Designing for Mixed Reality Urban Exploration

Salvatore Andolina[1,2], Yi-Ta Hsieh[2], Denis Kalkofen[3], Antti Nurminen[4], Diogo Cabral[5], Anna Spagnolli[6], Luciano Gamberini[6], Ann Morrison[7], Dieter Schmalstieg[3], Giulio Jacucci[2]

[1] University of Palermo, Italy
[2] University of Helsinki, Finland
[3] Graz University of Technology, Austria
[4] Aalto University, Finland
[5] ITI/LARSyS, University of Lisbon, Portugal
[6] University of Padova, Italy
[7] Auckland University of Technology, New Zealand

Abstract. This paper introduces a design framework for mixed reality urban exploration (MRUE), based on a concrete implementation in a historical city. The framework integrates different modalities, such as virtual reality (VR), augmented reality (AR), and haptics-audio interfaces, as well as advanced features such as personalized recommendations, social exploration, and itinerary management. It permits to address a number of concerns regarding information overload, safety, and quality of the experience, which are not sufficiently tackled in traditional non-integrated approaches. This study presents an integrated mobile platform built on top of this framework and reflects on the lessons learned.

Keywords: Augmented reality; mixed reality; virtual reality; haptic guidance; urban exploration; cultural heritage.

1 Introduction

Cities are amongst the most significant of all tourist destinations [1]. They attract people with a variety of primary purposes, ranging from leisure to business, conferences, shopping, and visiting friends and relatives [2]. In terms of tourist experience, urban environments pose peculiar challenges and opportunities. When visiting a city, people spend considerable time planning their activities, both before and during the visit itself. However, they tend to deliberately make plans that are not highly structured and specific, so that they can take advantage of changing circumstances [3]. Tamminen et al. [4] show how a typical behavior of urban visitors involves sidestepping from a predefined plan to engage in opportunistic exploration. Moreover, urban environments offer copious points of interest (POI), comprising sites, services, or cultural artifacts which are distributed in space and can also be encountered as dense assemblies in particular areas, thus providing great opportunities for exploration and personalization [5]. In addition, users require different modalities and interfaces, depending on the task at hand that can vary from planning, searching, navigating, or inspecting [6].

We define "urban explorer" as a city visitor whose behavior is more complex than the behavior of visitors of indoor settings, such as museums, where exploration options are more limited or predefined. The needs of urban explorers are dynamic and evolving, and require a wide set of tools to be met, as well as easy transitions between tools [6]. Today's multimedia technologies provide many useful tools for specific problems such as wayfinding, recommendation, or augmentations. Commercial solutions such as Google Maps, Google StreetView, and Google Earth permit the exploration of sites of cultural, historical, and geographic significance. Smart glasses and head mounted displays, such as Microsoft Hololens, allow the augmentation of cities with immersive data visualizations [7]. A wide range of touchscreen-based mobile virtual guides provide visitors with context-related facts and recommendations [8, 9]. However, the tools must be handled competently, since they are not designed to support serendipitous discovery during limited time. Furthermore, cognitive load can become a concern. When accessing through a smartphone, users may miss important aspects of the real-time experience or even risk accidents.

The aim of this work is to propose a comprehensive design framework for augmenting an urban explorer, who receives natural support from a multimodal display, consisting of visual, auditory, and haptic components. This form of mixed reality combines real and computer-generated digital information in the user's sensory perception, providing a bridge between physical and virtual reality with the intent to reduce cognitive load. Although mixed reality support in the tourism context has already been explored [10–12], a comprehensive consideration of MRUE has largely remained elusive in terms of what it means and what it should entail with regards to functionalities. This work aims to fill this gap.

1.1 Methodology

We contribute a design framework for MRUE based on a concrete implementation in an historical city (Figure 1). This research started from the consideration of typical urban exploration problems, to define desired design principles for a MRUE scenario. Then, it engaged in several research through design activities [13], to produce research prototypes that were tested with potential users in a historical city. The prototypes were used as probes to investigate how we could concretize the envisioned design principles into technological features that would successfully address the typical problems faced in MRUE [14–21]. The investigation led to the integration of different modalities, such as VR, AR, and haptics-audio interfaces, as well as advanced features such as personalized recommendations, social exploration, and itinerary management into the same mobile application running on a tablet. Finally, this study concludes by studying the integration and transitions of different modalities in an urban exploration scenario.

2 Related Work

Mixed reality has been defined as the "merging of real and virtual worlds along the virtuality continuum which connects completely real words to completely virtual ones" [22]. Moving on from reality, the mixed reality of AR takes what we can see around us

and augments it with virtual information that can be accessed both visually or with non-visual modalities, such as audio and haptic [23]. Toward the other end of the continuum, there is VR, where interaction happens in a virtual world. In the "Post-WIMP" world [24], reality, or the world, has become the interface of the future [25, 26]. Mixed reality has the potential of extending our awareness in a very natural manner. This particularly applies to an urban environment with dense content such as shops, homes, historical sites, cultural points of interest, and human experiences.

