A Mixed Reality Interface for Real Time Tracked Public Transportation

Antti Nurminen1*, Juha Järvi,1 Matti Lehtonen1
1Helsinki Institute for Information Technology HIIT, of Aalto University and University of Helsinki, Finland
*antti.nurminen@aalto.fi, Otaniemietie 19-21, 02150 Espoo, Finland

Abstract
Real time tracking of public transportation with open interfaces unfolds interesting possibilities for potential use cases. We present work towards full mixed reality interfaces, where virtual vehicles are a) embedded in a mobile 3D map, depicting a real city, and b) visualized directly with augmented reality on top of a real world view. Both these graphical application types aim at the same goal, reducing the cognitive distance between representation and the real world in contrast to traditional visualization, i.e. 2D maps. Our main contribution is the technical achievement of combining both representations into one application, with smooth visual transitions between them, together with the capability of representing the dynamically tracked entities in real time.

Keywords: REAL-TIME DATA, OPEN DATA, MIXED REALITY

Introduction
In the framework of public transportation, data visualization typically focuses in traffic management analysis using traditional 1D and 2D plots, charts and map interfaces [1, 2, 3]. On rare occasion, advanced 4D visualization methods have been applied, for example to incident analysis, to provide a better overall situational view and management of temporal data [4]. For individual travelers, journey planner web sites with map support are nowadays a commonplace. Requirements for such planners are well known, where complexity of data, especially in the case of multimodal routing, often leads to data reduction, leaving out features such as pedestrian guidance and navigational detail [5]. However, as the area of public transportation management is maturing, advanced features have begun to emerge. One of these is accurate real time tracking, improving perceived service reliability [6]. However, visualization of such data still relies on traditional means. For example, the Lissu service by the city of Tampere, Finland, is based on a traditional map interface, lacking small scale visual features (see Figure 1).1

Our research contribution for traffic visualization aims at exactness and detail both spatially and temporally, where, in contrast to abstraction, we provide visualization aiming at full realism in real time. Our primary case is an urban environment, namely the city of Tampere, Finland, dense with visual and navigational detail. We build our use cases to minimize the mental effort required for perceptual challenges of self-orientation and proximal visual search.

1Lissu service ©CGI Finland, map data ©MML 2014.
Mixed reality applications provide a basis for embedding content perceptually within the environment, either by direct augmentation (augmented reality, AR) or via a 3D model of it (augmented virtuality, AV). Both these application types use world as the user interface, thereby potentially minimizing the need for spatial inferencing and cue-based triangulation for simple short-distance target finding tasks.

In the following, we analyze the properties and challenges of conventional two-dimensional spatial interfaces, augmented reality and augmented virtuality. We proceed with analysis on relevant available local Open Data sources and interfaces. Then, based on our visual interface analysis, we build our use cases where mixed reality applications have most potential for end users. We then describe our technical implementation.

Traditional Spatial Interfaces

Traditional spatial interfaces rely mostly on two-dimensional representations, which can be cartographically abstracted or, as is becoming a common alternative, aerial photographs. Such representations are best at displaying configurational spatial information, which suits for example route planning. Although early maps were populated with intuitive figures, modern maps depend highly on compact symbols, annotations that require a legend, which explains the various visual conventions [7]. These conventions are often culturally bound, and visitors may need to put some effort in map reading.

When one is orienting him or herself with the world using an external information source, he/she first needs to match cues between the source (the map) and the target (the real world). Typically, this is challenged by the drastically different viewpoints and view content. An abstract map or even an aerial photo do not natively share many directly recognizable visual cues with the ground level view of the observer. An aerial photo lacks
building façades and their salient features. For an abstract map, the cartographer has pre-designed the set of cues suited for the expected spatial task. In the case of pedestrian navigation, street names are often given precedence. These cues may or may not be efficient, depending on context, task and situation. For example, road names need to be searched from the physical environment. After this initial search, they become more efficient during navigation.

