A Wearable Virtual Guide for Context-Aware Cognitive Indoor Navigation

Qianli Xu, Liyuan Li, Joo Hwee Lim, Cheston Tan, Michal Mukawa, Gang Wang
Institute for Infocomm Research, Agency for Science, Technology and Research
1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore
{qxu, lyli, joohwee, cheston-tan, stumam, gswang}@i2r.a-star.edu.sg

ABSTRACT

In this paper, we explore a new way to provide context-aware assistance for indoor navigation using a wearable vision system. We investigate how to represent the cognitive knowledge of wayfinding based on first-person-view videos in real-time and how to provide context-aware navigation instructions in a human-like manner. Inspired by the human cognitive process of wayfinding, we propose a novel cognitive model that represents visual concepts as a hierarchical structure. It facilitates efficient and robust localization based on cognitive visual concepts. Next, we design a prototype system that provides intelligent context-aware assistance based on the cognitive indoor navigation knowledge model. We conduct field tests to evaluate the system's efficacy by benchmarking it against traditional 2D maps and human guidance. The results show that context-awareness built on cognitive visual perception enables the system to emulate the efficacy of a human guide, leading to positive user experience.

Author Keywords

Indoor navigation; cognitive spatial model; context-aware; wearable camera; visual perception.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces – *interaction styles*; H.1.2 Models and Principles: User/Machine Systems – *human factors*.

INTRODUCTION

Indoor localization and navigation have great potential for personal assistance and context-aware services [23, 31]. Existing technologies are mostly based on sensing specific signals, such as, GPS, infrared, WIFI, and RFID [21]. There are several challenges faced by these sensor-based technologies in indoor environments [16, 31], e.g., the high cost of infrastructure construction, the accuracy of localization, and the reliability of signals in complex indoor environments. Moreover, sensor-based localization methods normally anchor the user's location on a 2D map, which requires the users to interpret the map and associate it with the egocentric perception of the environment [18].

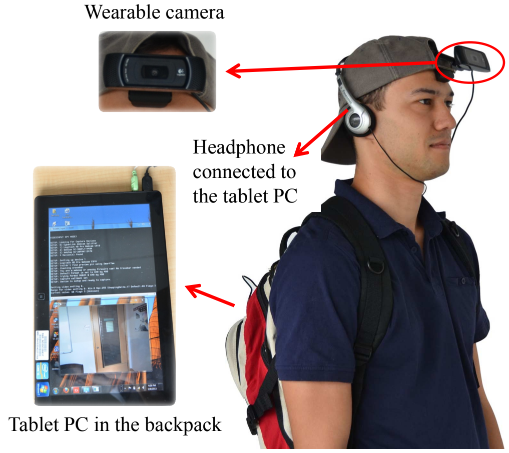
Wearable camera
Headphone connected to the tablet PC
Tablet PC in the backpack

Figure 1. Prototype system for a wearable navigation guide based on cognitive vision for indoor wayfinding

Recently, advancements in wearable cameras (e.g., Google Glass) and mobile computing provide a way to achieve an egocentric perception from the first-person-view (FPV) vision [19, 24]. Unlike sensor-based systems, a vision-based system does not require signal sending devices. Rather, it only requires a camera to receive the visual information. Therefore, it alleviates the need of special building infrastructure for spatial sensing and enables hands-free operations. However, light-weight wearable devices are not sufficient. From a user-centered perspective, there are two major challenges to build such a system: (1) the representation of navigation knowledge that is consistent with the human cognition of wayfinding, and (2) the design of interactions that specify the cognitive communications between the system and the user.

One could resort to human intelligence in assisting wayfinding when designing an intelligent, vision-based navigation guide. If a human usher accompanies a visitor to an indoor destination, she will first (often implicitly) retrieve knowledge of the building and destination, and forms a mental model of the route. Then, along the route she will give instructions typically with reference to some spatial positions (e.g., "*Go through the glass door and turn left*") based on the egocentric visual perception of the current scene, as well as the knowledge of the route [3].

MobileHCI 2014, September 23–26, 2014, Toronto, ON, Canada.
Copyright © ACM 978-1-4503-3004-6/14/09...$15.00.
http://dx.doi.org/10.1145/2628363.2628390

This makes the instructions effective and incurs less mental effort, thus reducing the visitor's stress in a new environment. Existing sensor-based methods could not provide such navigation service because they do not possess the cognitive knowledge of wayfinding built on an egocentric perception of scenes.

The motivation of this research is to provide aided wayfinding using a wearable virtual guide. Inspired by the cognitive process of human-aided wayfinding, we intend to build a cognitive model of knowledge representation for indoor navigation and develop methods to provide context-aware navigation instructions. We propose a hierarchical cognitive model aligned with the concepts of egocentric visual scenes to represent the cognitive knowledge of a building. Based on this model, vision-based indoor navigation is defined as a state transition problem where a complex route is represented as a sequence of trip segments. These trip segments embody the cognitive concepts of indoor scenes. Further, by analyzing the human behaviors in indoor wayfinding, we design navigation instructions based on the egocentric perception of scenes and the changing contexts along a route. Such a capacity makes the system adaptable to the status of the user and the conditions of the environment, which is known as context-awareness [23]. Moreover, this research proposes a novel interaction protocol for indoor navigation that consolidates multiple aspects of contexts, including the physical space, time, and user's cognitive ability. Therefore, the navigation instructions are designed for human cognitive states based on egocentric perception. Hence, we call it cognitive context-aware indoor navigation.

To evaluate the effectiveness of the proposed model, we built a prototype system and conducted user studies involving a few navigation tasks in a complex office building. It is shown that the system was able to provide accurate and robust navigation aid. Moreover, it was perceived as a useful and intelligent tool that enhanced user experience, similar to a human guide in certain aspects.

In this paper, we focus on the contexts related to human cognitive knowledge of wayfinding in indoor environments based on egocentric visual perceptions. In comparison with existing indoor navigation tools, notably those built on mobile devices [6, 7, 9, 28, 31, 32, 38], our system has two main advantages. First, owing to the cognitive concepts of indoor scenes, our method provides more comprehensive navigation knowledge beyond simple decision points. It also alleviates the need of landmarks or fiduciary markers for localization. Second, it is wearable and supports hands-free interaction for context-aware navigation. In this regard, the main contributions of this paper are:

1. A new model for representing egocentric-based cognitive knowledge for indoor navigation;
2. A novel interaction protocol for context-aware navigation that employs egocentric views.

