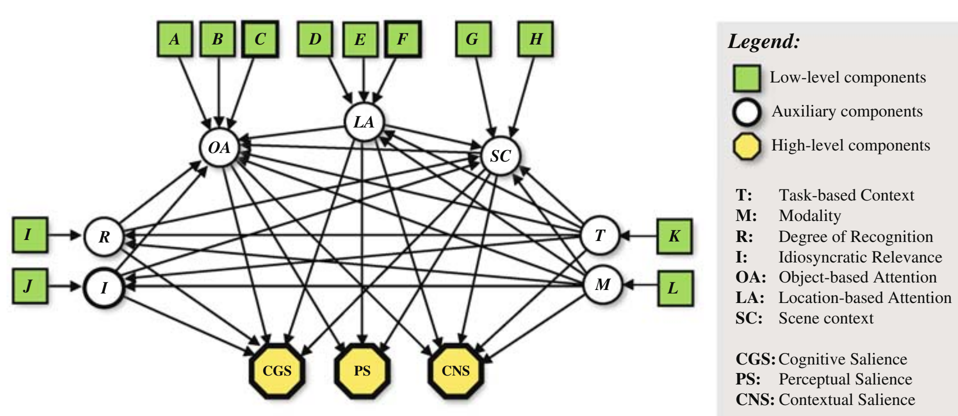

Fig. 12 The structure of the Bayesian network used for simulating the salience assessment process. The low-level components are derived directly from input data and serve as evidence. The auxiliary components account for the different types of attention, and the high-level components describe the resulting saliency of the observed spatial object

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.

Summary

The aim of this paper is to propose a conceptual framework for the assessment of salience of spatial objects tailored to the requirements of navigation of sighted humans. As such, the framework provides the ontology that is necessary to formally model landmarks and implement them in information systems for navigation. To achieve this goal, we conceptualized our understanding of salience, contrasted it with views in the literature, investigated which factors influence the prominence of spatial objects, and proposed a probabilistic approach for integrating the different factors in order to determine the object's salience. We introduced the concept of 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 the components into a single computational model.

Discussion

The primary goal of this paper was to review literature on landmark saliency assessment and analyze which components add to the relative importance of spatial objects for navigation. Knowing what factors influence landmark saliency is important for accurate assessments of the saliency, and hence, the discrimination of landmarks. The result of this analysis is a framework that considers different types of salience in an integrative way. This section critically discusses the framework in terms of implications, restrictions, and scope.

Implications

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 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.

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.

Restrictions

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 object's saliency as an attentional target (De Graef et al. 2000). We do not account for this a priori knowledge of the semantics of spatial features, but the framework is structured such that the incorporation of additional factors is easily possible. Another issue to consider is that of identification of spatial objects (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. Our framework does not account for such an emergence of landmarks.

The proposed human information processing cycle abstracts the ease of encoding and memorizing single objects [e.g., typical objects are hard to remember while atypical 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.

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, especially for those groups that do no rely on visual input, but on other sensory input. A shift of sensory input, however, implies a shift of strategies for spatial orientation (Golledge 1999b). 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. The framework was developed with adaptation and flexibility in mind, and therefore, once evaluated for vision, may be extended by incorporating additional sensory input.

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 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.

A final issue to consider is in terms of practical applicability of the framework and concerns the collection of appropriate data and the level of detail. The evaluation of environmental features proposed in this paper is based on the egocentric frame of reference of the observer. The practical implications of this approach are manifold. First of all, a scene containing the spatial features perceived from a specific location needs to be computed from a source dataset, and the second point to consider is that the spatial scene needs to feature a level of detail that allows for extraction of the low-level features (i.e., color, orientation, etc.). These critical points need to be considered when collecting the data.

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 complete. 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 (1999a), 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, culminating in a superior structure often referred to as the cognitive map (Golledge 1999a; Kuipers 1982; Miller 1956) 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 we do model previous knowledge in our framework, we do no 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 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. Hence, the framework treats landmarks as navigational aid, rather than as an organizing concept

Outlook and future work

The main contribution of this paper is a review of relevant literature and the definition of a conceptual framework for the assessment of landmark 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. We define a conceptual framework of landmark properties and contributing components, and set the frame for a computational model for the assessment and integration of these components. 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.

On the base of this framework, we are currently working on a prototype implementation for the assessment of landmark saliency. The prototype application includes a refined computational model and will serve as test-bed for future research. Our framework defines the overall structure of the assessment, but leaves open how the low-level components of saliency (e.g., degree of recognition, task-based context) are derived. The prototype application will provide support in answering these open questions. Refining the computational model and the prototype application will also help answering questions related to usability and performance, and provide insight into technical and infrastructural questions, such as feasibility and acquisition of appropriate data.