Several mixed reality approaches in urban environments have been proposed for spatial augmentation, sensing, and wayfinding. The first trials with fiducial-marker-based AR apps with PDAs required learning regarding the tracking itself [27]. Later, work toward marker-less tracking mitigated these issues [28]. Similarly, rendering issues related to large urban 3D virtual environments in mobile devices have been handled with optimizations [29]. These developments have been followed by focused studies, such as navigation experiments [30, 31]. Fröhlich et al. [32] discuss the usage of handheld devices, AR, and 3D maps for geo-reference data retrieval, while Schmalstieg and Höllerer [23] present different AR systems for spatial wayfinding. Recently introduced, the Walkable MxR Map [11] uses interactive, immersive, and walkable maps to allow users to interact with cultural content, 3D models, and different multimedia content at museums and heritage sites. It allows users to interact with virtual objects via maps that are virtually projected on the floor and viewable through mixed reality devices, like head-mounted displays. However, these approaches did not target the particular case of urban exploration, where people divert from a predefined plan and engage in serendipitous discovery.

Typical solutions for urban exploration leverage recommendations derived from like-minded users. However, in a recent field study on touristic exploratory navigations, Vaittinen and McGookin [6] found that most of current tourism systems are based on recommendations visualized on maps and lists (e.g., TripAdvisor), which do not work for all touristic/urban exploration needs. Other approaches to support mobile urban exploration have compared 2D and 3D maps, list and category views, as well as tag clouds [33] but have missed the study of AR and haptics-audio interfaces. Recent research like the Uncover [34], which compares 2D maps with radar visualizations of POIs for urban exploration, is still limited to regular mobile applications. McGookin et al. [35] have also developed a mobile app with traditional 2D maps combined with haptic and sound notifications intended to reveal cultural heritage to visitors of an island. Its main goal was to respect the varied reasons individuals visit the island and support free exploration. More complicated approaches for urban exploration have combined AR and 3D models [36] in the form of Wizard-of-Oz prototypes, showing that AR is more pleasant and realistic, as well as easier to associate with the environment.

Although existing studies and systems highlight the advantages of various features, such as AR, multi-modal interaction, and recommendations for urban exploration, there is no agreement in what it means for designing MRUE. In this work, we start from identified problems and opportunities to propose a framework for MRUE, based on the experience gained in designing a full working prototype, including multimodal MR support for urban exploration in a historical city with a large amount of cultural content.

3 Design Principles

Several factors need to be considered when designing for MRUE. Based on our literature review, we identified four main design principles (Figure 1), as described below.

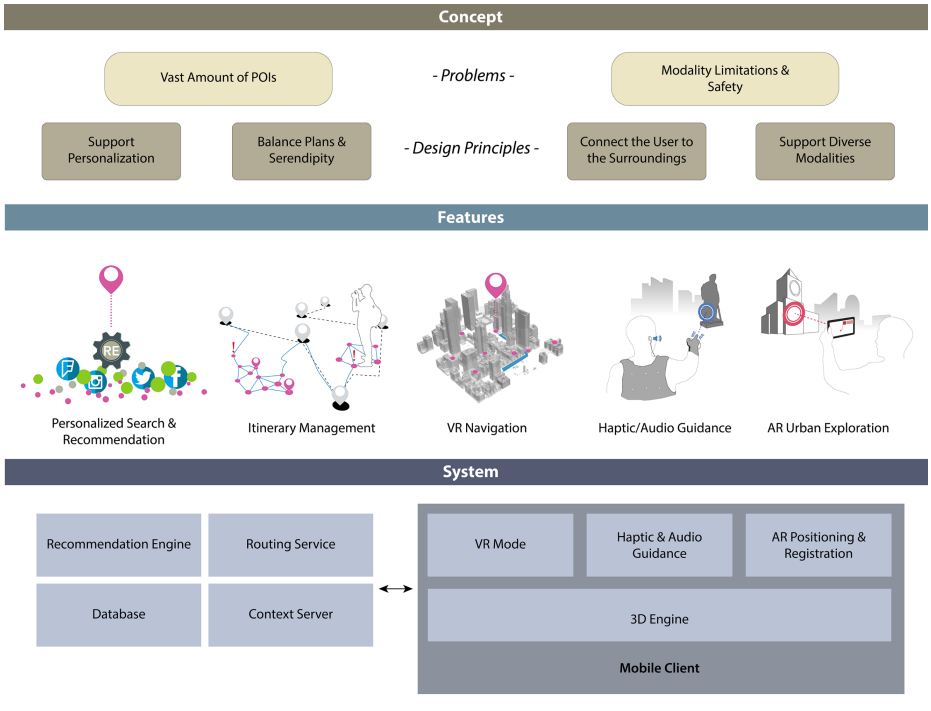

Fig. 1. The Mixed Reality Urban Explorer framework.

3.1 Support Personalization and Social Discovery

A rich urban environment can contain a large number of POIs. Yet, identifying relevant information is challenging with common search tools, resulting in high mental load and impoverished experience. For example, the same top-rated POIs would be pushed to all nearby users. It is crucial to combine search engines with personalized recommendations based on a user's profile, social ratings, and contextual information. A technology for urban exploration needs to retrieve information that matches user interests and external factors (e.g., opening hours, weather forecast, etc.), with the POIs available in a particular place. This requires a model of users and their contexts, selecting the most appropriate content and delivering it in the most suitable way [5]. We also see a need to acknowledge variability in a visitor's profile. Users should be

allowed to transfer from one profile to another, rather than being confined to a static (or not easily changeable) profile.

