Visual data mining aims at finding the hidden potential of data. In this paper, functionalities for a visualization tool, CommonGIS, are developed. The focus is on visualizing spatiotemporal data. Two data sets were collected, both containing location data of several users or entities and various events attached to these entities. Different techniques for bounding the data sets for a more illustrative visualization were derived. The results are shown on a user interface. The tool revealed important information from both data sets. Also, it performs well regardless of the diversity of the events in different data sets.

1. INTRODUCTION

Visual data mining aims at finding the hidden potential of data. Visual data mining has been studied mainly for time invariant data. Demographic data sets have been studied in [1], [2], and [3], among others.

Geographic Information Systems are tools for efficient analysis- and visualization in landscape architecture. There are several such tools available, [4], and [5], for example. We have chosen the CommonGIS [6], developed by Andrienko for our research. The CommonGIS has a core one cannot change, but it is flexible and easy to expand for different purposes. In their research, Andrienko et. al have tested geo-visualization in outlier removal, visual comparison, dynamic classification, dynamic query, and dynamic linking, for example. They also have a web application for simple spatiotemporal data mining, where they follow the flight paths of storks [7]. In our research, we develop methods for spatiotemporal visual data mining.

In this paper, we have two data sets available for visual study. The main data set has been collected during a field trial on Rotuaari, a pedestrian street in the city center of Oulu, Finland. There are three main themes in the Rotuaari testbed, the development and implementation of a service platform, field testing with end users, and a value chain: technology providers contribute to the process with their (business) activity. Other similar mobile service test-beds are Testbed Botnia [8] and Elisa mobilemall [9].

The other data set has been collected at the campus area of the University of Oulu. The data is related to the development of a context-aware service architecture [10], and routine learning algorithms [11]. Both environments enable the collection of the location data of the users among other measurements. The tool developed here will aid in data mining by allowing an analyst to visually determine the interesting points from the data. Clearly, the project partners and especially the businesses around the Rotuaari can benefit from such a tool, also.

2. DATA

2.1. Rotuaari Field Trials

The first field trials for Rotuaari project were carried out during 28.8.-30.9.2003. The tested applications were map-based guidance, mobile advertisements, and TimeMachine Oulu, which is an interactive three-dimensional virtual model of historical Oulu. The headquarters for the field trials was a building initially meant for Oulu city guidance and info. The field trials were advertised on the setting and in newspapers.

The basis for data collection in Rotuaari is a service architecture SmartRotuaari [12]. The architecture is a distributed platform for mobile context-aware services. The users were ordinary shoppers in Rotuaari area, and they were given PDA’s for their use for two hours. In the PDA’s, Smart-Rotuaari client offered the services. All in all, there were eight PDA’s available for the customers. Each Compaq iPAQ has a Wireless LAN card. A Wireless LAN covers the Rotuaari area and the location of the user is achieved using Ekahau [13], which tracks the location of the Wireless LAN card (the location of the iPAQ). The users were sent mobile advertisements according to profile and proximity criteria. The user could either accept or discard these advertisements. The content providers for the advertisements consisted of several (over 20) shops, restaurants, companies etc. along the Rotuaari street.

The data measured during a trial is stored into an XML document as key-value pairs. The measurements are grou-
aped as attributes of entities, where entity is one of User, Place, or Advertisement. An example of a log entry for one User is presented, as follows:

```xml
<logentry>
  <header>
    <date>30-09-2003T14:29:44</date>
    <module>
      <name></name><version></version>
    </module>
    <session>
      <id>216</id>
      <username>seppo</username>
    </session>
  </header>
  <body>
    <userAttributeChangeEvent>
      <location>
        <longitude>25.468917078116988</longitude>
        <latitude>65.0110523987453</latitude>
        <altitude>0.0</altitude>
        <floor>0</floor>
      </location>
    </userAttributeChangeEvent>
  </body>
</logentry>
```

Login events of the users are saved, but no logout information is maintained. The location information is saved as latitude-longitude coordinates. The log files include information about the advertisements also: when it is sent, when it is read, or if it is discarded. Also, the "mood" of the user is saved. This is an attribute that the user sets himself. The moods are normal, hungry, party, shopping, lonely, sightseeing, busy, business. Furthermore, weather information for the test days is available.

2.2. Office space data

Another test data was collected in the premises of the University of Oulu. A wireless LAN covers the whole office space of our research group, and the location of users can be obtained as well as in Rotuaari. Also, the testees had a PDA profiler to change the profile of the PDA into Silent, Meeting, Outdoors, or General modes depending on the location. Three persons collected data for this experiment for over 60 hours. They changed the profile of the PDA like it would happen in real life with an ordinary mobile phone. A more detailed description can be found from [14].

3. PHASES OF DATA VISUALIZATION

We propose functionalities for a visualization tool to study spatiotemporal data sets. In our point of view, the phases of visual data mining are described in Figure 1.

Visualization starts by loading data. After loading, constraints are defined for the raw data set. These constraints determine the subset, the active data set to be actually visualized. As the data is time-variant, an interval is an obvious constraint. The area to be studied can be constrained as well. An important data type to be visualized is an event informing that something has happened. The user can e.g. change the mood to "hungry" or send a message to her friend. Or, a mobile advertisement from a pizzeria can be received by the user’s mobile device.

The raw data set is constrained by specifying start and stop conditions for data. The conditions are applied for each user’s data set separately, for each data set in the order the data was measured. Data is included into the active set from the first datum fulfilling the start conditions till the datum preceding the first datum fulfilling the stop conditions. For example, the start condition can be "User is inside area A" and the stop condition "Event B occurs". Primitive conditions can be combined with Boolean operations, for example "After time T and inside area A". Although simple, this representation offers a wide set of possibilities for constraining the data set.

