Attention Approximation of Mobile Users towards their Environment

Johann Schrammel

CURE – Center for Usability Research & Engineering

Modecenterstr. 17, Obj. 2, 1110 Vienna, Austria

schrammel@cure.at

Georg Regal

CURE – Center for Usability Research & Engineering

Modecenterstr. 17, Obj. 2, 1110 Vienna, Austria

regal@cure.at

Manfred Tscheligi

ICT&S, University of Salzburg, 5020 Salzburg, Austria &

AIT – Austrian Institute of Technology GmbH, 1220 Vienna, Austria

manfred.tscheligi@sbg.ac.at   or   manfred.tscheligi@ait.ac.at

Abstract

Public environments are increasingly equipped with interactive features, such as electronic maps for way finding, dynamic information displays or animated advertisements. Understanding the attention patterns of users within these environments is important for the design and evaluation of such interactive elements. We present a modelling approach based on dynamic adaptation of the users' field-of-view, bottom-up visual saliency calculation and task-dependent semantic interest modeling that allows approximating the attention of users with regard to a photorealistic 3D-model. This improved availability of attention information can help to design more usable navigation systems, identify problems for user groups with special needs and support the design of seamless attention switches between information elements.

Author Keywords

Attention approximation; gaze modelling; navigation

Introduction

To know the mobile users' focus of attention is of high relevance for the design and evaluation of interactions in mobile contexts. Mobile eye tracking technologies have been used to address this problem, e.g. [11]. Unfortunately, eye tracking is resource intensive and has several other disadvantages (e.g. sensitive to lightning conditions, obtrusive equipment). The goal of

our work is to develop a system that is able to approximate the users' attention towards his/her environment based on typical attention patterns of mobile users, the visual saliency of the scene and the task-specific modelling of the effects of vision-based heuristics (e.g. typical color coding) on attention.

In this paper we describe the development and application of an attention model in an example scenario of a train station. The aim of this scenario is to support the design of better navigation systems, to identify problems for specific user groups early in the design (e.g. inappropriate placing of signs caused by different viewpoints of wheelchair users), and to better orchestrate and guide attention switches between information elements e.g. between dynamic screens and navigation signs by placing them in spaces, which receive attention from the same viewpoint.

Related Work
Visual saliency has been used successfully to model human attention towards images [5]. This approach provides a helpful bottom-up characterisation of the visual scene, but research has shown that it can only account for a low percentage of fixations [10], and that different types of task-based information (episodic & scene-schema knowledge [4]) are needed to successfully describe visual attention processes.

Besides studying attention in relation to static images recently also attention patterns and the relationship of body movements, head motion and gaze direction in naturalistic settings were researched [3], [11]. More specifically, also the problem of walking and the walker's need to deploy attention for navigational tasks e.g., collision avoidance, has been addressed [7]. Also, through the use of advanced scene reconstruction and

eye tracking equipment [8] more and more detailed attention data becomes available.

Jay et al. [6] have shown that in the context of dynamic content on web pages up to 80% of the user's attention can be approximated correctly. Recently attention approximation has also been applied in the context of studying way-finding problems in buildings, but the used models typically do not realistically model the variations of the field of view and the dynamics of head motions for different context situations. Only [2] considers legibility of text from different angles.

The Modelling Approach
The goal of our work is to develop a model that allows approximating human attention. Figure 1 on the next page provides an overview of the different elements. Modelling is based on several stages: First the specification of the starting conditions and environment (user, task, context, and annotated 3D-model), next the calculation of user trajectories, and finally the application of the attention model (7-9 in Figure 1).

User (1)
Basic **mobility characteristics** relevant for route-planning are defined in the user model, specifically the usage of elevators, stairs, and escalators. These user characteristics are used as input for the trajectory planning (see *Calculating user trajectories (5)* below).