3.2 Balance Plans and Serendipity

Urban exploration often involves stepping aside from a predefined plan to engage in opportunistic exploration [4]. Most current mobile solutions provide little support for exploration and serendipitous discovery; they focus more on providing a route from the visitor position to a particular POI (e.g., a museum, restaurant, or church). Such approaches usually rely on geographical visualization of the environment (such as a 2D map), complemented with a route and POI list. Such information is suitable for wayfinding but not so much for the broader issues of exploration, like getting an overview or choosing a worthwhile destination. Facilitating serendipity becomes a significant issue in designing for exploration rather than wayfinding. Users should be allowed to have significant experiences by chance. One possible way to facilitate the process is to provide users with an extended awareness of the surroundings; this can be achieved by proposing mixed reality cues which combine information from the vast digital information with the physical environment, inviting the user to explore through personalized suggestions.

3.3 Connect the User to the Surroundings

While exploring a place, a user's focus of attention should be on the surroundings where the unique content is located. However, very often, virtual information (e.g., 2D maps or POI lists) draws much of a user's attention to a device screen and spares few cognitive resources for exploring the surroundings, which may result in not only hampered experience, but also safety concerns. The commonly adopted touchscreen-based interaction (eyes monitoring where the fingers touch) further concentrates a user's cognitive resources onto the handheld device. This almost forces users to immerse themselves into the virtual world and disconnects them from the surroundings. Technologies for urban exploration need to redirect the visitor's attention from the device to the surroundings, taking into account the allocation of cognitive resources on the interface. This is addressed through mixed reality affordances that provide access to information by interacting in the physical environment.

3.4 Support Different Modalities along the Virtual Reality Continuum

Leveraging modalities, other than vision, is a straightforward approach to avoid users from being disconnected with surroundings. By providing the necessary redundancy, multimodality can transfer cognitive resources from the device toward the environment, an aspect especially important in a mobile setting [37]. For example, the user's visual attention can be focused on the environment while receiving audio cues from the system. To achieve this, a desirable feature to consider is to merge virtual objects with the real world; this will help to avoid switching attention between the device (in which the virtual objects reside) and the environment. In particular, supporting various degrees of virtuality and immersion along the VR continuum could provide MRUE users with the flexibility needed to interact with the dynamic urban environment.

Depending on the situation, MRUE users should be allowed to choose the modality providing the most suitable level of safety, obtrusiveness, social acceptance, level of detail, and so on. For example, haptic and audio interfaces may be more suitable for users standing by a road, where visual attention shall be spent on monitoring the traffic. Multimodal interfaces need to generalize across multiple display dimensions (visual, audio, and haptics) and decide where and how information items should be displayed for the best results.

4 Framework Features

Our framework for MRUE (Figure 1) is based on the identified design principles and on a concrete implementation in the historical city of Padua, Italy. Below, we describe the implemented features of our framework and discuss their integration.

4.1 Recommendations and Flexible Itineraries

We addressed the problem of the vast amount of POIs by providing a level of personalization through the implementation of a recommendation engine; this engine ranks a POI, based on social media, visitor profiles, and query inputs.

Social rating, as used in TripAdvisor or Booking.com, is a Web 2.0 technology, which analyses the experience of other visitors to single out highly rated POIs. However, the reasons for certain destinations becoming highly popular are obscure and may be owed to reasons that do not concern the inquiring visitor, especially if the visitor deviates from the evaluation group.

Therefore, recommendations in our system also always consider the user's profile. Weighting by both crowd-sourced ratings and by personal preferences better facilitates serendipity, since recommendations can be based on either. We found it necessary to explicitly deal with profile variability. Consequently, we initiated by analyzing tourists' profiles by reviewing tourist office statistics, interviewing more than 250 tourists in Padua, Italy, and talking to tourism stakeholders. Our goal was to identify factors that had a significant effect on tourists' visiting preferences and habits. We initially selected age, provenance, length of stay, companionship, interests, and budget. Soon, we realized that visitors' profiles could not be assumed to be tantamount, since a visitor may change his or her profile several times during a visit when budgets change, companionship shifts, or interests drift. Therefore, we made the main factors easily re-configurable from the interface, so that recommendations can reflect the user's current situation [20]. Technically, our approach is based on a graph-based retrieval model that utilizes different data sources as graph layers associated with content: the social ratings, tags, and personal user profiles. This is where the relevance estimation model performs random walks with restarts on the graph overlays and computes a ranking for the information items. We evaluated our approach through an experiment with the largest publicly available social content and review data from Yelp [19], and found that it improves result rankings, visualization of results to communicate relevance, and the amount of like-minded user activity around the results.

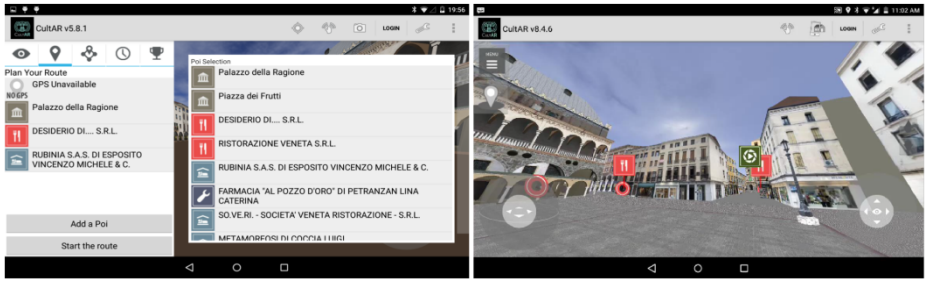

Fig. 2. (a) Itinerary building. Recommended POIs can be selected to create a custom path. (b) Virtual Exploration.