A GPS can assist in localization, and a magnetometer (compass) in orientation; but still, a set of matching cues are needed to establish a mental scale. For a local person, who already holds an accurate mental map, this orientation process is fast, if even necessary.

**Augmented Reality**

In augmented reality (AR), artificial, computer generated content is used to enrich or annotate the real world. AR techniques are inherently egocentric. Augmentations lie either directly in the view of the user (using see-through displays) or on top of a camera view of a mobile device [8]. An AR system has three key characteristics: (a) it mixes real and virtual imagery, (b) registers the digital data to the real world and (c) provides interactivity in real-time [9]. One of the main challenges is accurate registration, where the virtual content is accurately positioned onto the real world and tracked. Two fundamental methods for this are commonly utilized: 1) sensor-based registration and 2) computer vision based registration. For the first case, spatial sensing is utilized to solve the current pose of the viewer, for example using embedded GPS and orientation sensors (see Fig. 2, left). Typical computer vision based AR systems use fiducial markers for tracking (see Fig. 2, right). Recently, research has focused in markerless environments. These approaches have their pros and cons. Sensor-based registration is globally applicable, but depends on sensor accuracy. For markerless recognition, visual features from areas and targets to be recognized need to be pre-extracted.

![Figure 2: Sensor-based mobile AR (from [10]) and AR with a fiducial marker (from [11]).](image)

Further challenges in AR are related to perception, especially the hard problem of inter-position [12]. There is no depth information in the video feed of a mobile device, and when virtual content is simply rendered on top of the view, human vision cannot directly resolve proper depth order; the content appears to be in front of everything.

**Augmented Virtuality: 3D Maps**

In the case of augmented virtuality, a virtual environment is augmented with data emerging from the real world. However, as our virtual environment actually represents the real
world, we prefer to describe this case as a 3D map. 3D maps rely on 3D models, which are built to resemble the real environment. They could be abstracted with similar task and usage based goals as traditional 2D maps, but in our case, we have set the goal to be realism. Figure 3 presents the case for Tampere, Finland, with a traditional 2D map, focusing the attention to names and labels (left). The 3D map, using the Tre-D model, focuses attention to the accurately textured façade surfaces (right).

Figure 3: A 2D map with bus stops as red dots (©City of Tampere and NLS Finland) and a 3D map of the same area on our platform with a view from East to West.

In a realistic mobile 3D map, the primary technical challenges are in rendering a potentially very large, detailed 3D environment with limited resources. Our original research solved this problem by an elaborate optimization scheme [13]. While the in-core resources of current mobile devices have significantly improved, the issues of scalability and efficient use of battery power remain.

In contrast to augmented reality, 3D maps do not suffer from many of the perceptual issues. Interposition is resolved with the well known Z buffer solution as part of 3D rendering. Registration is resolved automatically as the city model and augmentations maintain the same coordinate system.

Interactivity has long remained a challenge in mobile 3D maps. In an earlier experiment, 3D maneuvering with simple hardware buttons was found cumbersome [14]. Along with spatial sensing and touch screens, 3D manipulation has now become easier. For example, interacting with a 3D map can follow the same pointing paradigm as the AR with spatial sensors. Here we have found that such a method applies well to on-site browsing, while the common touch screen based pan&zoom paradigm suits off-site tasks [15].

**Tampere Use Cases**

Spatial tasks that involve long distances and understanding of spatial configurations, for example route planning, benefit from traditional visualization methods. On the other hand, in the short range, proximal visual cues and visible content might be better expressed with mixed reality techniques.

Our case city, Tampere, Finland, is a medium sized city with some 220,000 residents. The central plaza, "Keskustori", of Tampere, contains over 10 proximal bus stops (see Figs. 1 and 3). Most Tampere region buses drive via this area. For a casual visitor, the situation may be complex. Even with the help of a journey planner, one may easily fall to an exhaustive strategy to find the correct bus stop by physically searching through each sign.
for the appropriate bus line number. In addition, buses are arriving from three different
directions, occluding each other.