A very important question in this context is concerned with the evaluation of the framework. We will use the prototype application for evaluation of our framework and plan to divide the evaluation process in two steps, namely: (1) Verification and (2) Validation. We understand the process of verifying the framework as confirmation by examination and provision of objective evidence that specified requirements have been fulfilled. Verification will answer questions related to inner correctness and performance of the prototype. Validation, on the other hand, is understood as the process of *ground-truthing*, and will determine if the framework can be properly applied as intended. Validation will answer questions related to performance in real world scenarios and fine-tuning with respect to human performance. Successful evaluation of the prototype is crucial for further research and will have to be performed accordingly.

The prototype implementation is designed such that it can be integrated in agent-based simulations. Agent-based simulations are increasingly becoming a popular tool for various lines of research and applications, including research on human cognition and uncertainty, information retrieval, and environmental design. Agent-based simulations incorporating our framework will help answering questions related to user-group refinement or taxonomic cataloguing of landmarks, as well as incorporation of landmarks in the route generation process, which was one of the main objectives of our work.

Acknowledgments The authors would like to thank Urs-Jakob Rueetschi for his valuable input. This work is supported by the Swiss National Science Foundation under grant number 2151-06529101.

References

Aivar MP, Hayhoe MM, Chizk CL, Mruczek REB (2005) Spatial memory and saccadic targeting in a natural task. J Vis 5(3):177–193

Allen GL (1997) From knowledge to words to wayfinding: issues in the preduiction and comprehension of route directions. In: Hirtle SC, Frank AU (eds) Spatial information theory: a theoretical basis for GIS, international conference COSIT '97. Springer, Laurel Highlands, pp 363–372

Anderson ML (2003) Embodied cognition: a field guide. Artif Intell 149(1):91–130

Appleyard D (1969) Why buildings are known—predictive tool for architects and planners. Environ Behav 1(2):131–156

Ball K, Owsley C et al (1993) Visual attention problems as a predictor of vehicle crashes in older drivers. Invest Ophthalmol Vis Sci 34:3110–3123

Biederman I (1972) Perceiving real-world scenes. Science 177(4043):77–80

Burnett G (2000) "Turn right at the traffic lights": the requirement for landmarks in vehicle navigation systems. J Navig 53(3):499–510

Busquets D, Sierra C, De Màntaras RL (2002) A multi-agent approach to fuzzy landmark-based navigation. J Multivalued Log Soft Comput 9:195–220

Busquets D, Sierra C, De Màntaras RL (2003) A multi-agent approach to qualitative landmark-based navigation. Auton Robots 15:129–154

Caduff D, Timpf S (2005a) The landmark spider: representing landmark knowledge for wayfinding tasks. AAAI 2005 Spring Symposium. AAAI Press, Stanford

Caduff D, Timpf S (2005b) The landmark spider: weaving the landmark web. STRC'05—5th Swiss Transport Research Conference. Monte Verità, Switzerland, ETH, CD-ROM

Chater N, Tenenbaum JB, Yuille A (2006) Probabilistic Models of Cognition: Conceptual foundations. Trends Cogn Sci 10(7):287–291

Couclelis H, Golledge RG, Gale N, Tobler W (1995) Exploring the anchor-point hypothesis of spatial cognition. In: Gaerling T (ed) Urban cognition. Academic, London

Daniel MP, Denis M (2004) The production of route directions: investigating conditions that favour conciseness in spatial discourse. Appl Cogn Psychol 18(1):57–75

De Graef P, Lauwereyns J, Verfaillie K (2000) Attentional Orienting and Scene Semantics. Psychological Reports Nr.268. Laboratory of Experimental Psychology, University of Leuven, Belgium

Denis M, Pazzaglia F, Cornoldi C, Bertolo L (1999) Spatial discourse and navigation: an analysis of route directions in the city of Venice. Appl Cogn Sci 13:145–174

Downs RM, Stea D (1977) Maps in minds:reflections on cognitive mapping. Harper & Row Publishers, New York, p 284

Elias B (2003a) Determination of landmarks and reliability criteria for landmarks. Fifth Workshop on Progress in Automated Map Generalization. ICA Commission on Map Generalization, IGN, Paris

Elias B (2003b) Extracting landmarks with data mining methods. In: Kuhn W, Worboys M, Timpf S (eds) International conference on spatial information theory, COSIT 2003. Springer, Kartause Ittingen, pp 375–389

Eriksen CW, St James JD (1986) Visual attention within and around the field of focal attention: a zoom lens model. Percept Psychophys 40(4):225–240