A subset of POI locations retrieved by the recommendation engine can be selected and used to create a customized itinerary (Fig. 2a). The itinerary is built with the help of an external routing service (Nokia). However, regular wayfinding is replaced by a joint space-time optimization concerning the recommended POI locations: *Which set of POI locations could be visited this morning?* and *Which area of the city should be visited today?* As an alternative, our system provides itineraries curated by professional tour guides, which highlight a topic of the urban environment (e.g., religious places, art museums, famous nightlife places, or restaurants) or tell a story (e.g., life of Galilei). None of these itineraries force a visitor to follow the suggested path. Instead, the itinerary is merely an initial suggestion, from which the visitor can divert and return to afterwards. Visual or haptic cues regarding the location of the next POI to be visited can be provided as requested by the visitor.

4.2 Wearable Haptic and Audio Guidance

Fig. 3. Mixed reality cues and affordances can be provided through haptic user interfaces: (a) the vest; (b) the glove; and (c) audio-haptic exploration, in which an MRUE user points at a POI with the haptic glove and listens to the associated audio description.

We deployed aural and tactile channels as a complement to the visual channel into the system. Eyes-free interaction techniques avoid drawing users' attention onto the interface, enabling higher awareness on the surroundings and less concern of safety. For example, we developed a vibrotactile vest (Fig. 3a) that offers a hands- and eyes-

free navigation, leveraging both locations on the body and produces distinctive vibration patterns to indicate cues such as direction, degree of turn, speed, or user error, which literally enables embodied interaction with the environment. The wearers are steered towards their chosen POI locations, prompted in a "natural-enough", pleasant, and easily understood manner [18]. To complement the information provided by the vest, we leveraged a headset capable of 3D audio.

This not only provides essential textual information, but also spatial information about the environment, such as distance, increasing or decreasing proximity, and direction of a POI. The categories of POIs (e.g., cultural, shopping, or service) can also be presented through auditory icons.

Moreover, we developed a sensor- and actuator-equipped glove (Fig. 3b); it leverages hand gesture recognition for direct interaction with the real environment. To fetch the information associated with a specific POI, the user can directly point at the landmark with the hand and make a selection gesture (Fig. 3c). Although a POI is distant and untouchable, its tangibility is represented through the vibrotactile cues on the hand. The non-visual multimodal interaction is complemented visually by the real view as seen by the user. When the glove is set in exploration mode, MRUE users get a vibration cue anytime a recommended POI is nearby. Users then accept or reject the information by bending the index finger or the thumb, respectively. Vibrotactile guidance cues on the hand then lead users to point in the direction of the recommended POI. We conducted a comparative study with a context-aware mobile app and found that although experiencing similar performance of different evaluation metrics, smartphone users spent on average 70% of their time looking at the screen when exploring the urban area, whereas users wearing our haptic glove were able to have a good exploration experience while leaving their visual attention on the surroundings [16].

4.3 Augmented Reality of Urban Artifacts

Mobile AR browsers have recently become very popular [38]. They commonly augment an urban POI, using GPS in combination with the orientation sensor of modern smartphones. Although this approach works well for distant POI locations surrounding the user, the applications often suffer from an imprecise user localization for artifacts in the close proximity of the user.

Therefore, we have built an image-based localization pipeline; it can precisely estimate the 6-degrees-of-freedom camera pose (i.e., position and orientation) of a user's AR device. Our implementation includes a localization system based on an analysis of visual features and an incremental tracker. Similar to Arth et al. [39], our system localizes a query image (the current video frame, acquired by the camera) within a known 3D world, represented as a 3D-point cloud. To more precisely estimate a user's pose after its initial localization, we have developed a fast incremental tracker, based on rotational movement only (i.e., panning the device around while staying in place). Combining the absolute 6-degrees-of-freedom camera pose (which we retrieved from the localizer) and relative 3-degrees-of-freedom, camera rotation yields a full 6-degrees-of-freedom pose for every camera frame. Here, efficiency is crucial to be able to process the input images at frame rate. Furthermore, the ability to cope with blurry

frames or frames that contain only few trackable feature points is also very important for the overall robustness of an AR application. Our system addresses these issues by adaptively combining device sensors and pose information retrieved from the panoramic tracking system. In particular, our approach supports running and coordinating multiple instances of the panoramic tracker, each having its own panoramic map, and its combination with all other sensors able to retrieve the orientation on the current mobile device.