We will focus on two simple cases: a) locating a bus stop and b) spotting an arriving bus. We assume a general proximity.

**Open Data**

Governments and municipalities are increasingly making public transportation data available on the Internet. Such Open Data interfaces provide this information freely without restrictions from copyright, patents or other mechanisms of control [16]. As typical use cases, this data can be used to explore and characterize current and historical service levels or to forecast operations in the immediate future [17]. Opening previously closed data sources or creating such services for the public comes with a cost. However, there are several occasions where there are clear benefits, such as creating more compelling advanced searches, mashups with other data sources and data visualization [18, 17]. Our work falls within the last category.

**OpenStreetMap**

OpenStreetMap (OSM) is an appraised case of crowd-sourcing for collecting spatial data. It follows the nearly 2000 year old idea of separating data from its visualization [19]: contributors are encouraged to focus on creating accurate real world data, while other tools are developed to yield polished renderings. The project is maintained by over 15000 active users and anyone can register for free to start modifying the data.

The OSM database consists of **nodes**, **ways** and **relations**, all of which can be enriched by associating them with any number of **tags**, key-value pairs. A node is simply a pair of coordinates. A way is an ordered list of nodes and is used to define any complete or partial physical feature that is formed by connecting all consecutive points together. A relation is an ordered list of nodes and ways that are somehow related to each other, each with a specific role which can be any text. This structure is illustrated in Figure 4. Public transit lines are defined as relations containing roads, with nodes defining passenger drop off and pick up points.

Figure 5 presents a tram stop (node 314048973), and its platform in XML format. It is referenced from the platform, and several relations describe tram lines. Three other nodes of the platform are shared by footways connected to nearby sidewalks. Thus the pedestrian road network and public transit network are connected.

OSM data is available either in XML or the more compact PBF, Google’s Protocolbuffer Binary Format [20]. A map database of 3.3GB XML data can be expressed with PBF with only 159MB (the case of Finland), together with faster process times.
A Mixed Reality Interface for Real Time Tracked Public Transportation

Figure 5: Example data downloaded from the OpenStreetMap database.

Public Transportation: Tampere SIRI Interface

Tampere SIRI feed covers the local area public transportation with approximately 200 vehicles, most being tracked. Real-time bus locations from Tampere are available in the SIRI Lite format [21] encoded in JSON (JavaScript Object Notation). The interface provides a single file containing vehicle monitoring responses for all vehicles in circulation. The file changes once per second with an additional latency of one second before location observation updates from GPS receivers in the vehicles are centrally received and written into the file.

Mixed Reality Implementation

Our mixed reality platform is based on our previous research for mobile 3D city maps [22]. It has two main components: the Back End, handling data processing and transmission optimizations, and the Front End, currently an Android client, rendering the mixed reality visualization.

Back End

Our Back End hosts a database for 3D content, OSM and any other static data sets available to us. It also hosts real time data management processes, including interfaces to online data and real time state management. Entity state updates sent from Back End to Front End are utilized for both AR and 3D map visualizations.

Our networking solution is based on a dual view of the networked world. We consider on-line and off-line heterogeneous data sources as external to our system, integrated to our platform at the Back End. Our communications between the Back End and Front Ends are internal, homogenized with our efficient tokenized binary TCP/IP protocol, which we define using an extendable XML scheme [23].

Tampere Data processing

Figure 6 presents our Tampere real time public transportation data pipeline. OSM data is processed off-line. A data set covering Tampere region is downloaded, roads extracted and the data set compressed in our in-house open source tool OSM Squeezer. Public transit
A Mixed Reality Interface for Real Time Tracked Public Transportation

Figure 6: Tampere real time data management and transmission pipeline.

stops and routes are fetched in GTFS format, and extracted. The real time SIRI feed is received in JSON format, and processed using the Node.js JavaScript environment. Currently, we only extract vehicle numbers and positions. At this stage, we also map the raw GPS data to the nearest OSM street segments.

