An important activity in urban three-dimensional (3D) mobile navigation is browsing the buildings in the environment and matching them to those in the 3D city map. There are different factors affecting the recognition process such as changes in the appearances of buildings, weather, and illumination conditions. The current aim was to study the salience of different types of visual cues in the recognition of buildings in 3D maps in suboptimal conditions. A pilot laboratory experiment was conducted, in which test participants recognized buildings in a 3D city map using systematically prepared photographs as stimuli, and their cognitive processes were studied using the think aloud protocol. The results suggested that buildings in a 3D city map can be recognized based on a variety of different visual cues ranging from small details such as textual signs to the shape of the building and landmark features such as towers. The results also suggested that buildings are recognized relatively much based on their location and other buildings and objects in their surroundings.

1. INTRODUCTION

Mobile navigation systems have existed for a while and are now commonly used in personal and professional navigation. One of the most important recent trends in the area of mobile navigation has been the introduction of 3D graphics and photorealistic views. Mobile 3D maps and navigators such as Google Earth and Navitime [1] have enabled their users to view 3D city maps either remotely or while navigating to a destination. Photorealistic systems such as Google Street View are becoming increasingly popular. Together with these recent developments in location-aware computing and 3D mobile graphics, 3D mobile maps have become an important research area, and the volume of related research has also increased.

Much of the existing research on mobile 3D maps has concentrated on technological aspects such as methods for rendering 3D environments or interaction aspects such as controlling the viewpoint in 3D mobile maps [2, 3]. In the area of building recognition, automated methods for identifying buildings from images have been studied more than the corresponding human visual cognition. For example, Zhang and Kočeká [4] developed a hierarchical approach for automatic recognition of buildings in an urban area. There are still important aspects of mobile 3D navigation, which have been studied hardly at all. For example, the salience of different kinds of visual cues in 3D city maps has not been paid much attention to. It is known that landmarks are important in mobile navigation [e.g. 5], and Bessa et al. [6] found that buildings act as important cues in the recognition of urban scenery, but they did not study different visual cues in buildings more closely. Thus, studying visual cues in photorealistic 3D buildings seems justified. A fully photorealistic mobile 3D map would require adapting the rendered output according to the physiological characteristics of the human vision system (HVS). However, a fully photorealistic and interactive 3D map is not feasible on a mobile device yet. Selective rendering techniques attempt to focus the available computational resources on the most important parts of the scene at hand [7]. Visual attention is classically divided to two processes, the stimulus driven bottom-up and task driven top-down. Both these processes, involuntarily and voluntarily, attempt to spot the relevant parts from the scene. In a navigation situation, where the 3D environment represents the immediate surroundings, these cues are extracted and matched in both environments. This process is essential for self-orientation [8].

It has not yet been proven that a photorealistic 3D environment would provide the best support for 3D navigation. Given the limited computational resources of mobile devices, identification of the relevant visual cues in a scene could be utilized to create a more resource aware rendering system, if not providing even more efficient navigation. Unfortunately, not much research has been done on the most relevant visual cues in the context of navigation. Bessa et al. [6] took photographs from an urban environment, and participants, dislocated nearby, were asked to point the spots where the photographs were taken from, and were requested to mark the real world objects assisting in these.
tasks. They found that only a few features were used, along with the overall geometry, and that these features were not the most salient ones to the HVS. Manipulation of the environment or the photos did not alter the performance, although the positioning error slightly increased. Their study, however, concentrated on studying only the objects based on which the locations were recognized, not their specific details or characteristics such as color, shapes, or textures. In the current work, we contribute to this research area by presenting results from an experiment, in which we aimed at obtaining an understanding of which kind of visual cues are the most salient in 3D city maps. We especially concentrated on studying the recognition of 3D buildings.

2. EXPERIMENT

2.1 Participants

There were 9 participants (4 females and 5 males, mean age 36.0 years, range 24-61 years). The participants’ previous experience of using mobile navigation systems ranged from low to high and their familiarity with the city of Tampere also ranged from low (an occasional visitor living elsewhere) to high (living in the city). The younger participants’ familiarity with the city was higher as compared to the older participants’ familiarity with the city. The participants received a movie ticket as compensation for their participation in this research.