RELATED WORK

Wayfinding and Navigation

Wayfinding and navigation is a persistent requirement of daily life and has received much attention from cognitive scientists and interaction designers [35]. Traditionally, a user uses a map to find the way in a new environment. Recent advances in wayfinding and navigation resorted to spatial models that accommodate an egocentric perspective and human cognition [2, 21, 22, 31]. Irrespective of the medium used for spatial representation, three sources of information are involved in a wayfinding model [17], namely, the environment (physical reality), the topological representation in the form of a map, and the mental representation (cognitive knowledge) possessed by an agent. An agent perceives the egocentric view of the environment and associates the egocentric observations with the allocentric representation for spatial localization and wayfinding. Such a process usually relies on various mental capabilities and efforts [17].

According to the classical cognitive model of wayfinding [40], three levels of cognitive knowledge are required: (1) *survey knowledge* (knowledge about the environment), (2) *route knowledge* (knowledge about the path), and (3) *destination knowledge* (knowledge about the target). The survey knowledge is considered as a mental representation of a familiar area, including information about regions, routes, topography, and landmarks. The route knowledge is the mental representation of a path to a destination that is beyond the immediate field of vision. The destination knowledge is about landmarks along the path to the destination. Many theories have been proposed, focusing on the relationship between the environments and agents [17, 35, 40]. For example, the route knowledge is represented as a sequence of decision points/action pairs, where a landmark is usually associated with a decision point [35].

Existing cognitive models for wayfinding mainly deal with the topological representation of the environment and design of landmark/action pairs for route directions in outdoor environments [35]. Current vision-based navigation methods require a sequence of front views along a route in memory and perform image-matching for navigation [25]. As such they are not computationally effective for indoor navigation because: (1) few salient landmarks are available indoors [37], and (2) there is a misalignment between the allocentric nature of the model and the egocentric perception of the user. Due to these limitations, existing systems usually provide sparse navigation guidance at specific points (e.g., landmarks). It is difficult to provide navigation assistance in real-time to accommodate the user needs. For example, it may require a user to self-locate herself based on the current scenes [3]. It is desirable to develop a computational model for indoor wayfinding based on the egocentric perception and mental concepts of indoor scenes. Such a capacity enables the cognitive-level communication between a system and a user in navigation.

Context-aware Interaction

Context can be defined as characterizing the situation of entities that are relevant to the interaction between a user and an application [14]. A context-aware system can adapt the information content to the users' status [27]. Early work on context-aware interaction focused on mobile computing that was sensitive to the location of users and objects [26, 36]. Recently, more advanced sensors, powerful mobile devices, and enhanced computational intelligence have enriched the implications of context. The scope of context has been expanded to include the physical environment, users' social setting [8], their activities [5, 27], past experience [15], time, weather, the display medium [12], etc. Of particular interest is the user's cognitive ability and familiarity with the environment [6, 9]. In fact, context-awareness was considered as an essential element for building the information models in navigation applications [26]. A few systems with context-awareness include the Aixplore [1], GUIDE [11], HIPS [34], etc. However, in the respective work, the navigation information was mostly tailored to the capability of the system (e.g., its sensibility and accuracy) rather than the user-specific factors.

Some systems have been built to customize interaction modalities and information content to the user's needs [13, 39]. However, these systems usually address only the static and long-term user needs, which are not particularly useful in real-time task-specific navigation. Some researchers proposed activity-based navigation [7, 32], which resorted to recognizing users' activities (e.g., walking or standing still) to inform the format and content of instructions. However, these simple activities may not effectively reflect the user's actual needs. Moreover, for interaction design, users are expected to find navigation instructions on the displays of mobile devices, which require additional cognitive efforts and shifts of attentions. Finally, it requires the user to hold the device upright in most circumstances [31]. Therefore, it is suitable for sporadic usage rather than long-term and frequent use.

METHODOLOGY

Cognitive Knowledge Representation

To achieve egocentric perception of an indoor environment with a wearable vision system, we propose a hierarchical context model for spatial perception based on cognitive conceptualization of FPV indoor scenes (Figure 2).

The model has three layers. At the top layer, a root node is used to represent a specific building. At the next layer, the nodes represent functional areas in the building. In the subsequent case example, it contains three nodes that represent the following cognitive spatial concepts: (1) *Shopping* area, (2) *Transition* area (paths connecting the shopping area and other functional areas), and (3) *Office* area. The bottom level contains sub-classes of locations in each area. For example, the conceptual locations in the *Shopping* area are *Lobby*, *MainEntrance*, *Shop*, *GateToLiftLobby*, etc.; the locations in *Transition* area are *LiftLobby*, *InLift*, *Entrance*, etc.; and the *Office* area has *MeetingRoom*, *Junction*, *Corridor*, *Entrance*/*Exit* etc. The nodes within each level are connected if there is a direct path between them, as indicated by the dash lines in Figure 2. For example, in layer II, the nodes of *Shopping* and *Transition* are connected, and in layer III, the *MainEntrance* and *Lobby* are connected.

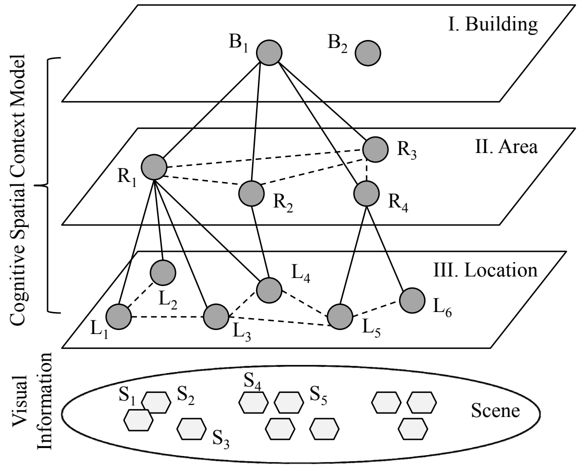

Figure 2. A structure of cognitive spatial context model

The above spatial concepts and their relations collectively constitute the cognitive spatial context model. It represents the survey knowledge for wayfinding. Further, a scene layer is defined as visual information in the form of image sequences (videos) (Bottom of Figure 2), which correspond to the nodes in the cognitive spatial model. In other words, these scenes can be mapped to the location/area/building concepts as defined above. The mapping relationship is established using scene recognition based on image classification algorithms [29].

Next, a cognitive route model can be generated from the hierarchical context model for a specific navigation task in the following process. Given an origin and a destination, two chains of nodes can be generated in the area and location levels, respectively. That is, the route model consists of a chain of area nodes and a chain of location nodes. The area nodes are connected to their child location nodes according to the hierarchical structure. Importantly, we define a trip segment as all information (including the visual scenes) associated with two connected locations. In this way, navigation is reduced to a state transition problem, where a complex indoor route is represented as a sequence of trip segments.

In an actual navigation task with a given origin and destination, scenes (i.e., image sequences) are captured continuously using the wearable camera. The scenes are categorized into area nodes and location nodes, which are compared with the nodes in the cognitive route model. Once a match is found between a recognized node and that in the cognitive route model, specific navigation instructions are retrieved according to predefined rules.

