URQUELL: Using wrist-based gestural interaction to discover POIs in urban environments

Andreas Komninos, Stylianos Sarakinos, Lydia Marini, John Garofalakis
Computer Engineering & Informatics Department
University of Patras
Rio, Greece
[akomninos, sarakinos, lmarini, garofala]@ceid.upatras.gr

***Abstract*—In this paper, we discuss the implementation of a prototype application (URQUELL) for the opportunistic discovery of POIs in a urban environment, and the multimodal, non-visual interaction with the system to retrieve details about these. The application works by pointing a smartphone or a wrist-worn sensor kit towards a physical venue location of interest. By retrieving information about the venue pointed at over various APIs, the application presents these to the users via high quality text-to-speech synthesis, negating the need to interact with the smartphone, which can be kept in a user's pocket at all times. Here, we discuss the implementation details of our prototype and outline future research directions.**

Keywords*—*multimodal interaction; wrist sensors; location based services; gestural interaction

I. Introduction

Location-based applications and services (LBS) are a frequently encountered instance of pervasive computing. Though originally only the subject of research experiments, LBS are now a mainstream commercial application type, used for various purposes (e.g. tourism, routing and navigation, taxi hailing etc.). All LBS applications involve a common interaction pattern. A digital map is shown on a mobile screen device, and the map is overlaid with icons, denoting the position of points of interest (POIs). To discover the information pertinent to a POI, the user typically taps on the icon, resulting in the display of additional information (e.g. via popups, or navigating to a "details" screen and away from the map).

This interaction pattern supports exploratory behaviour well, in the case where the user is interested in discovering the POIs that might be of interest in an area, and to make subjective decisions based on the provided information (e.g., choosing the best café from the several available at a location). However, the interaction pattern fails to optimally support opportunistic, in-situ discovery of information. Supposing that a user is walking down the street of a city they are visiting, if something piques their interest, they have to bring out their device, unlock it, open the LBS app, locate the POI on the map, tap, read information, and then put the device back in the pocket. In this case, the overhead, i.e. the amount of interactions with the mobile device required to bring up the necessary information, is significant and consumes a disproportionate part of the overall interaction sequence to complete the information seeking task.

In this paper, we describe the implementation of a prototype system aimed at reducing the interaction overhead while exploring POIs in a urban environment. Our prototype works on two principles – reducing visual interaction with the device and LBS application, through the use of audio, and using wearable technology that is already part of the user's personal device ecosystem, as input to the system. The resulting prototype uses a wrist-worn sensor, which could be part of a smartwatch, to allow the user to opportunistically point at real locations and retrieve information about these via audio (headphones).

II. Related Work

In the previous scenario, a major pain-point for interaction with LBS is distinctly visible: the user must match the map and overlaid POI icons, with the real world around them. This is a potentially complex and time-consuming task, depending on the current map zoom level (so the POI may not be visible), the amount of nearby POIs (which may occlude the real target), the user's ability to cognitively match the map symbology to their surroundings, to name but a few factors. To resolve this issue, researchers have long been interested in using the mobile, or external devices, act as a means for pointing at real world objects, and then having the application bring up information about them on the screen.

The earliest example of this technique that we have found is the XWand by Wilson and Pham [1], a device comprising of multiple sensors that allow the pointing at real world objects through gesture recognition. Use of such external devices are also investigated by Simon & Frolich [2, 3], Carswell et al. [4], Strachan & Murray-Smith [5], Lei & Coulton [6] and Jacob et al. [7]. These research efforts introduced the concept of synergetic operation of multiple devices on the user's personal device ecosystem to solve the information seeking task, however, one significant drawback is that they required a user's full attention, occupying both hands, since the user had to hold the external device with one hand, while looking at their mobile screen (held by the other hand).

As mobile devices began to become enriched with sensor hardware, other researchers investigated the use of the mobile itself, as a pointer to the real world, doing away with the external pointing device. Such work is reported by Baillie et al. [8], Pombinho et al. [9], Magnusson et al. [10], Poppinga et al. [11]. The research community's interest in the research topic seems however to wane after Meek et al. in 2013 [12]. In these research efforts, the user is still faced with the issue of having to divert their attention from the real world to the mobile device, effectively context-switching between the two, in order to get the required information.