2.2 System

The experimental system featured the Tre3D model of the city of Tampere and the m-LOMA mobile 3D map system, which is one of the first full-featured mobile 3D maps [9]. m-LOMA, mobile LOcation-aware Messaging Application, delivers a real urban virtual environment to mobile users, with optimized wireless networking and efficient rendering. It allows entire cities to be rendered in real-time in mobile devices, even those lacking 3D hardware. It supports both vector and raster 2D maps, and navigation features such as GPS and routing. It also acts as a platform for location-dependent information. The application runs on many different devices including mobile phones, PDAs, and personal computers. The Tre3D city model was originally developed in the “3D City Info for mobile users” project [10]. The photorealistic style of describing the content of the database of city information was chosen because the whole system was intended to be as credible as possible, describing a real place (the city of Tampere). In the project, the assumption was that with a photorealistic style to describe the city, the users are better supported in navigation and finding information, because they can recognize places more easily [10, 11]. These two systems were combined for the current research, and the participants used the system with a Nokia N95 smart phone with a 2.6” display and a resolution of 240 x 320 pixels. The system was fast enough to display the 3D model in real time. The 3D model was deliberately dated a few years back (it was not exactly up-to-date), in order to study the recognition of buildings in the context of both correct and missing features. Most of the changed features were stores and signs, but it some cases also colours and the exact shapes of the building had changed. The granularity of details in the 3D model was about 10-20 cm.

2.3 Tasks and materials

The experimental tasks consisted of 17 tasks for each test participant. Each task consisted of browsing the 3D city model using the navigational features of the m-LOMA system and recognizing a correct building based on a photograph of the scene to be found. The buildings were mostly modern or classical highrise buildings, but special purpose buildings such as the railway station, the city hall, and an old church were also included. The order of buildings presented was preselected and they were distributed to different parts of the city center (area approximately 1000m x 500m) so that buildings in subsequent tasks were not right next to each other or in the near vicinity. The experiment started from the central square. The times the pictures were taken ranged from daytime to early evening, and they represented the city in early winter. Examples of pictures shown for the participants in the tasks and the corresponding 3D map views are presented in Figure 1 below.

2.4 Procedure

The current experiment was conducted as a pilot laboratory experiment in order to study salient cues in 3D city maps without the effects of environmental variables present in field experiments. When the participant entered the laboratory, the researchers introduced the experiment, after which the m-LOMA system and the controls for moving were introduced. The Nokia N95 arrow keys were used for moving forwards and backwards and turning around in the 3D city map, and the participants could also control the viewpoint on three different axes (height, tilt, and step left/right) using the number keys of the smart phone. Street names (or any other augmented information) were not displayed in the 3D map to support the current focus of studying salient visual cues of 3D buildings. When the actual experiment began, the participants were shown a picture of the building to be recognized and their task was to find the picture using the m-LOMA system. The photographs were projected onto a screen near the participants. The stimulus photographs and the key map of the m-LOMA controls were available for the user during...
the tasks. The navigation task was completed, when the participant had found the correct building pictured approximately from the same angle as in the stimulus picture. The participants were instructed to think aloud during the experiment specifically in relation to the following aspects in different phases of each experimental task: • Attention. Which kind of visual cues drew your attention in the stimulus photograph? • Recognition. Which kind of visual cues enabled the recognition of the correct building in the 3D city map? • Improvements. Which cues (inaccurate or missing) would help in recognizing the building, if improved? The experimental time ranged from 40 to 90 minutes per participant depending on their speed in learning the use of the m-LOMA system, carrying out the navigational task and the amount of the qualitative comments given using the think-aloud protocol.

2.5 Data analysis

The qualitative research data were analysed separately in relation to the three above mentioned research questions studying attention, recognition, and improvements. First, the relevant categories for the different kinds of visual cues data were identified in the whole data using affinity diagramming. As a result of this analysis, ten categories for visual cues emerged: landmarks, signs with text, other details, building shapes, colours, textures, building as a whole, part of the building, location/environment, and augmented information. Second, all the data were analysed once more and the data related to all the three research questions were coded into the categories, which were identified in the first phase of the analysis.

3. RESULTS

The results regarding the salience of different kinds of cues in the different phases of the research (cues drawing attention in the photograph of the building, cues which enabled the recognition of the correct building in the 3D map, and suggestions for improvements) are presented in Table 1.