Interaction Design with Context-awareness

This research accounts for three types of contexts for the design of context-aware cognitive navigation service. First, the system recognizes the egocentric scenes so that it achieves localization of the user in the environment. This is called the *spatial context*. Second, the system must determine the appropriate time when the related navigation instruction is provided. Instructions given too early or too late may confuse the user, resulting in reduced usability of the system [4]. This gives rise to the consideration of the *temporal context*. Third, the system is designed to be adaptable to the user's cognitive status and capability, such that instructions are given only when the user needs them. Too few instructions may lead to poor task performance; and too frequent instructions may make a user feel bored or agitated. We consider an individual's status and spatial cognition capability as the *user context*.

Data collection

To align the different types of contexts with the indoor scenes along trip segments, we conducted an experiment to collect visual information related to the typical indoor segments and the actual user behaviors. The experiment involved subjects who were new to the building to go to three destinations in sequence: (1) from the main entrance outside the building (denoted as M_0) to a meeting room M_1 in the *Office* area, (2) from M_1 to meeting room M_2, and (3) from M_2 to meeting room M_3. M_0~M_3 are located at different floors and different parts (towers) of the building. Six subjects were involved in the experiment.

At the beginning of the experiment, a subject's self-reported ability of spatial cognition was collected using the Santa Barbara Sense-of-Direction (SBSOD) scale [20]. The SBSOD score is used to adjust the system's behavior to accommodate the subject's cognitive spatial capability. Before each task, a subject was informed of the next destination and was asked to find it. Along the route, the FPV video was recorded using the vision system, which comprises a webcam connected to a tablet PC running Windows 7 OS. The head-mounted webcam captures color images with a resolution of 640×480 pixels at 8 frames per second. These images are streamed to the tablet PC using a USB cable. In the meantime, a human usher who was familiar with the route accompanied the subject and provided assistance when needed. Such a need is recognized if a subject explicitly said that s/he needed help, or if the usher observed subtle signs of confusion of the subject. Notes were taken of the times and locations of the need for assistance.

Localization using cognitive visual scenes

Videos captured in the navigation tasks were analyzed to build the task-specific route knowledge. Sample scenes in the route from M_0 to M_1 are shown in Figure 3, with 10 scenes/locations belonging to 3 areas (i.e., *Shopping*, *Transition*, and *Office*), respectively. The trip segments, including the scenes, constitute the video-based, task-specific domain knowledge.

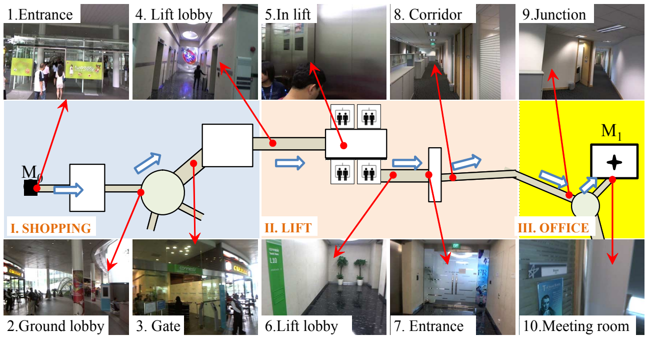

Figure 3. Scenes (images) and locations along a task route (M_0-M_1)

Once the route model is built, it can be used to determine the location of a user in an actual navigation task. Specifically, when a user goes along a route, the visual scenes are captured continuously and are used to infer the locations. Two cues are used to determine the location, namely, image categorization and time. The research adopts the extended spectral regression method to infer the location category based on image sequences [29]. This method uses a few learning algorithms to map the raw low-level image features into a manifold subspace with reduced dimensionality. It significantly reduces the load of feature computing, making it possible to recognize scenes using a lightweight computational device. According to [29], the image-driven localization can achieve 84.1% accuracy.

To further improve the robustness of localization, our model includes temporal inference (time). This is intended to tackle the dynamics of a wayfinding task. For example, it takes less time if a user walks faster than others; and the time in a lift may be unusually long when many people are entering/leaving the lift at different floors. The mean travel time of each trip segment is determined roughly on the distance of the segment. First, the time when a location (L_i) is reliably identified (based on image recognition) is denoted as t_i^0. Next, the travel time of a trip segment (T_i) is defined as the average interval between two adjacent locations (L_i, L_{i+1}), i.e., $T_i = avg\left(t_{i+1}^0 - t_i^0\right)$. The information is used in the subsequent probability model for localization (Figure 4).

Next, a probability distribution function is built for each trip segment to specify the probability that a visual scene is associated with that segment. In this study, we adopt a normal distribution $N\left(t_i, \sigma_i\right)$, where $t_i = t_i^0 + \sigma_i$ and σ_i is initialized empirically as a value that is slightly larger than the minimum travel time of the segment. It changes dynamically during a specific navigation task.

In a navigation task, after a scene is registered, the probability distribution function for the respective trip segment is retrieved and used to determine the user's location, i.e., the probability that a user is located in trip segment *i* is dictated by the current scenes and how long the scene is in the view since scene registration (t_i^0). With the increase of a user's staying time in the segment, σ_i increases gradually until the next scene is correctly recognized as the subsequent location. For example, at time t_i^0, the scene recognition module identifies the present scenes as location L_i. Accordingly, the distribution function $N(t_i,\sigma_i)$ is retrieved. Note that the distribution function of the previous location $N(t_{i-1},\sigma_{i-1})$ is still active, i.e., there is typically an overlap between the probability distribution function for adjacent locations. With the probabilities $p_{i-1}(t)$ and $p_i(t)$ computed according to the Gaussian distribution models, the current location is determined based on:

$$L(t) = \arg\Big(\max\big(Rand\big(p_{i-1}(t)\big), Rand\big(p_i(t)\big)\big)\Big) \tag{1}$$

where *Rand*(*p*) represents a random number in the range of [0, *p*]. At the time t_i^0, the likelihood of recognizing the location as *i* and *i*+1 is identical. As time passes, there is higher probability that the location is recognized as *i*+1. The mechanism is similar to dead reckoning in a typical navigation process. However, unlike activity- or movement-based dead reckoning which suffers drift and loses accuracy over time [33], the time step parameter is calibrated each time a new location is correctly recognized.

Context-aware navigation instructions

The above process allows the system to determine the location of a visitor when s/he traverses the route. The next issue is how to provide effective navigation aids based on the location information. We design an interaction protocol such that navigation instructions are given timely and unambiguously according to the cognitive needs of the user and the egocentric visual perception.