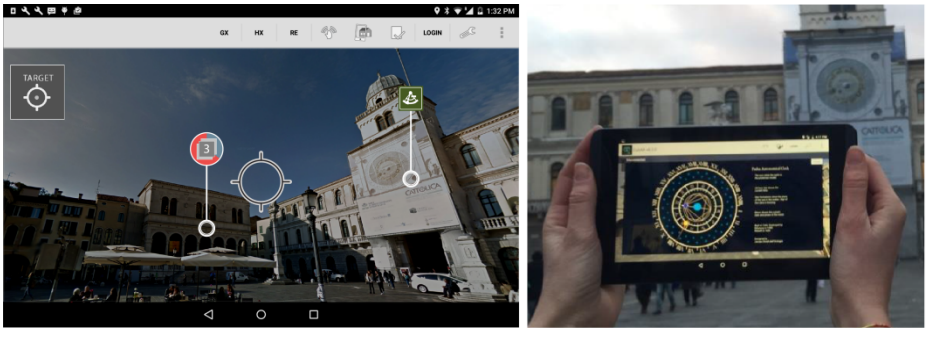

Fig. 4. Mixed reality cues and affordances through computer-vision-based AR provide visual augmentations of a POI. (a) The AR visualization points the user to relevant objects, which he or she can select for detailed information. A green icon indicates that animated content is associated with the astronomical clock. Next to the center, an abstract visualization indicates a cluster of search results, including the overall amount and the type encoded in color. The reticle in the center of the screen can be used to aim at POI icons by moving the tablet. When the reticle matches with a POI, users can tap with the thumb on the target button on the edge of the screen to confirm selection. (b) The animated explanation of the astronomical clock.

While we enable precise spatial registration of annotations to POI in the environment, we cannot simply place the annotation exactly over the POI. Finding a good placement of object annotations in a visual display is difficult, since a clutter must be avoided. A common approach is to choose locations such that a minimum number of salient objects are occluded. Such view management is commonly done by computing force fields [40] that repel annotations from important objects. Unfortunately, this force field must be computed for every frame, such that the chosen locations do not suddenly change and the annotations appear to jump erroneously. Our coherent view management [17] evaluates the force field in 3D space, even though a 2D image layout is finally computed. This move to a higher dimension makes sure a temporally continuous motion emerges, where annotations naturally follow the objects they are assigned to. We also reliably suppress clutter by restricting the maximum annotation density on the display. If the POI list from the recommendation engine is too long and does not comfortably fit on the screen, conventional view management works by just omitting low-ranking annotations, making serendipitous discovery impossible. Adaptive view management [21] folds similar annotations of low importance into a single “group” POI, which initially takes less space but can be unfolded by the user on demand. This feature supports the creation of mixed reality

cues and affordances through visual augmentations of POI (Figure 4). We performed a user experiment, comparing our approach to a conventional AR browser, using a search task [21]. The results show that participants made fewer errors when using our interface; this indicates that the reduced clutter allowed them to focus on finding relevant items. Therefore, we recommend using adaptive view management AR in for applications which require exploring a large amount of data in-place.

4.4 3D Engine and Virtual Reality

As an alternative to AR, we provide a VR view of the environment. The AR view has the fundamental property of being egocentric. While this provides intuitiveness due to the short cognitive distance between the real view and the view on the screen (they are always aligned), it also poses limitations. Augmented content is overlaid on top of video feed with no capability to resolve real world occlusions, such as people walking by. Labels that are supposed to be on a building facade will hang in front of everything, contradicting human depth perception, where occlusions provide the strongest depth cue. Similarly, content that belongs to the far side of a building would be also visible. Indeed, incorrect depth interpretation is the most common perceptual problem in AR applications [41]. In our system, the second case is handled by an online visibility test hosted in the client device, utilizing our 3D city model as a virtual occlude. For the first case, we offer an alternative.

Table 1. Implementation details of system used to inform the development of the MRUE framework

City of Padova model	More than 3,000 *points of interests* (1,000 of which high quality), 8 animated explanations (Fig. 4b), 5 hectare of *area covered*; More than 100k vertices and 184 textures in the resulting *3D model*. *Point cloud* created through Ladybug 3 360 camera with six 1600x1200 Sony CCD sensors
Rec. engine	Performance: 183 ms per query. Graph with 3,276 nodes and 16,995 edges;
AR and VR	Performance with Tegra 4-based Asus TF701T (max refresh rate 30 fps): 40fps in ***sensor-based* AR** tracking; 16.6 fps in ***computer vision panoramic* AR** tracking; 100-200 fps in **VR**
Haptic Vest	Includes 40 precision microdrives pancake actuators, custom Arduino Mega board with two RN42 Bluetooth control modules, InvenSense MPU9150 Inertial Measurement Unit, modular power board using 2-3 PCA9685 PWM I2C controllers
Haptic Glove	Constructed used thin elastic fabric. Includes two Arduino microcontrollers (Arduino Pro Mini), 9-axis Inertial Unit (IMU, InvenSense MPU-9150), 3 flexible bend sensors (Spectra Symbol's flex sensors), 3 vibrotactile actuators (Precision Microdrives 10 mm Shaftless Vibration Motor).

AR's egocentric view does not allow easy virtual exploration or visual navigation planning. Our realistic 3D city model compensates for this, allowing free motion of the viewpoint in the full 3D environment (Fig. 2b). Furthermore, the default interaction mode for the 3D map follows the AR pointing metaphor: the initial view position and orientation are aligned with the position of the user and the orientation of the device. Users can choose either AR or 3D representation, with a smooth transition. The 3D map implementation is based on the m-LOMA mobile 3D city map engine, utilizing visibility pre-processing, level-of-detail management, and temporal coherence for optimal resource usage [29].