One plausible explanation for this decline in interest is the onset of Augmented Reality for smartphones. Given significant advances in hardware (processing power, sensing accuracy and camera quality), AR has become a popular interaction pattern to explore a user's surroundings. Such work is reported by Yovcheva et al [13], Lee et al. [14], Vert & Vasiu [15], Langlotz et al. [16] and Kamilakis et al. [17], to name but a few of the very sizable body of researchers working in this area. A recent and thorough review of the area and its application can be found in Chatzopoulos et al. [18]. Using AR-browsers, i.e. applications that overlay the smartphone camera image with POI markers, which the user can tap to view more information, removes the problem of context-switching, but typically requires that the device is held with two hands. Another drawback is that the user typically has to be stationary to use the app, since their field of vision is essentially restricted to the contents of the mobile screen.

Quite a different scenario and application, but still relevant to the topic, is the work of Rümelin et al [19], which we will mention separately. In this paper, the researchers investigated how car drivers can point at objects they see through the windshield, and have information about these objects be presented to them through the car's in-dashboard information display. From this paper we extract two important takeaways. First, although most participants slowed the car down while pointing, they continued to drive, hence showing that the pointing gesture has a cognitive impact, but not one that would entirely disrupt a safety critical task. Second, the participants felt that this type of interaction was natural and part of what they might do intuitively, e.g. if conversing about a landmark they see with a passenger.

From the previous review of literature, we extract the following conclusions. Gesturally pointing a device as a way to input an information search request to a mobile system, is a natural way to match the real world with information systems, with low cognitive cost. However the main issue encountered in all above systems, is the provision of information visually, which ties the user's attention to the mobile (or ambient, in the car's case) device. Multimodal provision of information for navigation via haptic (e.g. Szymczac et al [20] or Pielot et al [21]), or non-speech audio cues (e.g. Heller et al. [22], Komninos et al. [23]) has been examined in the past – however this relates to providing low resolution (i.e. non content-rich) information, which is not enough for the purposes of discovering necessary details about POIs.

III. URQUELL – Opportunistic Discovery of POIs

To address the issues mentioned above, we present the design and development of our system, called URQUELL: Urban QUErying using Location and Line-of-sight. The motivation behind URQUELL is the coming of age of two important technologies in mobile devices. First, the advent of smartwatches and wearable fitness bands. For users, carrying a sensor-laden device other than their personal mobile was something unfamiliar at the time of previous research, however, nowadays such external devices are very common. Secondly, the availability of high quality text-to-speech (TTS) libraries in modern mobile device SDKs allows for the easy implementation of natural sounding speech-based interfaces, which can be used to convey much richer information compared to the simple audio cues in previous research.

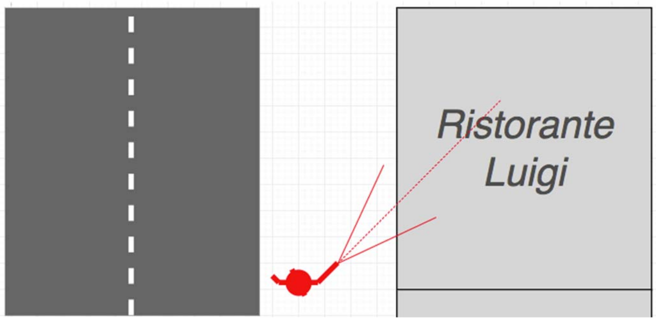

Fig. 1. Operating concept of URQUELL. As the user walks down a street (top-down view shown here), they are able to point at real world locations and retrieve information through speech about these places.

The operating concept of URQUELL is simple: The user wears a smartwatch that contains a gyroscope, accelerometer and magnetometer, which provides information about the direction of the user's arm to the mobile app. Upon detecting a pointing gesture (i.e. lifting up of the arm and keeping it pointed towards a location), the app attempts to discover if there are any POIs in the line-of-sight of the arm, by querying external APIs. The user's location is known via the device GPS. If a POI is discovered, its details are fetched, and conveyed to the user via TTS. Further gestures can control playback of the audio, for example, lowering the arm in a specific gesture, or pointing at another direction where another POI is discovered. With URQUELL, the user can keep their main device (smartphone) in their pocket, while all that's required is an opportunistic interaction via pointing, to retrieve the desired information.