First, we define a decision point along a trip segment as the place where a user needs certain guidance. The decision points are extracted based on the subjects' behaviors in the wayfinding experiment. We found that a need for guidance may or may not be associated with an intersection. For example, when walking along a main corridor, the users did not hesitate in continuing along this direction even if there were sideways leading to an irrelevant place (e.g., a toilet or a pantry). Conversely, users may need certain kind of help even if there were no alternative routes. This happened when the user was unsure if s/he was on the correct route after a prolonged period of unassisted walking. In such a situation, it is useful to affirm the user to continue in that direction. This user-centered approach is consistent with the heuristics of instruction chunking and controlled granularity for lowering the user cognitive load in wayfinding [35]. However, by exploiting the actual user behavioral patterns, our approach alleviates the need to explicitly define the heuristics to optimize the instructions.

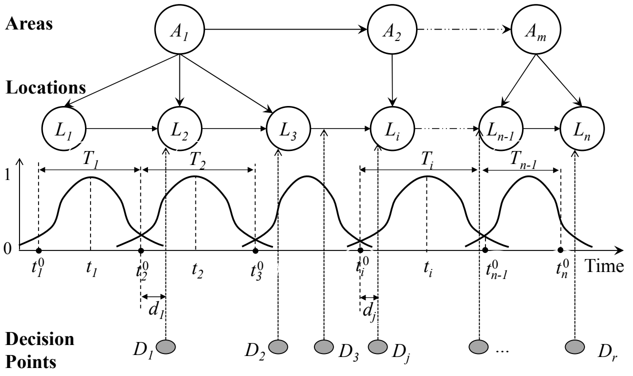

Figure 4. A stochastic model that accommodates multiple aspects of the context

Based on our experiment, if one or a few subjects explicitly requested navigation guidance or showed signs of confusion or hesitation at a position, the respective location was considered as a decision point (D_j). Such decisions reflect the cognitive need of a human subject in wayfinding. A probability value (P_j) is defined on the number of subjects (*n*) who requested aids at D_j.

$$P_j = \begin{cases} 0.2n, 1 \le n \le 5 \\ 1 \qquad n = 6 \end{cases} \tag{2}$$

This model shows that the more subjects required assistance in the vicinity of a decision point, the more likely an assistance instruction is needed. This probability will be used to determine if an instruction is given at an appropriate position. Based on the training data, 10 decision points are identified for task 1, 11 for task 2, and 10 for task 3.

Next, the starting time of the navigation instruction is determined based on the user's cognitive behavior in a trip segment. We design a model that specifies the time attribute of a decision point based on the average interval between the time when a location is identified (t_i^0) and the time of a subject's request for help. The mean value of the time interval between t_i^0 and decision point (D_j) is d_j. Accordingly, the navigation instruction should be provided at time $t_i^0 + d_j$.

To tackle the possibility that a user does not comply with the instructions or there is a significant delay in a user's action due to unforeseen circumstances, the system repeats the instructions (the instructions are rephrased to avoid being considered to be too rigid). We use TT_{j-1} to denote the narration at decision point D_j, and TT_{j-2} to denote the rephrased narration. As an example, when a user reaches the lift lobby, a voice instruction TT_{4-1} is given (Table 1). In case the user stays in the lift lobby for a time longer than expected, another instruction TT_{4-2} is given. Table 1 shows the list of instructions associated with the trip segments along the route M_0→M_1 (ref. Figure 3). These navigation instructions are generated and presented as voice instructions using the text-to-speech module.

Location	Instructions
1: Main entrance	$TT_{1.1}$*-Welcome. Let's begin. We will go to Bayes at level ten. Please walk through the glass doors.* $TT_{1.2}$*-Please go into the glass doors.*
2: Ground Hall	$TT_{2.1}$*-Please turn left towards the South Tower gate.* $TT_{2.2}$ *-Please walk to the South tower gate.*
3: South tower gate	$TT_{3.1}$*-You are facing south tower gate. Please go ahead to the lift lobby.* $TT_{3.2}$*-Please go to the lift lobby and take a lift.*
4: Lift lobby	$TT_{4.1}$*-Take a lift on the left side; and go to level ten. The lift on the right does not go to level ten.* $TT_{4.2}$*-Take a lift to level ten. Please look at the top right of the lift door to see if the lift stops at level ten.*
5: Inside lift	$TT_{5.1}$*-Please remember to alight at level ten.* $TT_{5.2}$*-Please remember to alight at level ten.*
6: Lift lobby, Info Board	$TT_{6.1}$*-Please stand in front of the information board and check if you are at level ten.* $TT_{6.2}$*-Proceed to the glass door and get in.*
7: Entrance	$TT_{7.1}$*-Turn right after you enter the glass door.* $TT_{7.2}$*-Please make a right turn after the glass door.*
8: Corridor	$TT_{8.1}$*-Go straight ahead for about 30 meters.* $TT_{8.2}$*-Please go ahead until the junction.*
9: Junction, meeting room	$TT_{9.1}$*-We are close to Bayes. It is the room on the left side of the junction.* $TT_{9.2}$*-Bayes room is around. Please check the name on the door.*
10: Meeting room door	$TT_{10.1}$*-Here we are. You have reached meeting room Bayes. Congratulations!* $TT_{10.2}$*-You have reached meeting room Bayes. Congratulations!*

Table 1. Navigation instructions along route 1: M_0-M_1

In addition, the orientation of a user is critical for correct guidance. In our system, since the scenes are defined on egocentric view, the orientation information is embedded into the scene categories. For example, when a user walks from *LiftLobby* to *Office* along the entrance, the scenes are categorized as *EntranceToOffice*; if the user walks along the same link from *Office* to *LiftLobby*, the scenes are categorized as *ExitToLiftLobby*. We have also considered this issue at decision points where different orientations will lead to different instructions. As an example, in a lift lobby, our system identifies the different information boards on the opposite sides of the lift lobby.

Finally, to account for user's mental capability of spatial cognition, we used the SBSOD score as a parameter to modulate the probability that an instruction is provided at the decision point. Let C_p be a subject's spatial cognition level which is derived from the SBSOD. At any time t_k, a navigation instruction is provided according to three rules:

1. If: $t_k = t_i^0 + d_j$ AND $P_j \geq \min\left(1, \frac{C_p - C_L}{C_H - C_L}\right)$, Then: $TT_{j\text{-}1}$;
2. If: $t_k = t_i^0 + 2\sigma_i$ AND $t_k > t_i^0 + d_j$, Then: $TT_{j\text{-}2}$;
3. If: $t_k \geq t_i + 3\sigma_i$; Then: $TT_{j\text{-}3}$.

Note that in rule 1, C_L (=0.36) and C_H (=0.67) are the smallest and largest scores of the spatial cognitive capability level, respectively, as derived from the six subjects who participated in the data collection experiment. P_j is defined in Eq. 2. This rule determines if an instruction should be given or not depending on the user's cognitive spatial ability. In rules 2 and 3, $\sigma_i = {}^{T_i}\!/_{2}$ is used to specify the time property. $TT_{j\text{-}3}$ ("*You seem to be lost. Please talk to the experimenter.*") is used to allow the user to recover from a situation when the location is persistently not identified or if it does not change for an exceptionally long time.