To manage Tampere buses as as individual entities, we created a proxy, a Server adapter, that parses the SIRI update file and yields coherent entity updates. Our Back End can apply an interest management scheme to optimize network updates. For this, we optionally provide a visibility based culling scheme, but by default transmit all updates from a pre-defined region within the city center of Tampere.

Front End

Our Front End is typically a mobile device, a smart phone or a tablet. Our current work is based on Android phones and tablets running Android 4.1.x or newer. Our main device is the ASUS T700T, running on Tegra 3 1.6GHz quad-core CPU with 12-core GPU and 1GB of RAM, connected to Back End via a Wi-Fi/3G bridge. Our implementation splits resources between Java and C++ via the Java Native Interface. All hard computation is performed at C++ side, including 3D rendering, while the user interface components are generated with Java.

Our Tampere 3D model is based on the original Tre-D model [24]. We built a textured bus model and are currently updating the Tre-D model with more recent textures and detail geometry such as bus stops. We are using an advanced Android version of the original m-LOMA engine [13] for rendering. Our augmented reality solution is based on embedded spatial sensors and the camera of the device.

Results

Our platform is a research prototype for testing advanced journey planning visualization possibilities. We have made preliminary evaluation of it on site by measuring tracking
A Mixed Reality Interface for Real Time Tracked Public Transportation

accuracy, with our own subjective observations and by conducting limited field experimentation.

Mixed Reality Interaction

As we have earlier shown, in proximal situation the 3D map interaction scheme can follow the same natural pointing paradigm as augmented reality: orienting the device orients the view, similar to the camera embedded in the device [15]. We support this sensor-based interaction together with touch screen interaction, allowing the user to raise up for an overall view or browse the entire environment by panning and zooming. To maintain coherence between AR and 3D map views, we provide a smooth transition.

On the ASUS, we maintain 30fps rendering speed in most situations for both augmented reality and the 3D map.

Tampere SIRI: Preliminary Data Accuracy

We estimated the positioning accuracy of local buses by projecting location observations from the SIRI feed to nearest road center lines along the actual bus routes (from OSM). The measured deviations are provided in Table 1. These estimates are slightly exaggerated due to simplified OSM data. 95% of vehicles are tracked within 10m from the actual positions, which is a fairly good result.

<table>
<thead>
<tr>
<th>Error</th>
<th>Portion</th>
<th>Error</th>
<th>Portion</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1m</td>
<td>21%</td>
<td>&lt;6m</td>
<td>88%</td>
</tr>
<tr>
<td>&lt;2m</td>
<td>39%</td>
<td>&lt;7m</td>
<td>91%</td>
</tr>
<tr>
<td>&lt;3m</td>
<td>57%</td>
<td>&lt;8m</td>
<td>93%</td>
</tr>
<tr>
<td>&lt;4m</td>
<td>72%</td>
<td>&lt;9m</td>
<td>94%</td>
</tr>
<tr>
<td>&lt;5m</td>
<td>82%</td>
<td>&lt;10m</td>
<td>95%</td>
</tr>
</tbody>
</table>

Table 1: GPS tracking error distribution.

We approximated the temporal accuracy using arrival at a bus stop as the ground truth. The temporal accuracy was observed to be 4±1s, of which 2.5±1s results from GPS reaction time (bus stopped/started to move) and 1.5±0.5s from networking and processing.

These results are based on local measurements, GPS time stamps and visual mapping on a single location where satellite visibility should be decent. We consider our results indicative, as accuracy in urban canyons tends to vary widely.

Locating a bus stop with mixed reality – simple field trial

Use case: John has just come out from a fast food restaurant, and wishes to continue his travel. Using the 3D interface, he spots a nearby bus stop, marked with a large arrow. Moving toward it along the sidewalk, a bus arrives and he switches to the AR mode. Luckily it is not yet his – no hurry! (see Fig. 7).