Table 1. Visual cues mentioned in the data by category

<table>
<thead>
<tr>
<th>Category</th>
<th>Attention</th>
<th>Recognition</th>
<th>Improvements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landmark</td>
<td>13 (8.9%)</td>
<td>0 (0.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Sign with text</td>
<td>42 (28.8%)</td>
<td>13 (17.8%)</td>
<td>46 (23.5%)</td>
</tr>
<tr>
<td>Other detail</td>
<td>46 (31.5%)</td>
<td>7 (9.6%)</td>
<td>60 (30.2%)</td>
</tr>
<tr>
<td>Building shape</td>
<td>12 (8.2%)</td>
<td>11 (15.1%)</td>
<td>30 (15.1%)</td>
</tr>
<tr>
<td>Colour</td>
<td>8 (5.5%)</td>
<td>5 (6.8%)</td>
<td>5 (2.5%)</td>
</tr>
<tr>
<td>Texture</td>
<td>2 (1.4%)</td>
<td>5 (6.8%)</td>
<td>11 (5.2%)</td>
</tr>
<tr>
<td>Building as a whole</td>
<td>2 (1.4%)</td>
<td>3 (4.1%)</td>
<td>3 (1.5%)</td>
</tr>
<tr>
<td>Part of building</td>
<td>1 (0.7%)</td>
<td>1 (1.4%)</td>
<td>3 (1.5%)</td>
</tr>
<tr>
<td>Location/environment/nearby building</td>
<td>20 (13.7%)</td>
<td>17 (23.3%)</td>
<td>39 (19.6%)</td>
</tr>
<tr>
<td>Augmented information</td>
<td>-</td>
<td>-</td>
<td>5 (2.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>146 (100.0%)</td>
<td>73 (100.0%)</td>
<td>199 (100.0%)</td>
</tr>
</tbody>
</table>

Most of the qualitative comments given could be easily coded into the developed categories. Examples of visual cues in the categories included “the city hall tower” and “the watchtower” (landmarks), “the glass window”, “the clock on the wall”, “Xmas lights”, and “the balcony” (other detail), “the textures of the wall” (textures), and “the colours of the surface materials”
(colours). Other examples include “the location near the bridge” (location/environment), “the appearance of the upper part of the building” (part of building), and “a railway station icon” and “street names in the 3D map” (augmented information).

4. DISCUSSION

In the current experiment, we studied the salience of different kinds of visual cues in the context of 3D city maps. The result suggested that in urban buildings (represented by photographs in this experiment), the users pay relatively much attention to details, especially to signs with text. Details became important again, when the participants were asked, which kind of improvements they would suggest to support better recognition. In building recognition, the location, nearby environment, and building shapes and landmarks also become important. By highlighting the importance of recognizable details, these results give preliminary support for the use of photorealistic textures and views in mobile 3D navigation. At the same time they stress the importance of updating the map materials constantly in order for the details to be correct. On the other hand, the results suggest that the users can also use cues such as the shape of the building or other buildings in the environment to aid the recognition. Theory-wise, these preliminary results most closely match the currently prevailing viewer-centered theory of object recognition [12], which states that humans store many different views and angles of an object and can recognize the object through a general analysis of its whole and its parts. Future research is needed to study application of this theory in mobile 3D navigation more closely.

In this study, the aim was to obtain an overview of different visual cues used in the recognition of 3D buildings in different conditions. For that purpose, we varied a number of variables such as the level of correctness of the visual cues in the buildings and the time of the day (daytime – evening) in the photographs. We conducted this initial experiment in a laboratory, which enabled controlling environmental factors (e.g. social context) affecting building recognition in mobile contexts. Furthermore, one typical way of using mobile 3D maps is in the context of preparing remotely for a trip, for example, finding an interesting destination (with a picture) in an advertisement and then finding its location using the 3D city map – a procedure quite similar to that used in the current study. The current work represents work in progress aiming at understanding the most important visual cues in mobile 3D navigation, especially in the context of 3D city maps. Future work includes a corresponding study in a mobile context with a larger number of participants. We believe that the current work may act as an important step towards new kind of original research, which has been largely omitted in the existing research on 3D mobile navigation.

5. ACKNOWLEDGMENTS

The authors would like to thank all the test participants and the reviewers of this paper for their constructive comments.

6. REFERENCES