EXPERIMENTAL EVALUATION

We tested the context-ware cognitive navigation guide (denoted as CNG) by comparing it with two other types of navigation aids, namely, a 2D map (denoted as MAP) and a human guide (denoted as HUG). For the MAP condition, the subject was given the floor plans of the routes for which three destinations were marked out. The subjects were supposed to find the destinations without further external aid. In the HUG condition, an experimenter followed the subject and gave instructions when there was a need.

Participants

We recruited 12 participants (6 male, 6 female; mean age 27.7 years with a standard deviation of 8.6 years). The experiment was carried out for individual subjects. Each navigation mechanism (MAP, CNG, and HUG) was used once for each person. The order of the tasks was identical across all test sessions, namely, task1: $M_0 \rightarrow M_1$, task2: $M_1 \rightarrow M_2$, and task3: $M_2 \rightarrow M_3$. On the other hand, the order of navigation mechanisms was different and counter- balanced for the order of usage. This is to cancel out the learning effect, as well as the effect of task difficulty.

Experiment Design

A general procedure of the experiment is as follows. First, the SBSOD questionnaire was used to evaluate the subject's self-reported spatial cognition. The subject's SBSOD score was input into the system as a configuration parameter. Next, the subject was led to the starting point (M_0), where the wearable system was attached. Based on the predetermined order of assistance for that experiment session (Note that the order is different for different subjects), the voice feedback was provided through earphones if the CNG was to be used for the respective task. Otherwise the system ran in a recording mode.

Upon completion of an individual task, a questionnaire was given to the subject to collect the feedback on the respective navigation assistance. After three tasks were completed, a visual memory retention test was carried out whereby the subject was shown a series of 80 photos of the interior of the building. Half of the photos (40) were taken along the task route, and the other half were not. Each image was shown in the tablet screen for a maximum of 5 seconds during which the subject decided if s/he had seen the scene along the route. This was intended to estimate how much visual information the subject could remember.

Finally, the subject was requested to backtrack to the destinations ($M_3 \rightarrow M_2 \rightarrow M_1 \rightarrow M_0$). No navigation assistance of any kind was provided. The experimenter recorded the time that the subject finished each backtracking task. When the participant returned to the origin (M_0), s/he received a $15 shopping voucher as a token of appreciation.

Measures

The efficacy of the system is evaluated using both objective and subjective measures. Objective measures include:

- The time taken to complete an individual task ($O_{\_Time}$),
- Time taken to backtrack in a task ($O_{\_BTime}$),
- Memory retention of scene along the route ($O_{\_Mem}$).

For $O_{\_Time}$ and $O_{\_BTime}$, we normalized the time for each task. The normalization was intended to cancel out the effect of route length and other conditions. From the data collection experiment, we computed the mean (μ_i) and standard deviation (σ_i) of the time for each task (i=1, 2, 3). The normalized time t_i^n for a subject in task i was computed as: $t_i^n = (t_i - \mu_i)/\sigma_i$. The same normalization scheme was applied to the backtracking time.

Subjective evaluation of user experience is conducted on seven measures, each consisting of 3-4 items.

- Usefulness ($S_{\_Use}$) – how useful the guide/tool is for the navigate task.
- Ease-of-use ($S_{\_Eou}$) – how easy it is to use the guide/tool in the task.
- Cognitive load ($S_{\_Cog}$) – how much mental effort the subject has to make to complete the task.
- Enjoyment ($S_{\_Enj}$) – how enjoyable it is to use the guide/tool.
- Stress level ($S_{\_Stress}$) – how stressful the subject feels in the course of the task.
- Intelligence ($S_{\_Int}$) – to what extent the guide/tool can make decisions like a human.
- Trust ($S_{\_Trust}$) – how trustful the advices/information are, as provided by the guide/tool [30].

Hypotheses

We expect that the wearable navigation guide emulates a human guide to achieve equivalent performance with respect to both objective and subjective measures. At the same time, it should outperform the 2D map in most measures, except those related to backtracking time ($O_{\_BTime}$), scene memory ($O_{\_Mem}$), and trust ($S_{\_Trust}$). For $O_{\_BTime}$ and $O_{\_Mem}$, it is expected that the subjects who used MAP would be able to memorize more information of the route because: (1) a map provides an allocentric representation of the entire space (topological knowledge) which may lead to better spatial memory [41], and (2) the users may make more effort to observe the environment in the wayfinding [17]. In comparison, both CNG and HUG provide navigation directions and require less mental effort of the subjects. Therefore, the time taken for backtracking should be shorter in the MAP condition, and more scene images would be correctly recognized. For trust, we expect that the users will trust MAP and HUG more than CNG. This is because a map gives objective information that is supposed to be credible; a human guide with knowledge of the place is supposed to give trustful information.

Measurements		Hypotheses	
Task completion time	(O_{Time})	H1	MAP>CNG=HUG
Backtracking time	(O_{BTime})	H2	MAP<CNG=HUG
Scene memory	($O_{\_Mem}$)	H3	MAP>CNG=HUG
Usefulness	(S_{Use})	H4	MAP<CNG=HUG
Ease of use	(S_{Eou})	H5	MAP<CNG=HUG
Cognitive load	(S_{Cog})	H6	MAP>CNG=HUG
Enjoyment	(S_{Enj})	H7	MAP<CNG=HUG
Stress	(S_{Stress})	H8	MAP>CNG=HUG
Intelligence	(S_{Int})	H9	MAP<CNG=HUG
Trust	(S_{Trust})	H10	MAP>CNG<HUG

Table 2. Hypotheses (MAP - 2D map, CNG - Cognitive navigation guide, HUG - Human guide)

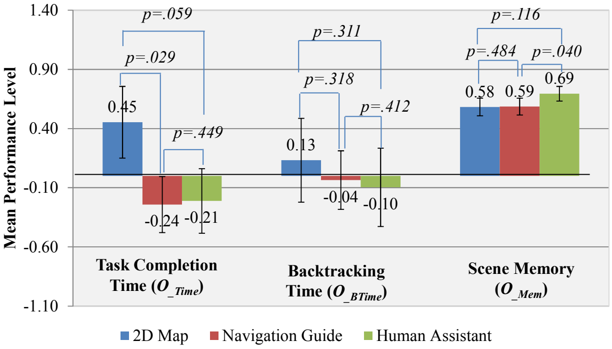

Figure 5. Comparison of task performance using objective measures.