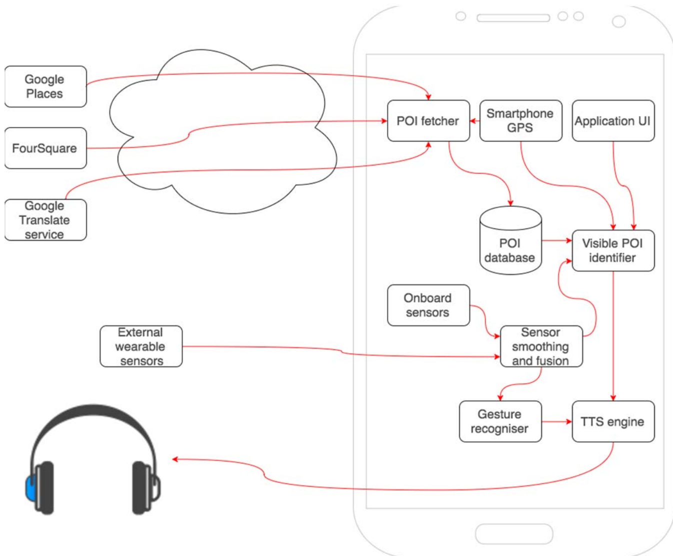

Fig. 2. URQUELL system architecture

A. Implementation

To develop URQUELL, we used the Android SDK (version 23). The overall system architecture is shown in Fig. 2. In order to keep the system independent from third-party SDKs offered by various smartwatch manufacturers, we selected to use only a wearable sensor set and for this purpose we used the Dialog Semiconductor DA14583 IOT development kit (Fig. 3).

This kit offers an integrated 6-axis inertial measurement unit including a gyroscope and accelerometer with sensor fusion and 3-axis geomagnetic field sensor. It also offers an environmental sensor suite (pressure, humidity and temperature), which are not relevant to our purpose. Communication with the smartphone takes place via a low-power Bluetooth system-on-chip module, using the Bluetooth Low Energy protocol. The sensor set is extremely wearable, measuring a mere 16 x 15 x 5mm. The unit benefits from very low power consumption.

Fig. 3. The Dialog DA1483 IOT sensor board, showing its tiny size.

To draw POI information, URQUELL interfaces with two location-based services APIs, namely Google Places and FourSquare, depending on user choice. Both APIs provide nearby venues using geographical coordinates and a distance radius as query parameters, and return information about venues, including category, rating and user comments. For the purposes of the system, and to conserve network and battery resources, we query the APIs with a user selectable radius (e.g. 200m). During retrieval of the results, we check for the language that the information is written in. Because information (usually tips and reviews) can often be in multiple languages, we send the information to Google's Translate API service, and retrieve back the resulting English text, which is then stored.

The retrieved results and relevant information are cached locally on the device in an SQLite database. As the user walks, when they reach a distance that exceeds a user-specified percentage (e.g. $3/4^{ths}$) of the specified query radius from the coordinates where the first query was issued, the system automatically issues another query to the API, without further user intervention. Previous POIs are cached, until a maximum of a user specified number of POIs (e.g. 500), after which limit earlier retrieved POIs are removed from the database.

When a user pointing gesture is recognized, the system attempts to determine whether a suitable POI exists in the database, based on the identified field of view of the sensors. To be more exact, we don't consider just a straight line, but rather a triangle extending from the user's position, at a user-specified angle and distance (Fig. 4). This allows the user to be less specific with the pointing gesture compared to a mere straight line, however, it introduces some problems when distant POIs are positioned closely together. To determine which POIs fall within the sensors' field of view, we select the POIs in our database which fall within a bounding box which contains the FOV triangle, and which is easy to calculate since we can calculate the FOV triangle coordinates (from the user specified angle and distance) and the user's position coordinates. For the POIs present within the bounding box, we then calculate the azimuth towards each POI and discard all those which fall outside the range of azimuths contained in the user specified angle. If multiple POIs are identified as falling within the sensor FOV triangle, we select the one which is closer to the user in terms of distance.