Also selected **physical characteristics** of the user (eye level, max. movement speed, space requirements) can be specified in the user model in order to be able to correctly parameterize the trajectory planning.
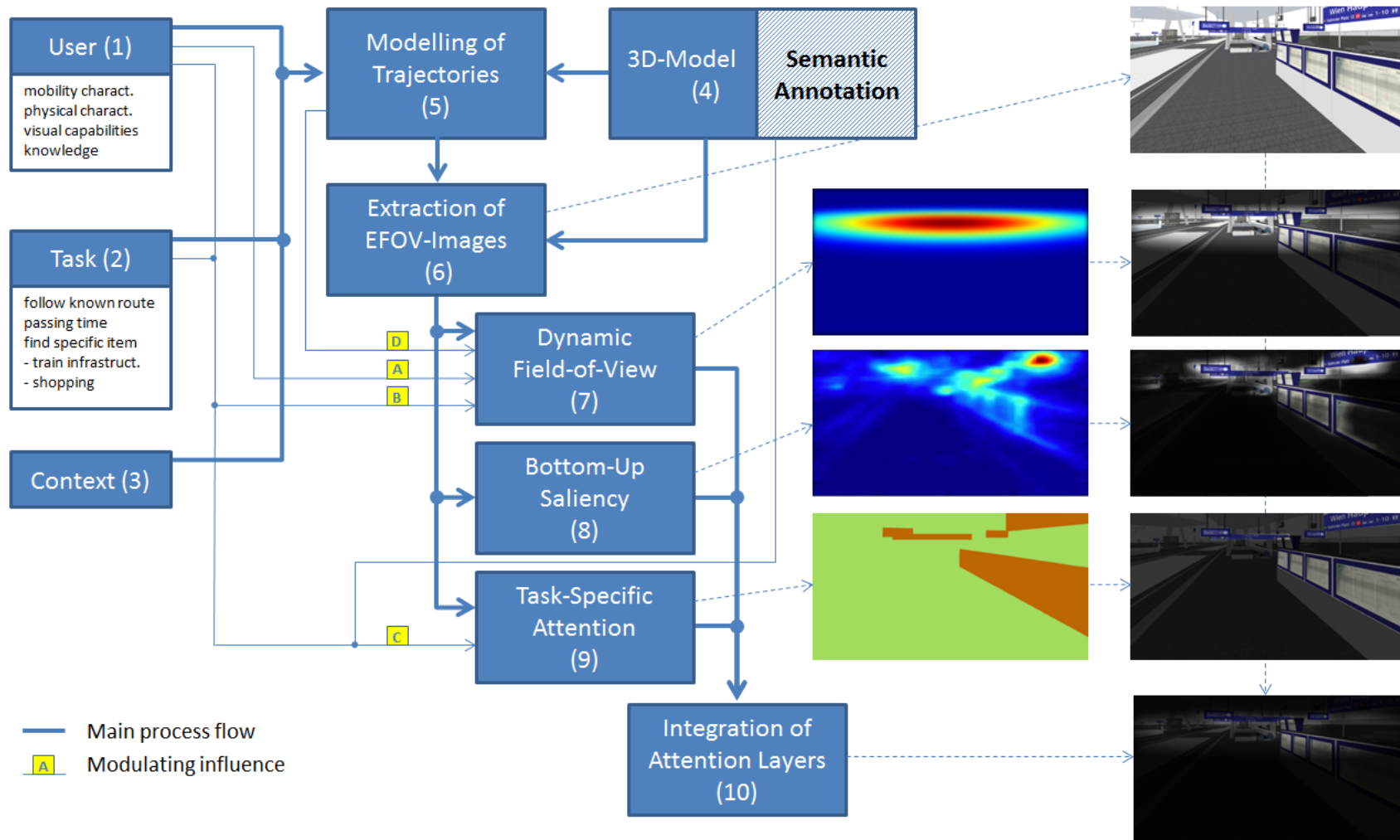
Figure 1. Overview of the model. The main processing flow starts at the top left and is indicated by bold blue lines. Slim lines indicate a modulating influence of different elements on main processing steps in the model and are marked with labels 'A' to 'D' for easy referencing. The right side of the figure shows the results of the individual processing steps applied to a sample frame taken from the 3D-model.

Furthermore relevant parameters of the **visual capabilities** of the users are specified: a) the general parameters for the users' useful field of view; this allows us to better model the attention of user populations as, for example both, useful field of view and head motion possibilities decrease with age. b) different kind of vision impairments (colour-blindness, short-sightedness); the use of specific simulators for these impairments allows to simulate their impact. E.g. for short-sightedness, elements further away can be blurred to simulate the effects of vision limitations.