4.5 Integration and Transitions in an Urban Exploration Scenario

The integrated system was deployed on an nVidia Shield tablet. The integration involved communication with a backend (recommendation engine, content databases, and an external routing service), gathering the visual components (3D maps plus AR) and communicating with the audio-haptic devices (glove, headset, and vest) through the Bluetooth protocol. All the haptic devices include inertial measurement unit (IMU) for computing the relative direction of the device, while the mobile device computes the direction between the user and the POI. Table 1 reports implementation details and type and amount of curated cultural content produced as part of this work, including 3D models, POI locations, and animations.

A typical scenario highlights how users can benefit from the integrated system:

Eric is visiting Padua for the first time. Having little knowledge about this city, he simply selects a few major POI locations of his interest from the main interface. Combining his profile data, current location, and time, the recommendation engine suggests additional POI locations that could be of Eric's interest. Based on this recommendation, Eric adds some additional POI locations to his initial selection and uses the system to get a personalized itinerary for a comfortable walking route. Thanks to the non-visual modality on the system interface, as he is proceeding along the proposed itinerary, he is able to see the city better with his own eyes while still being guided towards the next destination. Not needing to check the visual interface constantly, he walks safely in the busy traffic ahead. The system is now directing him to make a right turn by triggering a vibration cue (tapping on the shoulder, a pattern indicating intensity of turn on corresponding hip) on his body. However, his eyes are caught by an old building with a Baroque-style façade. Out of curiosity, Eric raises his hand and points at the façade. Sensing vibration on the finger and hearing a church choir-like auditory icon on the headset, he understands that the place might have a religious meaning, and more information is available. After "clicking" on the façade with his finger, he hears the story of the place, as if he was using an audio guide in a museum. He decides on a detour to explore the area more. The system instantly re-calculates a new itinerary with new local recommendations. When approaching a square, he spots an interesting-looking clock tower. He wonders how such a clock worked and decides to learn more about it. After retrieving his tablet, he switches to a visual interface and aims it at the tower. Through the AR app, the astronomical clock

starts animating, while a narrating voice describes the clock and its principles. Feeling satisfied to have learned something new, he replaces the tablet in his bag and continues exploring. He feels immersed in the city, since he is able to naturally interact with the real surroundings. Now, he senses vibration on his index finger and hears the sound of an espresso machine. It is around 3 pm, and the recommendation engine notifies him of a highly rated café nearby. The tactile guidance on his hand directs him to his right-hand side. Looking at where he is pointing, he spots the café and enters it.

To make possible the complex scenario described above, we used a user-centered approach in the whole design process, where the final outcome derived from iterative refinement of prototypes was tested with potential users. The resulting integrated user interface allows users to easily input personal preferences, switch between AR, 3D, and audio-haptic interfaces, access POI content, and create flexible itineraries. To design more specific and complex interactions we engaged in co-design with potential users. For example, we used a focus group with students from Padua to design an interaction technique for selecting a POI while in AR mode, without needing to hold the tablet single-handed. The resulting selection technique enabled easy operation with both hands holding the tablet. The reticle in the center of the screen can be used to aim at POI icons by moving the tablet (Fig. 4a). When the reticle matches with a POI, users can tap with the thumb on the target button on the edge of the screen to confirm selection.

To further improve safety and experience, we designed a smooth interaction technique for modality switch. In our system, switching from visual mode to haptic mode simply requires placing the tablet in a bag (Fig. 5). The orientation and proximity sensors on the tablet can effectively detect the device status. When the tablet is retrieved from the bag, the system automatically switches back to the visual mode.

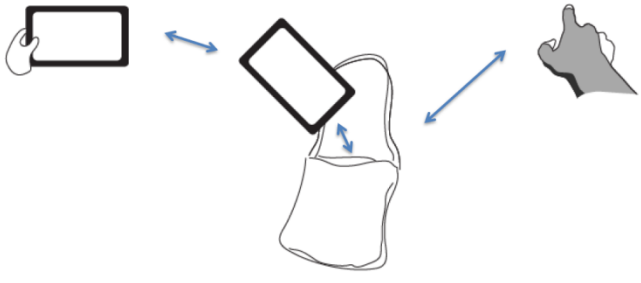

Fig. 5. The modality switching technique. When the mobile device is in landscape orientation and the screen is not covered, visual AR is activated, and audio-haptic exploration is disabled to avoid false positives. To switch from visual AR to audio-haptic exploration, the user simply covers the device screen (e.g., by storing the device in a pocket). To return to the visual AR interface, the user can simply uncover the tablet. This mechanism was reliably done with on-device accelerometers and proximity sensors that are commonly available on modern mobile devices.

5 Studying Feature Integration and Transitions in MRUE

To understand how the different features of the integrated system could co-exist, we conducted a user study where participants used our system in the central area of a historic city [15]. However, for practical reasons, we only investigated the effect of having the visual AR (cf. 4.3) and the wearable audio-haptic interfaces (cf. 4.2) integrated into the same application. The 3D engine (cf. 4.4) was running in background to enable visibility checks on surrounding POIs and avoid receiving content regarding a POI while pointing at an irrelevant building that occludes the designated POI. We disabled smart recommendations, flexible itineraries, and VR for the purpose of the study. The technique described in Figure 5 was used for switching between the different modes of exploration.