To ensure a consistent and smooth user experience, low-pass filters are applied to the sensor readings (accelerometer, gyroscope, magnetometers), and the sensor values are then passed to a sensor fusion function, from which we are able to determine azimuth, pitch and roll from the fused values.

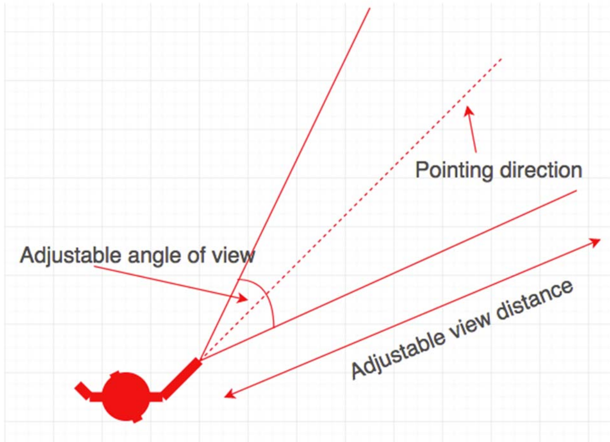

Fig. 4. Sensor FOV of URQUELL. The user can specify the angle and distance for the sensors' field of view. A lower distance works well in tight urban environments while for more open spaces, it can be adjusted accordingly.

In any case, once a POI is identified as currently being pointed to, then its details are retrieved from the database, and passed to the TTS engine, so they can be spoken out to the user, via their headphones. Playback of the information is stopped if a user points at another POI (and that POIs information is then played back), or if the user performs a special gesture by bringing their arm towards their torso area, and then dropping it down.

To allow us to check the operation of the system, and to specify operating parameters as described above, we are also displaying all of this information visually, using a map and relevant graphics (POIs and the sensor field of view), as shown in Fig. 5 and 6. Having this user interface up is not required for operation, as the whole system runs as a background service and the device can be in the screen locked state.

B. System versions

We have developed two distinct versions of the system. URQUELL-S (smartphone) works using only the device onboard sensors (Fig. 5), while the more advanced version URQUELL-W (wearable) works in synergy with the wrist-worn sensor (Fig. 6). The purpose of this is to allow us to investigate the use of a synergetic device approach, versus using the smartphone device itself as a pointing device.

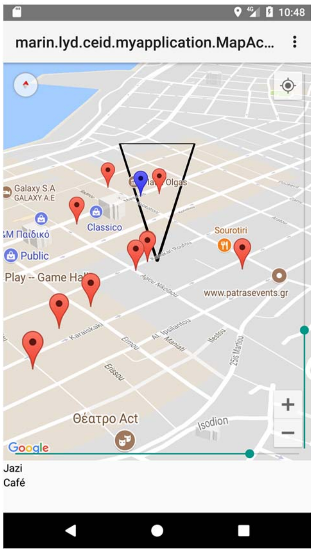

Fig. 5. URQUELL-S user interface, showing a 3D projection of the map. The map rotates automatically to point to the device's azimuth, keeping the triangle showing the device sensors field of view always fixed. In this case, two POIs fall within the FOV but the selected one (for which playback is ongoing) is marked blue since it's closer to the user. The venue details are shown below the map. Sliders on the map allow adjustment of the FOV angle (bottom) and distance (right).