Task-related **knowledge** is important for attention [4]. Especially prior knowledge of the location and generic semantic and spatial knowledge about a particular type of scene is relevant. In our model we differentiate the following types of knowledge:

a) *Location knowledge:* Does the user have any knowledge of the location i.e. can he/she use this prior knowledge to navigate to a certain point, or must the user rely on guidance signage or general heuristics to find his/her way around. We currently use 3 categories of location knowledge: no knowledge at all, passing knowledge, and detailed knowledge.

b) *Scene-schema knowledge:* Schema knowledge [4] includes information about the objects likely to be found in a specific category of scene (e.g. train stations contain navigation signs), and spatial regularities associated with a scene category (e.g. navigation signs are typically placed overheads). Therefore, whether users are familiar with a scene makes an important distinction for attention processes. In our model we currently only consider knowledge of the typical colour-coding scheme that is used. For example, experienced users will know that a specific kind of blue is always used for navigation signs in Austrian train stations.

Figure 2. View on the 3D-model from external position

Also, for many other tasks (e.g. find a certain shop) users will have specific expectations (e.g. look for the brand colours).

Task (2)
Within our target scenario we identified three main task types, which influence attention: Follow known route, find specific item, and passing time (without specific objective). Each of these task types influences interest towards elements in the environment (e.g. navigation signs, advertisements).

Context (3)
Context has been shown to be important for attention processes [10]. A first factor in our model is the density of persons in the environment, as this significantly influences the movement, visibility of elements and attention patterns of mobile users [2], [7]. A second important context factor is the condition of pavements, as this is relevant with regard to how much attention is needed for careful walking [8]. As context variables can vary depending on the current location they are expressed as 2D-layers, superimposed on the ground view of the study area.

3D model (4)
The basis for the attention approximation is a (near) photo-realistic 3D model (cf. Figure 2) of the relevant environment (in our case, a train station). The model includes all relevant structures, and the surfaces are represented in a realistic way. This model is manually annotated with semantic information driven by the tasks and the hypotheses related to top-down attention processes. Semantic annotations are stored as 2D-planes, defined by a polygon and the related code for the semantic category.
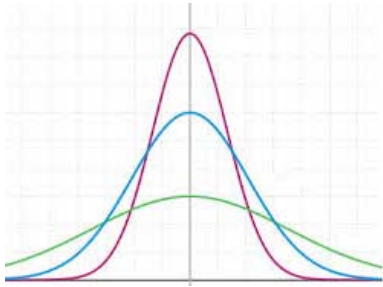
Figure 3. Illustration of Gaussian with different sigma.
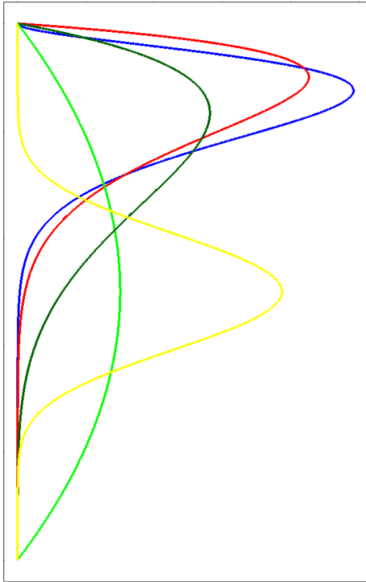
Figure 4. Illustration of beta distribution with varying parameterizations showing the flexibility of the function to map different conditions.

Calculating user trajectories (5)
Substantial advances have been made in the modelling of pedestrian behavior, a comprehensive review can be found in [1]. Also, different commercial tools are available that simulate the flow of people (e.g. Exodus, Aseri. Legion, Simwalk, PedGo). We are currently in the process of evaluating the different options with regard to their suitability for our work and the flexible integration with our attention modeling approach.

Extraction of single images (6)
Once the user trajectories are calculated, the analysis process continues with the generation of still images, which show the extended-field-of-view (EFOV) of a user along his trajectory. The EFOV is defined as the area the user might look at using head and eye movements but without changing his overall body position or orientation. EFOV-images are calculated from the 3D-model for regular time-based intervals (currently we use one second) facing towards the movement direction. This approach allows us to simplify processing as we only have to deal with one image per processing step.

Dynamic field-of-view (7)
In this step we calculate the probabilities of every point within the extended field-of-view to be looked at, based on general attention patterns of users on the move identified in related work.

The *horizontal attention* distribution within the extended-field-of-view is calculated as an overlay of typical head (relative to body/movement direction) and eye (relative to head) orientation patterns. In order to develop the model a Gaussian (cf. Figure 3) was fitted to the data from prior studies [3], [11].