Any type of events can be visualized. The XML format of the data makes it possible to list the event types to the user although the tool itself does not associate any meaning to the different types. This feature enables a general tool that can be used to visualize a wide set of data. As long as there is a location associated with an event, it can be shown on a map.

After determining the active data set, the user selects one of the visualization operations. In addition, the representation of the data can be changed. Location measurements can be either shown separately or they can be grouped. Grouping the measurements of one user produces a path. Another
grouping method is to define a matrix and show the number of measurement for each cell. This clearly shows the areas where the users have moved the most. Or, if mobile advertisements have been selected to be shown, it can be seen where the ads have been received. A key characteristic of this visualization model is that all changes made by the user affect instantaneously to the visualization seen by the user. The different bounds and the representations are described in Table 1.

<table>
<thead>
<tr>
<th>Constraints</th>
<th>Representations</th>
</tr>
</thead>
<tbody>
<tr>
<td>No limit</td>
<td>Symbol</td>
</tr>
<tr>
<td>Time</td>
<td>Route</td>
</tr>
<tr>
<td>Area</td>
<td>Cell</td>
</tr>
</tbody>
</table>

Table 1. The different constraint criteria and representations developed for the CommonGIS.

There are over ten different combinations to visualize the data sets.

4. RESULTS AND DISCUSSION

Several different possibilities for defining the boundaries and active operations for visual data mining were developed.

The first phase in visualization is loading the data. In Figure 2, the location data of all test users in Rotuaari is shown. From this picture, it is pretty clear that the visualization must be bounded somehow.

After selecting the data set, boundaries for selecting interesting (from the user point of view) data are set. In Figure 3, the loaded data are the location events from the measurement period, and the advertisements received for a single test user. The user can select different starting- and ending criteria for the visualization. The starting conditions are no limit, time, advertisement received, and area visited. These same criteria can be set for the stopping conditions. In Figure 3, the starting criteria is Rosso’s (a restaurant) advertisement received (symbol star), and the ending criteria area visited. The area is shown as a box in the figure. Here, the user has selected the interesting advertisements to be shown, and probably wants to see, if the user went to the restaurant after receiving the ad. The user has received two Rosso’s advertisements and has entered the neighborhood of the restaurant in both cases.

The visualization of the office space data is carried out in a similar manner as for the Rotuaari data. Although, here the interesting events are the PDA profile changes. It would be interesting to see, if there are some patterns in changing the PDA profile based on location. In Figure 4, the profile changes of one test user are shown. It can be seen that the meeting profile has been set on most likely in the meeting room, while in the corridors the profile has been set to general mode. The meeting room is on the left upper corner of the figure and the symbol representing the meeting profile is a circle. In the figure, the profile general is represented with a square. This kind of information could enable the automatic profile configuration of the PDA (or other mobile device). The profile could automatically be switched to a meeting mode when entering the meeting room, for example.

Another interesting viewpoint in visualizing this kind of location and event data is to present the information quantized. This can be achieved by forming a cell representation for the map. The number of users, the time spent in a cell, or the number of events in the cell can be color coded. This way, it is possible to determine which cells are the most active. Also, it would be interesting to determine in what order are the cells activated (the routes of the users).

In Figure 5, the location events of all users are shown in an above type of cell representation. The user can adjust the size of the cells, the transparency of the cells, and change the maximum value for the color code scaling.

5. CONCLUSIONS AND FUTURE WORK

In this paper, functionalities for the visualization of spatiotemporal data were developed. The tools were built on CommonGIS, an open, object-oriented, distributed system providing knowledge-based GIS services. The user can perform visual data mining on data if location data is available. The data is described as XML with key-value pairs. The visualization tool identifies events (attributes) from the data, and can visualize them on a map if location information has been attached to the entity. In our research, we have used two different data sets for testing the tool, the Rotuaari data and the office space data. Both data sets have their own characteristics and events, but the tool developed here performed well on both sets.

The visualization tool was tested within the Rotuaari project group, which includes research on human computer interaction, marketing and education, also. The visualization lead to new ideas on developing their own research, and they had a better understanding on all of the data they are able to process and analyse. It was also suggested, that a special version of the tool would be delivered to the businesses along the Rotuaari street. In this tool, the number of users that have and have not visited a shop after receiving a mobile advertisement, are shown. The design of this special version is already finished, and it will be implemented as a Java Applet, so it is will be easy to use anywhere.
Later on, we are about to develop methods for visualizing events that do not have location information attached to them. One possibility is to visualize the events on a time segment of a line. Furthermore, it would be interesting to identify certain routines of the user/users. Routines could include the gradients of mutual movements of several users, and the information on different routes when time goes by. Then, the events could be displayed as part of the route.

In another project, we have studied the routines of mobile users with the office space data. Important locations and profile changes have been recognized from the data. The visualization tool can be expanded so that these routines could be shown along with the raw data. Different symbols could be assigned to the routines, and the user could configure her own devices based on the information that has been learned from her behaviour and shown to her. Other interesting development would be animation of the events.

ACKNOWLEDGMENT

Jaakko Hietajärvi is acknowledged for implementing the visualization tools. This work was funded by TEKES. The authors would like to thank all the personell in Rotuaari project.

6. REFERENCES


Fig. 2. Location data for all users.

Fig. 3. Location data visualized as routes.
Fig. 4. The profile events for one test user. Profile events are visualized with different symbols.

Fig. 5. The number of location events are color coded into a matrix representation of the map.