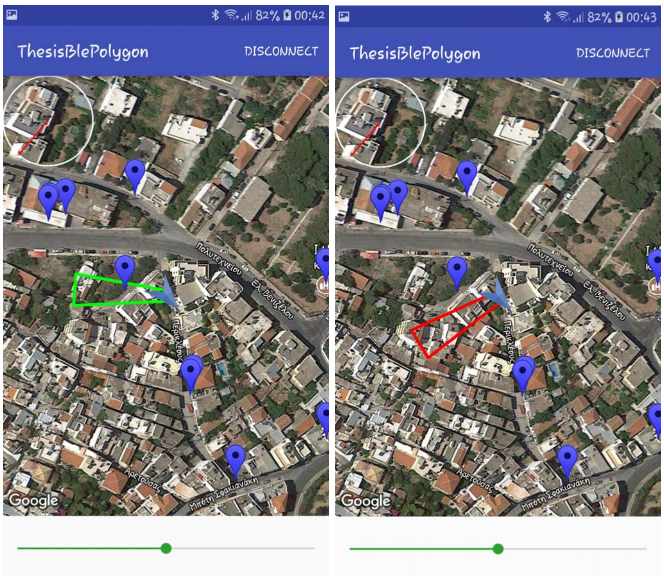

Fig. 6. URQUELL-W user interface. The arrow at the user's location shows the direction the user's mobile is pointing at, while the triangle extending from the user's location shows the direction the wearable sensor is pointing. On the left, a POI has been identified as "falling" into the sensor FOV, hence the triangle is shown with a green colour. On the right, no POIs are identified, so the triangle is shown in red.

IV. Technical Limitations

The URQUELL concept faces some technical implementation challenges, which merit further investigation prior to advancing with evaluations. These issues concern mostly the precision that can be achieved with the on-device or on-wrist sensors, and the impact on battery life. We outline these next.

One concern is the accuracy of the sensors, both in terms of user positioning and pointing direction. It is known that GPS signals can provide erratic information on user position in urban settings, and that additionally magnetometers might not always offer accurate estimates of orientation. The first problem can be likely mitigated by integrating a solution such as Google's Snap-to-Road API, which can overcome the problem of the user's position suddenly "jumping" in the middle of city blocks. Orientation issues are probably more an issue in URQUELL-S than URQUELL-W, since the latter platform uses on-board 9-DOF sensor fusion from accelerometer, magnetometer and gyro to provide orientation. A similar fusion technique is applied in some Android OS versions for orientation (known as virtual sensors), but of course the quality depends on the hardware and OS implementation, which may differ across devices.

As the user navigates their environment, it is plausible that the system might interpret their bodily position as input (pointing), generating many false-positives and ensuing in audio playback which can become annoying to the user. To mitigate this, we believe that proper gesture design can help. For example, using gyroscope data, we can accurately know where the device (phone or wrist) is pointing relative to the ground. The system can understand that user body posture should only be interpreted as input, if the sensors detect that the phone (or wrist) are in a position that is relatively parallel to the ground. This technique can be used to reduce the processing workload and battery consumption. When the device is not almost parallel to the ground, we can automatically drastically reduce the sensor & GPS data output frequency. The frequency can be raised again to the supported maximum, to allow for high-resolution gesture data, when the device is raised to point at a venue. Additionally, while the device is pointing towards the ground, processing to identify venue targets from the internal database can also stop.

V. Directions for Research

In the previous sections, we described the design and implementation of URQUELL, a system for the opportunistic discovery and multimodal presentation of information about points of interest in a city. The system is designed to overcome the limitations of previous research efforts, being designed in such a way as to minimize the user's interaction with their mobile device, while seeking information on the move. To achieve this, URQUELL leverages from the onboard and wrist-worn sensors from the user's personal device ecosystem, and presents information via synthesized speech.

To investigate the user interaction aspects with URQUELL, we are currently considering several research directions. Firstly, we would like to compare the two variants, URQUELL-S and –W, against a traditional map-based POI discovery application, in controlled field experiments. By logging the users' GPS and in-system behaviour (selecting POIs, controlling the app behaviour, retrieving information), we want to investigate not only the efficiency of URQUELL in terms of ad-hoc information seeking behaviour, but also behavioral aspects that concern the user's mobility. For example, we are interested in seeing how users' mobility patterns in terms of walking speed, stops and direction changes are affected by the use of the system.