According to data from literature [7] the value for sigma is slightly adjusted for different situations, such as walking speeds, tasks, people density, etc.

The *vertical distribution* of attention was modeled using data from [3]. Here the attention pattern does not approximate a Gaussian very well, and therefore was fitted to a more suitable beta distribution, which better reflects the asymmetry of data (cf. Figure 4). Again, this basic pattern is slightly modified based on findings from research, especially with regard to walking speed, density of people and pavement condition, see Figure 3 and 4 in the sidebar for details.

Bottom-up Saliency (8)
Visual saliency is calculated using the algorithm proposed by Itti et al. [5]. This method allows modeling the distribution of visual attention based on bottom-up visual features (e.g. color, intensity, orientation).

Task-specific Attention (9)
Next, the semantic annotation of the EFOV-image is analyzed and used to create a semantic-interest-map for the view. The semantic interest is influenced by the current task of the user, the semantic annotation of elements (relevant for task or not), and whether visible visual elements possess characteristics that indicate relevance for the task.

Integration of Attention Layers (10)
Finally, integrating this with the results of the other steps we receive the overall attention approximation.

Discussion and Conclusion
In this work we presented our approach to model attention patterns of mobile users. We extended existing concepts by including dynamic adaptation of the field of view of users, developed a method to consider basic task

and scene-scheme knowledge in the modelling process, and applied bottom-up saliency calculations. The results of this attention approximation process allow us to estimate 'occluded' positions with regard to different viewpoints and thereby help to improve the quality of guidance systems. The estimation of attention also allows identifying locations with high visibility, which can be used to guide the placement of advertisements and interactive features.

We expect that the application of the attention approximation methods will help to save resources in the design of information and navigation systems, as possible problems can be identified early in the design process, and because evaluations of different concepts can be done more easily. We now are in the process of optimizing and fine-tuning our work. Currently we are collecting eye tracking data for improved parameterization of the model and to compare the outcome of our approximation to data from the real world.

In the future we plan to integrate route planning and attention modelling by already considering attention information as input for the route calculations, similar to [2]. We also working on including currently unused sources of information in our model, e.g. eye-catching effects based on moving objects, face and text detection (in order to model the attention-drawing effects of these elements). We also want to include computer vision methods for scene analysis in order to automate some of the annotation work.

Aknowledgements

References

[1] Bierlaire, M., Robin, T.: Pedestrians Choices. In H. Timmermans, ed. *Pedestrian behaviour: Models, data collection and applications.* Emerald Publishing (2010)

[2] Brunnhuber, M., Schrom-Feiertag, H., Luksch, C., Matyus, T., Hesina, G. Bridging the gap between visual exploration and agent-based pedestrian simulation in a virtual environment. *Proc. VRST '12*, ACM Press (2012)

[3] Foulsham, T., Walker, E., Kingstone, A. The where, what and when of gaze allocation in the lab and the natural environment, *Vision Research* 51, 17 (2011)

[4] Henderson, J.M., Ferreira, F. Scene perception for psycholinguists. In Henderson, J.M., Ferreira, F., ed. *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, Psych. Press (2004)

[5] Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *Trans. Pattern Analysis & Machine Intelligence,* 20, 11, (1998)

[6] Jay, C., Brown, A., Harper, S. Predicting whether users view dynamic content on the world wide web. *ACM Trans. Comput.-Hum. Interact.* 20, 2 (2013)

[7] Jovancevic-Misic, J., Hayhoe, M. Adaptive Gaze Control in Natural Environments. *The Journal of Neuroscience*, 29, 19 (2009)

[8] Paletta, L., Santner, K., Fritz, G., Mayer, H., Schrammel, J. 3D Attention: Measurement of Visual Saliency Using Eye Tracking Glasses. Extended Abstracts CHI 2013.

[9] Patla, A. E., Vickers, J. N. How far ahead do we look when required to step on specific locations? *Experimental Brain Research*, 148, 1, (2003)

[10] Rothkopf, C., Ballard, D., Hayhoe, M. Task and scene context determine where you look. *Journal of Vision* 7, 16 (2007)

[11] Schrammel, J., Mattheiss, E., Döbelt, S., Paletta, L., Almer, A., Tscheligi, M. Attentional Behavior of Users on the Move Towards Pervasive Advertising Media. In *Pervasive Advertising* 2011