One main goal of the study was to understand the benefits of each of the main modalities under different tasks and to observe the switching between interfaces, specifically the circumstances under which the switching behavior occurred. Participants used the nVidia Shield tablet for the visual AR exploration (Fig. 4), the haptic glove and a headset for the audio-haptic exploration (Fig. 3a and Fig. 3b), and a pocket for storing the tablet and enabling the modality switch technique described in Fig. 5.

5.1 Procedure

A total of 18 volunteers were asked to use our system in three different conditions:

- *VAR*: visual AR only
- *HA*: haptic-audio guidance only
- *VAR+HA*: combined visual AR and haptic-audio guidance

The separation of modalities was used to understand what kind of augmentation is helpful to a given activity. The experiment was conducted in three nearby squares; this enabled the separation of the three interface modes. The assignment of squares to conditions was randomized, as well as the order of the VAR and HA condition. The third condition was always VAR+HA, where participants were free to choose which interface to use, also switching interfaces during the tasks. For each of the three conditions, participants performed two tasks: a *location task,* in which the participants were walked to a predefined starting point in the square and asked to locate an unknown POI that was selected in a preparatory phase, and an *information task,* in which they were asked to explore the available POIs and find information about a specific POI of their choice.

5.1 Data

Participants were asked to comment on their experience by answering four questions in a semi-structured interview. More specifically, they were asked what impressed them

about the experience, what were the strengths and weaknesses of the application, what were the most innovative elements, and whether they had the feeling to be more focused on the devices. We also extracted data about the efficiency of interaction. More specifically, we extracted the *task completion time* from the video recordings, while the *number of POI selections* in each task was derived from system log. Finally, we analyzed video recordings to identify the occasions and the circumstances (still or on the move) when interface switching took place in the combined VAR+HA condition.

5.6 Findings

From our quantitative findings, we observed that regardless of the task, using visual AR led to more POI selection than using haptic-audio guidance, but with similar task completion time. Qualitative findings from the interviews show how the system was overall well accepted, considered pleasant, and allowed a sense of presence in the information space. However, it was, at times, considered bulky and distracting. In the VAR+AR condition, we observed spontaneous switching of interfaces from most participants. Switching interfaces required participants to stand still in the majority of the cases. When using the integrated system, the participants spontaneously switched interfaces more frequently from the audio-haptic interface to visual AR than the contrary. Findings showed similar task completion time when using the preferred and less preferred interface, suggesting that participants could effectively use both modes despite their preferences. People switching from visual AR to the audio-haptic mode reported that they preferred to listen to the audio rather than reading the textual description of the POIs, while participants who switched from the audio-haptic mode to visual AR mentioned that they wanted to verify the directions provided by the audio-haptic interface, to get information faster or to be more in control of the application. In general, participants found that visual AR allowed them to get a good overview of all POIs available in the surroundings, with an accurate localization of the landmarks, while the audio-haptic interface was considered more amusing, innovative, and allowed users the liberty to obtain information without a screen interposed between them and the environment.

6 Conclusions

MRUE involves several challenges and opportunities. In this work, we propose design principles to consider when designing for MRUE, including supporting personalization, balancing plans and serendipity, keeping engagement with the urban environment, and supporting diverse modalities. We embodied these principles in a comprehensive MRUE framework including VR, AR, and haptic-audio interaction, and advanced features such as personalized recommendation, social exploration, and itinerary management. Having a large set of complex features in the same framework allows one to better address challenges and implement design principles. In our framework, for example, several components contribute to address the challenge of the vast amount of POI locations: the recommendation engine avoids showing items that are of no interest to the current user by providing a level of personalization, the 3D engine further allows

one to perform a visibility check that identifies and hides occluded POI locations, and the view management feature of the AR mode permits merging several items in single elements that can later be expanded, further reducing the mental load of the user. Similarly, several components contribute to serendipitous discovery by providing an extended awareness of the surroundings. Our study showed how personalized MR affordances and cues implemented through haptic-audio and visual interfaces complemented each other in addressing the complex and evolving needs faced by urban explorers, with visual AR used to get a good overview of all nearby POIs and the audio-haptic guidance used to explore individual POIs while keeping the focus on surroundings.

Open issues include designing how the different features are orchestrated. While multimodal MR interfaces can support each other, enriching the information delivered, how to properly orchestrate them requires considering the state of the user; for example, if users engage with a visual interface, interrupts from other modalities might not be considered helpful. In this work, we implemented the minimum requirement of allowing users to switch modality easily. However, this is just the first step toward acceptable interfaces for MRUE. The leap forward would come from appropriate orchestration of the full spectrum of modalities.

The current fast-paced technological advancement provides additional challenges and opportunities for future developments. Smart glasses have the potential to provide explorers with an even more immersive and safe experience once the technology would be mature enough. In addition, smart wearables and fitness gadgets are currently part of an exploding market rife, with small wearable technologies that track and respond to physiological data, which, more recently, are coupled with expanding assistive technologies. These technologies would allow to design recommendations that are more advanced and automatically assess when one modality could be more appropriate than another while allowing a safer and more engaging experience. Those considerations suggest that while novel technology will have the potential to change the way we experience urban sites, the fundamental needs and principles identified in our design framework would still remain valid and should be taken into account when designing novel interfaces and applications for MRUE.