Subsequently, we are interested in investigating the problem of selecting optimal distance and angle sizes under context. Obviously, depending on the characteristics of the urban environment, larger or smaller angles and distances for the sensors FOV might be more appropriate (e.g. compare walking down a narrow souk, vs. exploring in a large park or wide pedestrian shopping street). As shown in Fig. 7., the dynamic adaptation of these parameters can be useful even in cases when the user is facing across a street. For this purpose, we are exploring whether it's possible to dynamically determine the most appropriate settings, based on map feature dimension data, such as is available e.g. via OpenStreetMaps, or possibly by processing satellite imagery at the user's current location. For the latter option, a training set of optimal angles and distances for POIs at various ranges from a user's location can be built based on real user choices or simulated scenarios, to feed a machine-learning model such as a neural network.

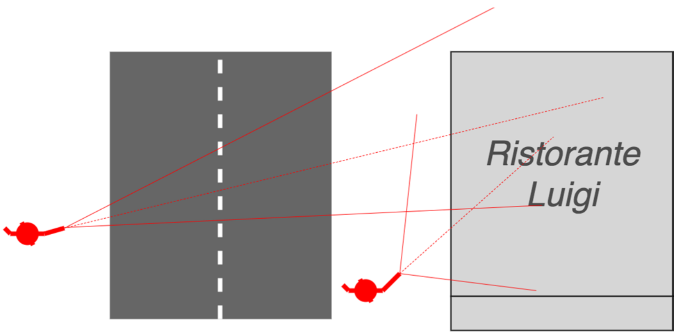

Fig. 7. Dynamically adapting the range and angle of the sensor FOV

Controlling the system via gestures (e.g. pointing, moving the arm to control speech playback) is another topic of research. Further from the problem of accurate gesture recognition, these gestures must be designed in such a way that they would feel unobtrusive and discreet, so they can be performed in public without the user feeling uncomfortable. Pointing at something while walking with another person (traveling companion) is certainly common, and an external observer would perceive it as a communication medium between the two travelers. However, pointing at something while alone is possibly a more socially awkward situation (for example, consider how people talking via Bluetooth headsets appear strange to others [24]).

A further option that we are exploring with this regard is the integration of the wearable sensor at locations other than the user's wrist. For example, given the sensor kit small size and weight, it could be plausibly integrated into the user's glasses, or headwear (e.g. cap, hat, beanie). Therefore, a more natural interaction might be achieved, as the system would take the user's head direction as input.

VI. Conclusion

The idea of using external sensors as input to a mobile application for the opportunistic discovery of POIs has been explored in the past, laying the groundwork in this interesting area of natural interaction with ubiquitous location-based services. Although the idea has been abandoned by the research community in the recent years, we believe that the proliferation of smartwatches and multi-device personal body-area ecosystems make the concept worth visiting again. Given the processing power available in modern mobile devices, the cheap wireless connectivity and abundance of spatial data for our urban environments, there is a strong argument that services should increasingly be decoupled from the smartphone's screen and delivered opportunistically and contextually with more support for multimodal interaction. URQUELL is a step in this direction and through the technical implementation presented here, we are able to raise several interesting research directions which move beyond the limitations of previous research.

References

[1] A. Wilson and H. Pham, "Pointing in intelligent environments with the World Cursor," in *In the proceedings of Interact 2003*, 2003.

[2] R. Simon, P. Fröhlich, G. Obernberger, and E. Wittowetz, *The Point to Discover GeoWand*. .

[3] R. Simon, P. Fröhlich, and T. Grechenig, "GeoPointing: evaluating the performance of orientation-aware location-based interaction under real-world conditions," *Journal of Location Based Services*, vol. 2, no. 1, pp. 24–40, Mar. 2008.

[4] J. D. Carswell, K. Gardiner, and J. Yin, "Mobile Visibility Querying for LBS," *Transactions in GIS*, vol. 14, no. 6, pp. 791–809, Dec. 2010.

[5] S. Strachan and R. Murray-Smith, "Bearing-based selection in mobile spatial interaction," *Pers Ubiquit Comput*, vol. 13, no. 4, pp. 265–280, May 2009.

[6] Z. Lei and P. Coulton, "A Mobile Geo-wand Enabling Gesture Based POI Search an User Generated Directional POI Photography," in *Proceedings of the International Conference on Advances in Computer Entertainment Technology*, New York, NY, USA, 2009, pp. 392–395.