Eriksen CW, Yeh YY (1985) Allocation of attention in the visual field. Exp Psychol Human Percept Perform 11:583–597

Escrig MT, Toledo F (2000) Autonomous robot navigation using human spatial concepts. Int J Intell Syst 15(3):165–196

Fontaine S, Denis M (1999) The production of route instructions in underground and urban environments. In: Freksa C, Mark DM (eds) Spatial information theory: cognitive and computational foundations of geographic information science, International Conference COSIT '99. Springer, Stade, pp 83–94

Funes MJ, Lupianez J, Milliken B (2005) The role of spatial attention and other processes on the magnitude and time course of cueing effects. Cogn Process 6(2):98–116

Gärling T (1999) Human information processing in sequential spatial choice. In: Golledge RG (ed) Wayfinding behavior: cognitive mapping and other spatial processes. John Hopkins University Press, Baltimore, pp 89–98

Gärling T, Böök A, Lindberg E (1986) Spatial orientation and wayfinding in the designed environment: a conceptual analysis and some suggestions for postoccupancy evaluations. J Archit Plann Res 3:55–64

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

Gigerenzer G, Murray DJ (1987) Cognition as intuitive statistics. Erlbaum, Hillsdale

Golledge RG (1991) Cognition of physical and built environments. In: Gaerling T, Evans GW (eds) Environment, cognition and action: an integrated approach. Oxford University Press, New York

Golledge RG (1992) Place recognition and wayfinding—making sense of space. Geoforum 23:199–214

Golledge RG (1999a) Human wayfinding and cognitive maps. In: Golledge RG (ed) Wayfinding behavior: cognitive mapping and other spatial processes. John Hopkins University Press, Baltimore, pp 5–45

Golledge RG (1999b) Wayfinding behavior: cognitive mapping and other spatial processes. John Hopkins University Press, Baltimore, p 428

Haken H, Portugali J (2003) The face of the city is its information. J Environ Psychol 23(4):385–408

Hayhoe MM, Shinoda H, Shrivastava A (2000) Attention in Natural Environments. Invest Ophthalmol Vis Sci 41:422–422

Henderson JM, Hollingworth A (1999) High-level Scene Perception. Annu Rev Psychol 50:243–271

Hollands MA, Patla AE, Vickers JN (2002) "Look where you're going!" Gaze Behaviour associated with maintaining and changing the direction of locomotion. Exp Brain Res 143(2):221–230

Itti L, Koch C, Niebur E (1998) A Model of Saliency-based Visual Attention for Rapid Scene Analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259

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

Janzen G, Turennout MV (2004) Selective neural representation of objects relevant for navigation. Nat Neurosci 7(6):673–677

Jensen FV (2001) Bayesian networks and decision graphs. Springer, New York

Kersten D (2002) Object perception: generative image models and bayesian inference. In: Bülthoff HH et al (eds) Biologically motivated computer vision second international workshop Lecture Notes in Computer Science, vol 2525. Springer, Heidelberg, pp 207–218

Kersten D, Mamassian P, Yuille A (2004) Object Perception as Bayesian inference. Annu Rev Psychol 55:271–304

Kersten D, Yuille A (2003) Bayesian models of object perception. Curr Opin Neurobiol 13(2):150–158

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

Klippel A, Richter K-F, Hansen S (2005) Wayfinding choreme maps. In: Bres S, Laurini R (eds) Visual information and information systems, 8th international conference, VISUAL 2005. Springer, Amsterdam, pp 94–108

Klippel A, Winter S (2005) Structural salience of landmarks for route directions. In: Cohn AG, Mark DM (eds) Spatial information theory—COSIT05. Springer, Ellicottville, pp 347–362

Koch C, Ullman S (1985) Shifts in selective visual-attention—towards the underlying neural circuitry. Human Neurobiol 4(4):219–227

Kosmopoulos DI, Chandrinos KV (2002) Definition and extraction of visual landmarks for indoor robot. Methods and applications of artificial intelligence: second hellenic conference on AI, SETN 2002. Springer, Heidelberg, pp 401–412

Kosslyn SM (1989) Understanding Charts and Graphs. Appl Cogn Psychol 3:185–226

Kubovy M, Cohen DJ, Hollier J (1999) Feature integration that routinely occurs without focal attention. Psychon Bull Rev 6(2):183–203

Kuipers BJ (1982) The "Map in the Head'' Metaphor. Environ Behav 14(2):202–220

Lacroix JPW, Murre JMJ, Postma EO, van den Herik HJ (2006) Modeling recognition memory using the similarity structue of natural input. J Cogn Sci Soc 30(1):121–145