The hypotheses in the study are presented in Table 2. The comparison is made with respect to the values of individual measurements. The formula "*X*>*Y*" means the value of the respective measurement is significantly higher in condition *X* than in condition *Y*. "*X*=*Y*" means there is no statistical difference in the measures in two conditions. For example, H1 is concerned with the task completion time ($O_{\_Time}$): MAP>CNG=HUG. It means that: (1) the task completion time is longer when the subjects use MAP than when they use CNG or HUG, and (2) CNG and HUG lead to equivalent task completion time.

RESULTS

Task Performance

We analyzed the effects of different navigation methods using one-way ANOVA with repeated measures for three objective measurements ($O_{\_Time}$, $O_{\_BTime}$ and $O_{\_Mem}$). Post- hoc analysis with paired *t*-test was conducted with a Bonferroni correction (a significance level set at $p < 0.017$). It is expected that the task performance is higher when subjects use CNG and HUG than when they use MAP. In addition, CNG and HUG would lead to identical performance. Hypothesis 1 (Task completion time) was partially supported, whereas H2 (Backtracking time) and H3 (Scene memory) were not supported. The results are shown in Figure 5.

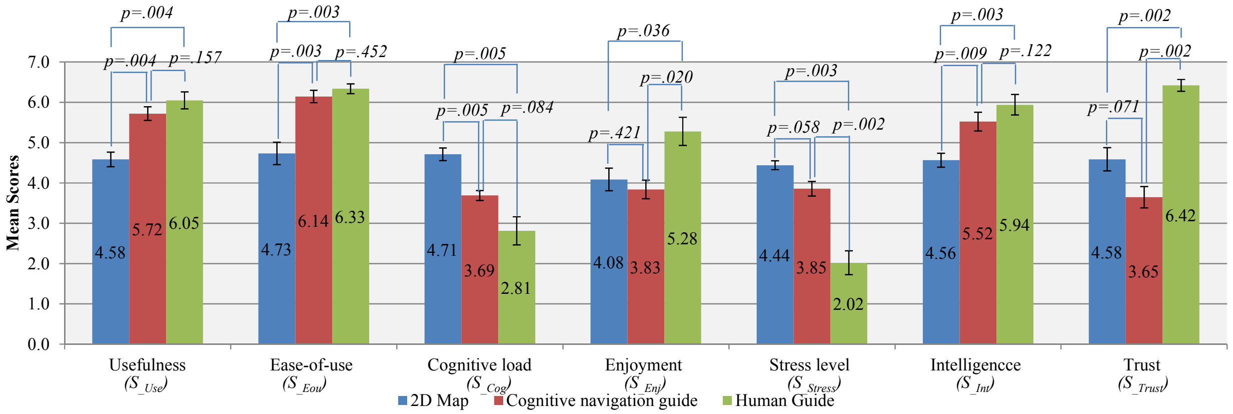

Figure 6. Comparison of user experience using subjective measures.

First, the task completion time was only marginally different among three conditions (*F*(2,22)=3.40, *p*=.051). Post-hoc test shows that it took slightly less time for the subjects to complete the navigation task using CNG than using MAP: $O_{\_Time}$ (MAP, CNG): *t*(11)=2.114, *p*=.029. There was no statistical difference between CNG and the HUG [$O_{\_Time}$ (CNG, HUG): *t*(11)=—0.131, *p*=.449], nor between MAP and HUG: $O_{\_Time}$ (MAP, HUG): *t*(11) = 1.694, *p*=.059.

Furthermore, no statistical difference was found among three conditions with respect to the backtracking time [$O_{\_BTime}$: (*F*(2,22)=0.567, *p*=.575] and the scene memory [$O_{\_Mem}$: (*F*(2,22)=1.907, *p*=.172]. For the backtracking task, the subjects generally completed the task fairly fast as seen from the near zero mean values of the normalized backtracking time. Similar observation was made for scene memory testing, whereas we found that the memory retention of the visual scenes was slightly worse in the CNG condition than in the HUG condition [$O_{\_Mem}$ (CNG, HUG): *t*(11)=—1.978, *p*=.040]. Overall, the task performance among the three conditions did not show much difference. The reasons will be discussed later.

Subjective Evaluation

We conducted Friedman test with post-hoc analysis using Wilcoxon signed ranks with Bonferroni correction (a significant level set at *p*<0.017) for evaluating the subjective measures. We found significant differences in all seven subjective measures across three conditions. Post-hoc analysis showed a positive effect of the CNG's context-aware capability on the user experience as compared to the MAP, except for the measurements of enjoyment, stress, and trust (Figure 6). In particular, users perceived the cognitive navigation guide to be more useful than the 2D map [$S_{\_Use}$ (CNG, MAP): *z*=2.849, *p*=.004]; it was easier-to-use [$S_{\_Eou}$ (CNG, MAP): *z*=2.937, *p*=.003]; it reduced users' cognitive load [$S_{\_Cog}$ (CNG, MAP): *z*=—2.82, *p*=.005]; and it was considered to be more intelligent [$S_{\_Int}$ (CNG, MAP): *z*=2.608, *p*=.009]. However, the enjoyment level was identical in both conditions [$S_{\_Enj}$(CNG, MAP):*z*=—0.804, *p*=.421]. No statistical difference was found with respect to the stress level [$S_{\_Stress}$ (CNG, MAP): *z*=—1.896, *p*=.058] and trust [$S_{\_Trust}$ (CNG, MAP): $S_{\_Trust}$: *z*=—1.806, *p*=.071.

As compared to the human guide, the CNG achieved equivalent performance in terms of usefulness [$S_{\_Use}$ (CNG, HUG): *z*=—1.414, *p*=.157], ease-of-use [$S_{\_Eou}$ (CNG, HUG): *z*=—0.751, *p*=.452], cognitive load [$S_{\_Cog}$ (CNG, HUG): *z*=1.730, *p*=.084], and intelligence [$S_{\_Int}$ (CNG, HUG): *z*=—1.548, *p*=.122]. It was inferior to a human guide in terms of perceived stress [$S_{\_Stress}$ (CNG, HUG): *z*=3.056, *p*=.002] and trust [$S_{\_Trust}$ (CNG, HUG): *z*=—3.070, *p*=.002]. A marginal difference was observed in favor of the human guide for the enjoyment level [$S_{\_Enj}$ (CNG, HUG): *z*=2.317, *p*=0.020].

DISCUSSION

From the evaluation, it was found that the wearable navigation guide with context-awareness effectively realized cognitive navigation. The measures for which the cognitive navigation guide emulated a human guide were related to the functional aspects of the system. In particular, its performance in usefulness, ease-of-use, intelligence, and cognitive load was equivalent to the human guide. Thus, it is functionally competent in providing navigation guidance for the tasks. This result shows that it is technically feasible to realize real time cognitive indoor navigation based solely on the FPV-based vision system.