[7] R. Jacob, P. Mooney, and A. Winstanley, "What's up that Street? Exploring Streets Using a Haptic GeoWand," in *Advances in Location-Based Services*, Springer, Berlin, Heidelberg, 2012, pp. 91–105.

[8] L. Baillie, H. Kunczier, and H. Anegg, "Rolling, Rotating and Imagining in a Virtual Mobile World," in *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*, New York, NY, USA, 2005, pp. 283–286.

[9] P. Pombinho, M. B. Carmo, A. P. Afonso, and H. Aguiar, "Location and Orientation Based Point of Interest Search Interface," in *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, New York, NY, USA, 2010, pp. 375–376.

[10] C. Magnusson, K. Rassmus-Gröhn, and D. Szymczak, "The Influence of Angle Size in Navigation Applications Using Pointing Gestures," in *Haptic and Audio Interaction Design*, 2010, pp. 107–116.

[11] B. Poppinga, M. Pielot, W. Heuten, and S. Boll, "Towards an Objective Comparison of Scanning-Based Interaction Techniques," in *Haptic and Audio Interaction Design*, 2012, pp. 111–120.

[12] S. Meek, G. Priestnall, M. Sharples, and J. Goulding, "Mobile capture of remote points of interest using line of sight modelling," *Computers & Geosciences*, vol. 52, no. Supplement C, pp. 334–344, Mar. 2013.

[13] Z. Yovcheva, D. Buhalis, and C. Gatzidis, "Overview of smartphone augmented reality applications for tourism.," *e-Review of Tourism Research*, vol. 10, no. 2, pp. 63–66, 2012.

[14] G. A. Lee, A. Dünser, S. Kim, and M. Billinghurst, "CityViewAR: A mobile outdoor AR application for city visualization," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, 2012, pp. 57–64.

[15] S. Vert and R. Vasiu, "Relevant Aspects for the Integration of Linked Data in Mobile Augmented Reality Applications for Tourism," in *Information and Software Technologies*, 2014, pp. 334–345.

[16] T. Langlotz, J. Grubert, and R. Grasset, "Augmented Reality Browsers: Essential Products or Only Gadgets?," *Commun. ACM*, vol. 56, no. 11, pp. 34–36, Nov. 2013.

[17] M. Kamilakis, D. Gavalas, and C. Zaroliagis, "Mobile User Experience in Augmented Reality vs. Maps Interfaces: A Case Study in Public Transportation," in *Augmented Reality, Virtual Reality, and Computer Graphics*, 2016, pp. 388–396.

[18] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui, "Mobile Augmented Reality Survey: From Where We Are to Where We Go," *IEEE Access*, vol. 5, pp. 6917–6950, 2017.

[19] S. Rümelin, C. Marouane, and A. Butz, "Free-hand Pointing for Identification and Interaction with Distant Objects," in *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, New York, NY, USA, 2013, pp. 40–47.

[20] D. Szymczak, C. Magnusson, and K. Rassmus-Gröhn, "Guiding Tourists through Haptic Interaction: Vibration Feedback in the Lund Time Machine," in *Haptics: Perception, Devices, Mobility, and Communication*, 2012, pp. 157–162.

[21] M. Pielot, W. Heuten, S. Zerhusen, and S. Boll, "Dude, Where's My Car?: In-situ Evaluation of a Tactile Car Finder," in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, New York, NY, USA, 2012, pp. 166–169.

[22] F. Heller, J. Jevanesan, P. Dietrich, and J. Borchers, "Where Are We?: Evaluating the Current Rendering Fidelity of Mobile Audio Augmented Reality Systems," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY, USA, 2016, pp. 278–282.

[23] A. Komninos, P. Barrie, V. Stefanis, and A. Plessas, "Urban Exploration Using Audio Scents," in *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, New York, NY, USA, 2012, pp. 349–358.

[24] J. L. Dvorak, "Social Issues of Wearables," in *Moving Wearables into the Mainstream*, Springer, Boston, MA, 2008, pp. 311–332.