Lee PU, Tappe H, Klippel A (2002) Acquisition of landmark knowledge from static and dynamic presentation of route maps. In: Gray W, Schunn C (eds) Twenty-fourth annual conference of the cognitive science society

Lee Y-C, Lee JD et al (2007) Visual attention in driving: the effects of cognitive load and visual disruption. Hum Factors 49(4):721–733(13)

Lewis D (1973) Causality. J Philos 556–567

Lovelace KL, Hegarty M, Montello DR (1999) Elements of good route directions in familiar and unfamiliar environments. In: Freksa C, Mark DM (eds) Spatial information theory: cognitive and computational foundations of geographic information science, international conference COSIT '99. Springer, Stade, pp 65–82

Lynch K (1960) The image of the city. MIT, Boston

Marcel A, Dobel C (2005) Structured perceptual input imposes an egocentric frame of reference—pointing, imagery, and spatial self-consciousness. Perception 34(5):429–451

May AJ, Ross T, Bayer SH (2003a) Drivers' information requirements when navigating in an urban environment. J Navig 56(1):89–100

May AJ, Ross T, Bayer SH, Tarkiainen MJ (2003b) Pedestrian navigation aids: information requirements and design principles. Personal Ubiquitous Comput 7(6):331–338

Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81–97

Montello DR (1997) The perception and cognition of environmental distance: direct sources of information. In: Hirtle SC, Frank AU (eds) Conference on spatial information theory: a theoretical basis for GIS, COSIT'97. Springer, Laurel Highlands, pp 297–312

Montello DR (2003) Navigation. In: Shah P, Miyake A (eds) Handbook of visuospatial cognition. Cambridge University Press, Cambridge, pp 257–294

Montello DR, Freundschuh S (2005) Cognition of geographic information. In: Mcmaster RB, Usery EL (eds) A research agenda for geographic information science. CRC Press, Boca Raton

Moulin B, Kettani D (1999) Route generation and description using the notion of object's influence area and spatial conceptual map. Spatial Cogn Comput 1(3):227–259

Müller NG, Kleinschmid A (2003) Dynamic interaction of object- and space-based attention in retinotopic visual areas. J Neurosci 23:9812–9816

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

Newman EL, Caplan JB, Kirschen MP, Korolev IO, Sekuler R, Kahana MJ (2007) Learning your way around town: how virtual taxicab drivers learn to use both layout and landmark information. Cognition (in press)

Nothegger C (2003) Automatic selection of landmarks. Institute for geoinformation. Technical University of Vienna

Nothegger C, Winter S, Raubal M (2004) Computation of the salience of features. Spat Cogn Comput 4:113–136

Olshausen BA, Anderson CH, Vanessen DC (1992) Computer-simulation of a dynamic routing model of visual-attention. Invest Ophthalmol Vis Sci 33:1263–1263

Parkhurst DJ, Niebur E (2003) Scene content selected by active vision. Spat Vis 16(2):125–154

Posner MI (1998) Foundations of cognitive science. MIT, New York

Presson CC, Montello DR (1988) Points of reference in spatial cognition—stalking the elusive landmark. Br J Dev Psychol 6(4):378–381

Raubal M, Winter S (2002) Enriching wayfinding instructions with local landmarks. In: Egenhofer MJ, Mark DM (eds) Geographic information science. Springer, Berlin, pp 243–259

Roge J, Pebayle T et al (2005) Useful visual field reduction as a function of age and risk of accident in simulated car driving. Invest Ophthalmol Vis Sci 46:1774–1779

Ruz M, Lupianez J (2002) A review of attentional capture: on its automaticity and sensitivity to endogenous control. Psicologica 23(1):283–309

Rüetschi U-J, Caduff D, Schulz F, Wolff A, Timpf S (2006) Routing by landmarks. STRC'06—sixth Swiss transport research conference. Monte Verita, Switzerland, ETH, CD-ROM

Schneider W, Shiffrin RM (1977) Controlled and automatic human information-processing.1. Detection, search, and attention. Psychol Rev 84:1–66

Scholl BJ (2001) Objects and attention: the state of the art. Cognition 80(1):1–46

Scholl BJ, Tremoulet PD (2000) Perceptual causality and animacy. Trends Cogn Sci 4(8):299–309

Serences JT, Schwarzbach J, Courtney SM, Golay X, Yantis S (2004) Control of object-based attention in human cortex. Cereb Cortex 14(12):1346–1357

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423, 623–656

Shiffrin RM, Schneider W (1977) Controlled and automatic human information-processing.2. Perceptual learning, automatic attending, and a general theory. Psychol Rev 84:127–190