The absence of differences in the performance indicators, namely, task completion time, backtracking time, and scene memory was unexpected. The effect of navigation tools seemed to have played little role in the users' memorability of the route and visual scenes. This can be explained in two aspects. First, in a new environment, the subjects' arousal level is typically high, i.e., they paid much attention to observe the scenes when traveling in the building, irrespective of the navigation tools used. Second, the cognitive load for wayfinding and scene memory might be confounded. In fact, the subjects needed to make much effort to reckon the correct route when using a 2D map (which can be seen from the higher level of cognitive load). Thus, the resource allocated to scene memory is reduced.

The lack of difference between MAP and CNG in enjoyment ($S_{\_Enj}$) turned out to be surprising. We had expected that the MAP, as a passive tool, would lead to lower enjoyment level.

However, the absolute value of enjoyment was higher in MAP (M=4.083, SD=0.975) than in CNG (M=3.830, SD=0.797). We suspected that the subjects might have confused the enjoyment of using the device with that of the task. It has been shown that a moderate level of challenge in completing a task may lead to positive enjoyment [10]. Considering that the MAP condition incurred more cognitive load, the subjects might have enjoyed meeting the challenge of the navigation task.

Further, we predicted that users have less trust in the CNG than in the MAP. This hypothesis was not supported with statistical significance. Nevertheless, with a relatively lower level of trust, there was a sign of users' reluctance in trusting the CNG, even though the device did not make any apparent mistakes. Some subjects showed their concerns during the post-experiment survey, i.e., they wondered how the CNG were able to determine the location and give instructions solely using a camera. Thus, they thought the navigation instructions were hardcoded and might be faulty. Such observations gave rise to an important design implication, i.e., unfamiliarity and lack of knowledge of the system may negatively influence users' trust. It is advisable to give users basic knowledge of the system in order to enhance trust and acceptance.

On the other hand, the vision system did not achieve a level of performance that is equivalent to a human guide with respect to enjoyment, stress level, and trust. These measures are largely concerned with the emotional aspects of users' experience. This can be partially attributed to the scope of the study, i.e., to ensure accurate and timely navigation based on the perceptual ability of the system. The design of context-awareness emphasized more on enhancing system performance than on delivering enjoyable experience. This inadequacy can be compensated by adding more features that enhance the users' emotional engagement.

Another observation is related to system reliability. There were no apparent errors during the navigation tasks in all three conditions. However, when using the 2D map, subjects occasionally paused and looked at the map for directions. This might have contributed to the slightly longer task completion time using the map. On the other hand, a few errors happened during the backtracking task when the subjects deviated from the correct route; and the number of errors were similar in three conditions (MAP: 2; CNG: 2, HUG: 3). It might have contributed to the larger variances in backtracking time, thus causing the lack of statistical difference in backtracking time.

Limitations and future work

Benchmarking strategy: In this research, the proposed system is benchmarked against a 2D map and a human guide, rather than with the latest technologies built on smartphones [6, 7, 9, 28, 31, 32, 38]. The reasons for choosing these two benchmarking navigation mechanisms are: (1) they are the prevalent navigation aids that are familiar to the subjects, (2) these two methods require different mental efforts in wayfinding, and (3) smartphone-based navigation systems are still not available in the test building. Moreover, the current work focuses on the technical feasibility of a wearable cognitive usher, i.e., the benefit over traditional maps and its competency compared to human. A comparison with smartphone-based navigation technologies will lead to details of usability evaluation, which is beyond the scope of this research. Notwithstanding, an apparent benefit of using the wearable device is its support to hands-free operation in navigation.

User-specific context: The current research has a simplified mechanism to deal with the cognitive context related to the user. In the proposed model, we trained the model to estimate the decision points where users were likely to request assistance. This strategy could only deal with the "average" needs of users. It falls short as an adaptable mechanism to account for the individual needs of users. In the future, we will collect visual cues that directly reflect a user's needs based on his/her activities [32], e.g., head movement patterns for showing hesitation, eye gaze for indicating intentions, etc. The ability to assess user's cognitive states more accurately and adaptively will endow the system with refined context-awareness for desired personalization. In addition, the current application required additional effort to collect the subject's SBSOD score. However, this is solely for purpose of experimental validation. In future, the respective feature may be designed as a customized function, i.e., users may optionally provide their level of spatial cognition. If so, the system will adapt its feature based on such information. Otherwise, the system provides standard instructions as if the user has "average" cognitive ability.

Scalability and learning: A few parameters of the proposed system are determined by the training data from six videos, and the current test includes a few specific routes. It might not support navigation along unexplored routes. However, the proposed cognitive spatial model is not contingent on the specific routes and can host alternative routes of interest. Data collection at a larger scale can be carried out by observing users' activities along other routes such that the navigational information can be incorporated into all the trip segments. Moreover, the system can be expanded to the entire building provided that wearable cameras, such as, Google Glass, become prevalent. Possibly, during the system development stage, visitors to a host building may voluntarily participate in the program. A glass-type device can be leased out and the navigation process can be recorded. Different routes will be covered in this process. The navigation information can be aggregated, such that route segments may be extracted from users who follow different routes. Therefore, the need for a huge data set for each and every route is exempted.

CONCLUSION

Wearable vision systems provide tremendous opportunities for real-time and context-aware services, e.g., navigation. In this research we investigated if vision-based wearable guides with proper cognitive knowledge and interaction modalities are beneficial to the users. We addressed two critical issues in the implementation of such a system, namely, the representation of cognitive knowledge for wayfinding based on egocentric visual information and the design of cognitive interaction with context-awareness. The hierarchical spatial information structure establishes the foundation for cognitive visual information processing. The interaction model with context awareness serves as a useful basis for developing personalized navigation services. Our research leads to useful implications for the design of wearable vision systems in navigation assistance in indoor environments.

ACKNOWLEDGMENTS

This research is partially supported by the Singapore A*STAR, SERC grant on HFE (#092 1530095).