We presented a similar use case on the same area to seven subjects (ages 20 to 65, 3 male, 4 female), varying the target (one of the bus stops) randomly, using a) egocentric 3D map at ground level, b) augmented reality and c) exocentric 3D map with bird’s eye view. The task was to point out the real target based on the view. Our qualitative findings are:
Subjects considered the task to be viable; locating the correct bus stop from the area is a real challenge and the mixed reality interfaces could assist in this.

All interfaces appeared intuitive, with a very fast (10–20s) task introductory phase.

*Bird’s eye 3D view* provided fastest and most accurate responses.

*Ground level 3D view* provided fast and accurate responses, but occluded targets were more challenging.

*Augmented reality* provided slowest and least accurate responses; it suffered from sensor inaccuracy and occlusions; some targets were misidentified.

Local subjects found targets faster and more accurately than non-locals.

Local subjects found targets even with large inaccuracies in presentation (AR view).

Non-locals faltered in spotting more distant targets (all views).

---

**Real time transportation visualization – subjective validation and discussion**

Figure 8 provides two views to our real time visualization. For a simple “x-ray” visualization, we render occluded buses in red. We extrapolate bus locations, but cannot cater for changes in speed. In Fig. 8 (right), buses N\textsuperscript{o} 15 (front) and N\textsuperscript{o} 22 (on the right) are rather accurately positioned but bus N\textsuperscript{o} 12 (behind N\textsuperscript{o} 15) is still waiting behind the corner in the 3D view, while the real bus is already turning to the street. Our observations are:

- Observing the moving 3D models or red “x-ray” presentations yields a reassuring feeling on the traffic situation.
- The simple red “x-ray” visualization allows us to perceive occluded vehicles quite intuitively.
- Associating bus markers with buses with inaccurate AR is difficult.
- Not all buses are tracked via SIRI and the sometimes occurring mismatch is confusing; identifying a certain bus may become uncertain.
- Tracking accuracy is location and situation dependent, and varies.
- Resolving bus order on a street with a ground level view is challenging; with a bird’s eye view or laterally with the “x-ray” visualization it is easy.
Again, we have observed the challenges of embedded sensor accuracy. The 3D map appears to provide a better solution than AR. Sensor-based interaction does not suffer so much from the wobbling of the orientation sensor data, as target annotations stay locked with the 3D model. Depth perception does not suffer from local occlusions and motion parallax can be observed in better magnitude with touch screen interaction.

Figure 8: 3D Maps visualizing real time tracked public transportation in Tampere.

Conclusions and Future

We have developed a mixed reality system capable of portraying real time tracked public transportation vehicles using both augmented reality and 3D map views. The platform is functional and performs well. Our use cases are simple, aiming at proximity target acquisition such as bus stop and bus vehicle identification. While our initial user feedback is encouraging, further studies and development are needed to verify the true usefulness of our technology.

With mixed reality visualization, latencies and inaccuracies of tracking become clearly noticeable. Vehicle GPS streams and our local AR view suffer from similar problems. Although our 3D model is partially outdated, the 3D map appears to be less prone to such problems.

In the future, we will integrate a full multimodal journey planner functionality to the platform, with crowd sourced real time data feedback loop and the ability to portray other real time updated sources such as pedestrians, bicycles, cars and potentially even traffic lights and data emerging from environmental sensors. We will improve our visualization methods in search of the best user experience. We are also working toward higher quality external orientation sensors and real time kinematic positioning (RTK), which will hopefully minimize most of the observed sensor based issues. We will also test the app with smaller devices.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement № 601139, project STREETLIFE. We acknowledge the collaboration of CGI Finland Oy, City of Tampere, and Ismo Rakkolainen and Teija Vainio for the permission to use the Tre-D model.
A Mixed Reality Interface for Real Time Tracked Public Transportation

References


A Mixed Reality Interface for Real Time Tracked Public Transportation