Shinoda H, Hayhoe MM, Shrivastava A (2001) What controls attention in natural environments? Vis Res 41:3535–3545

Siegel AW, White SH (1975) The development of spatial representations of large-scale environments. In: Reese HW (ed) Advances in child development and behavior. Academic, London, pp 9–55

Silva MM, Groeger JA, Bradshaw MF (2006) Attention-memory interactions in scene perception. Spat Vis 19(1):9–19

Sorrows ME, Hirtle SC (1999) The nature of landmarks for real and electronic spaces. In: Freksa C, Mark DM (eds) Spatial information theory: cognitive and computational foundations of geographic information science, international conference COSIT '99. Springer, Stade, pp, 37–50

Soto D, Blanco MJ (2004) Spatial attention and object-based attention: a comparison within a single task. Vis Res 44(1):69–81

Spelke ES (1990) Principles of object perception. Cogn Sci 14(1):29–56

Staal MA (2004) Stress, cognition, and human performance: a literature review and conceptual framework. Moffett Field, California 94035, National Aeronautics and Space Administration, Ames Research Center

Steck SD, Mochnatzki HF, Mallot HA (2003) The role of geographic slant in virtual environment navigation. In: Freksa C, Brauer W, Habel C, Wender KF (eds) Spatial Cognition III, LNAI 2685, Springer, Heidelberg, pp 62–76

Stevens Q (2006) The shape of urban experience: a reevaluation of Lynch's five elements. Environ Plann B Plann Des 33(6):803–823

Sweller J (1988) Cognitive load during problem solving: effects on learning. Cogn Sci 12(1):257–285

Tezuka T, Tanaka K (2005) Landmark extraction: a web mining approach. In: Mark DM (ed) Spatial information theory—COSIT05. Springer, Ellicottville, pp 379–398

Tolman EC (1948) Cognitive maps in rats and men. Psychol Rev 55:189–208

Tom A, Denis M (2004) Language and spatial cognition: comparing the roles of landmarks and street names in route instructions. Appl Cogn Psychol 18(9):1213–1230

Trahanias PE, Velissaris S, Orphanoudakis SC (1999) Visual recognition of workspace landmarks for topological navigation. Auton Robots 7(2):143–158

Treisman A, Gormican S (1988) Feature analysis in early vision: evidence from search asymmetries. Psychol Rev 95(1):15–48

Treisman A, Vieira A, Hayes A (1992) Automaticity and preattentive processing. Am J Psychol 105(2):341–362

Treisman AM, Gelade G (1980) Feature-integration theory of attention. Cogn Psychol 12(1):97–136

Turatto M, Mazza V, Umilta' C (2005) Crossmodal object-based attention: auditory objects affect visual processing. Cognition 96:B55–B64

Tversky B (1993) Cognitive maps, cognitive collages, and spatial mental models. In: Frank AU, Campari I (eds) Conference on spatial information theory: COSIT'93. Elba Island, Italy, Springer, Berlin, pp 14–24

Weissensteiner E, Winter S (2004) Landmarks in the Communication of route directions. In: Egenhofer MJ, Miller H, Freksa C (eds) Geographic information science 2004. Lecture Notes in Computer Science, vol 3234. Springer, Berlin, pp 313–326

Werner S, Krieg-Brückner B, Mallot HA, Schweizer K, Freksa C (1997) Spatial cognition: the role of landmark, route, and survey knowledge in human and robot navigation. In: Jarke M, Pasedach K, Pohl K (eds) Informatik aktuell, Springer, Berlin, pp 41–50

Williams LJ (1988) Tunnel vision or general interference? cognitive load and attentional bias are both important. Am J Psychol 101(2):171–191

Winter S (2003) Route adaptive selection of salient features. In: Kuhn W, Worboys M, Timpf S (eds) COSIT'03—spatial information theory: foundations of geographic information science. Springer, Heidelberg, pp 320–334

Winter S, Raubal M, Nothegger C (2004) Focalizing measures of salience for route directions. In: Meng L, Zipf A, Reichenbacher T (eds) Map-based mobile services—theories, methods and design implementations. Springer Geosciences, Berlin, pp 127–142

Wolfe JM (1994) Guided Search 2.0—a revised model of visual-search. Psychon Bull Rev 1(2):202–238
Wood S, Cox R, Cheng P (2006) Attention design: eight issues to consider. Comput Hum Behav 22(4):588–602
Xu YX (2006) Understanding the object benefit in visual short-term memory: the roles of feature proximity and connectedness. Percept Psychophys 68(5):815–828