REFERENCES

1. Aixplorer. http://www.aixplorer.de/.
2. Arikawa, M., Konomi, S.I. and Ohnishi, K. NAVITIME: Supporting pedestrian navigation in the real world. *Pervasive Computing*, *6*, 3, (2007), 21-29.
3. Baras, K., Moreira, A. and Meneses, F., Navigation based on symbolic space models. in *IPIN'10*, (2010), 1-5.
4. Barberis, C., Bottino, A., Malnati, G. and Montuschi, P. Experiencing indoor navigation on mobile devices. *IT Professional*, *14*,1, (2014), 50-57.
5. Bardram, J.E. Activity-based computing for medical work in hospitals. *ACM T. Comput-Hum. Int.*, *16*, 2, (2009), 10:11-36.
6. Baus, J., Krüger, A. and Wahlster, W., A resource-adaptive mobile navigation system. in *IUI'02*, (2002), 15-22.
7. Brush, A.J., Karlson, A.K., Scott, J., Sarin, R., Jacobs, A., Bond, B., Murillo, O., Hunt, G., Sinclair, M., Hammil, K. and Levi, S. User experiences with activity-based navigation on mobile devices *MobileHCI'10*, ACM Press (2010).
8. Bulling, A., Weichel, C. and Gellersen, H., EyeContext: Recognition of high-level contextual cues from human visual behaviour. in *CHI'13*, (2013), 305-308.
9. Butz, A., Baus, J., Krüger, A. and Lohse, M., A hybrid indoor navigation system. in *IUI'01*, (2001), 25-32.
10. Chatting, D.J., Action and reaction for physical map interfaces. in *TEI'08*, (2008), 187-190.
11. Cheverst, K., Davies, N., Mitchell, K. and Friday, A., Experiences of developing and deploying a context-aware tourist guide: The guide project. in *MobiCom'00*, (2000), 20-31.
12. Cheverst, K., Davies, N., Mitchell, K., Friday, A. and Efstratiou, C., Developing a context-aware electronic tourist guide: Some issues and experiences. in *CHI'00*, (2000), 17-24.
13. Davies, N., Cheverst, K., Mitchell, K. and Efrat, A. Using and determining location in a context-sensitive tour guide. *Computer*, *34*, 8, (2001), 35-41.
14. Dey, A.K., Abowd, G.D. and Salber, D. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction*, *16*, 2, (2001), 97-166.
15. Etzion, O., Skarbovsky, I., Magid, Y., Zolotorevsky, N. and Rabinovich, E., Context aware computing and its utilization in event-based systems. in *DEBS'10*, (2010), 270-281.
16. Fallah, N., Apostolopoulos, I., Bekris, K. and Folmer, E. Indoor human navigation systems - A survey. *Interacting with Computers*, *25*, 1, (2013), 21-33.
17. Freksa, C., Klippel, A. and Winter, S. A cognitive perspective on spatial context. in Cohn, A.G., Freksa, C. and Nebel, B. eds. *Spatial Cognition: Specialization and Integration*, Dagstuhl, 2007.
18. Giudice, N.A., Bakdash, J.Z., Legge, G.E. and Roy, R. Spatial learning and navigation using a virtual verbal display. *ACM Trans. Applied Perception*, *7*, 1, (2010), No. 10.
19. Hanheide, M. *A Cognitive Ego-Vision System for Interactive Assistance*, Ph.D Thesis, University of Bielefeld, Bielefeld, 2006.
20. Hegartya, M., Richardsona, A.E., Montellob, D.R., Lovelacea, K. and Subbiah, I. Development of a self-report measure of environmental spatial ability. *Intelligence*, *30*, (2002), 425-447.
21. Heiniz, P., Krempels, K.H., Terwelp, C. and Wüller, S., Landmark-based navigation in complex buildings. in *IPIN'12*, (2012), 1-9.
22. Hile, H., Grzeszczuk, R., Liu, A., Vedantham, R., Košecka, J. and Borriello, G. Landmark-based pedestrian navigation with enhanced spatial reasoning. *Lecture Notes in Computer Science - Pervasive Computing*, *5538*, (2009), 59-76.
23. Hong, J., Suh, E.-H. and Kim, S.-J. Context-aware systems: A literature review and classification. *Expert Systems with Applications*, *36*, 4, (2009), 8509-8522.
24. Kanade, T. and Hebert, M. First-person vision. *Proceedings of the IEEE*, *100*, 8, (2012) 2442-2453.
25. Kaneko, Y. and Miura, J. View sequence generation for view-based outdoor navigation. in *ACPR'11*, (2011), 139-143.
26. Kenteris, M., Gavalas, D. and Economou, D. Electronic mobile guides: A survey. *Pers. & Ubiquit. Comput.*, *15*, 1, (2011), 97-111.
27. Kjeldskov, J. and Paay, J. Indexicality: Understanding mobile human-computer interaction in context. *ACM Trans. Computer-Human Interaction*, *17*, 4, (2010), 14:11-28.
28. Kray, C., Elting, C., Laakso, K. and Coors, V., Presenting route instructions on mobile devices. in *IUI'03*, (2003), 117-124.
29. Li, L., Goh, W., Lim, J.H. and Pan, S.J. Extended spectral regression for efficient scene recognition. *Pattern Recognition*, *47*, 9, (2014), 2940-2951.
30. McKnight, D.H., Carter, M., Thatcher, J.B. and Clay, P. Trust in a specific technology: An investigation of its components and measures. *ACM T. Inform . Syst.*, *2*, 2, (2011), 1-15.
31. Möller, A., Kranz, M., Huitl, R., Diewald, S. and Roalter, L. A mobile indoor navigation system interface adapted to vision-based localization, in *MUM'12*, (2012), No.4.
32. Mulloni, A., Seichter, H. and Schmalstieg, D., Handheld augmented reality indoor navigation with activity-based instructions. in *MobileHCI'11*, (2011), 211-220.
33. Mulloni, A., Seichter, H. and Schmalstieg, D., Indoor navigation with mixed reality world-in-miniature views and sparse localization on mobile devices. in *AVI'12*, (2012), 212-215.
34. Opperman, R. and Specht, M., A context-sensitive nomadic exhibition guide, in *HUC'00*, (2000), 127-149.
35. Richter, K.-F. and Klippel, A. A model for context-specific route directions. in Freksa, C. et al.(eds). *Spatial Cognition IV. Reasoning, Action, and Interaction*, (2004), 58-78.
36. Schilit, B. and Theimer, M. Disseminating active map information to mobile hosts. *IEEE Network*, *88*, 5, (1994), 22-32.
37. Snowdon, C. and Kray, C., Exploring the use of landmarks for mobile navigation support in natural environments. in *MobileHCI '09*, (2009), No. 13.
38. Taher, F. and Cheverst, K., Exploring user preferences for indoor navigation support through a combination of mobile and fixed displays. in *MobileHCI'11*, (2011), 201-210.
39. Wang, S.-L. and Wub, C.-Y. Application of context-aware and personalized recommendation to implement an adaptive ubiquitous learning system. *Expert Syst. Appl.*, *38*, 9, (2011), 10831-38.
40. Wiener, J.M., Buchner, S.J. and Holscher, C. Towards a taxonomy of wayfinding tasks: A knowledge-based approach. *Spatial Cognition and Computation*, *9*, 2, (2009), 152-165.

41 Willis, K.S., Hölscher, C., Wilbertz, G. and Li, C. Comparison of spatial knowledge acquisition with maps and mobile maps. *Computers, Environment and Urban Systems*, *33*, 2, (2009), 100-110