Acknowledgments. This work was supported by the European Union Seventh Framework Programme (grant agreement nº 601139) and the Italian Ministry of Education, University, and Research (MIUR) PON AIM project (id: AIM1875400-1, CUP: B74I18000210006).

References

1. Hayllar B., Griffin T., Edwards D.: City Spaces - Tourist Places, Routledge, (2010)
2. Ashworth G., Page S.J.: Urban tourism research: Recent progress and current paradoxes, Tourism Manage., 32, pp. 1–15 (2011)
3. Brown B., Chalmers M.: Tourism and mobile technology. ECSCW 2003. pp. 335–354. Springer Netherlands (2003)
4. Tamminen S., Oulasvirta A., Toiskallio K., Kankainen A.: Understanding mobile contexts., http://dx.doi.org/10.1007/s00779-004-0263-1, (2004)

5. Ardissono L., Kuflik T., Petrelli D.: Personalization in cultural heritage: the road travelled and the one ahead, User Model. User-adapt Interact., 22, pp. 73–99 (2012)
6. Vaittinen T., McGookin D.: Phases of Urban Tourists' Exploratory Navigation: A Field Study. Proceedings of the 2016 ACM Conference on Designing Interactive Systems. pp. 1111–1122. Association for Computing Machinery, New York, NY, USA (2016)
7. Hockett P., Ingleby T.: Augmented Reality with Hololens: Experiential Architectures Embedded in the Real World., http://arxiv.org/abs/1610.04281, (2016)
8. Andolina S., Pirrone D., Russo G., Sorce S., Gentile A.: Exploitation of Mobile Access to Context-Based Information in Cultural Heritage Fruition. 2012 Seventh International Conference on Broadband, Wireless Computing, Communication and Applications. pp. 322–328 (2012)
9. Emmanouilidis C., Koutsiamanis R.-A., Tasidou A.: Mobile guides: Taxonomy of architectures, context awareness, technologies and applications, Journal of Network and Computer Applications, 36, pp. 103–125 (2013)
10. Bekele M.K., Pierdicca R., Frontoni E., Malinverni E.S., Gain J.: A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage, J. Comput. Cult. Herit., 11, pp. 1–36 (2018)
11. Bekele M.K.: Walkable Mixed Reality Map as interaction interface for Virtual Heritage, Digital Applications in Archaeology and Cultural Heritage, 15, pp. e00127 (2019)
12. Boletsis C., Chasanidou D.: Smart Tourism in Cities: Exploring Urban Destinations with Audio Augmented Reality. Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference. pp. 515–521. Association for Computing Machinery, New York, NY, USA (2018)
13. Zimmerman J., Forlizzi J., Evenson S.: Research through design as a method for interaction design research in HCI. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 493–502. Association for Computing Machinery, New York, NY, USA (2007)
14. Hsieh Y.-T., Jylhä A., Orso V., Andolina S., Hoggan E., Gamberini L., Jacucci G.: Developing hand-worn input and haptic support for real-world target finding, Pers. Ubiquit. Comput., 23, pp. 117–132 (2019)
15. Hsieh Y.-T., Orso V., Andolina S., Canaveras M., Cabral D., Spagnolli A., Gamberini L., Jacucci G.: Interweaving Visual and Audio-Haptic Augmented Reality for Urban Exploration. Proceedings of the 2018 Designing Interactive Systems Conference. pp. 215–226. Association for Computing Machinery, New York, NY, USA (2018)
16. Jylhä A., Hsieh Y.-T., Orso V., Andolina S., Gamberini L., Jacucci G.: A Wearable Multimodal Interface for Exploring Urban Points of Interest. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. pp. 175–182. Association for Computing Machinery, New York, NY, USA (2015)
17. Madsen J.B., Tatzqern M., Madsen C.B., Schmalstieg D., Kalkofen D.: Temporal Coherence Strategies for Augmented Reality Labeling, IEEE Trans. Vis. Comput. Graph., 22, pp. 1415–1423 (2016)
18. Morrison A., Knoche H., Manresa-Yee C.: Designing a Vibrotactile Language for a Wearable Vest. Design, User Experience, and Usability: Users and Interactions. pp. 655–666. Springer International Publishing (2015)
19. Orso V., Ruotsalo T., Leino J., Gamberini L.: Overlaying social information: The effects on users' search and information-selection behavior, Inf. Process. Manag., (2017)
20. Orso V., Varotto A., Rodaro S., Spagnolli A., Jacucci G., Andolina S., Leino J., Gamberini L.: A two-step, user-centered approach to personalized tourist recommendations. Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter. pp. 1–5. Association for Computing Machinery, New York, NY, USA (2017)
21. Tatzgern M., Orso V., Kalkofen D., Jacucci G., Gamberini L., Schmalstieg D.: Adaptive information density for augmented reality displays. 2016 IEEE Virtual Reality (VR). pp. 83–92 (2016)

22. Milgram P., Kishino F.: A taxonomy of Mixed Reality visual displays, IEICE Trans. Inf. Syst., E77-D, pp. 1321–1329 (1994)
23. Schmalstieg D., Hollerer T.: Augmented Reality: Principles and Practice, Addison-Wesley Professional, (2016)
24. van Dam A.: Post-WIMP user interfaces, Commun. ACM, 40, pp. 63–67 (1997)
25. Jacob R.J.K., Girouard A., Hirshfield L.M., Horn M.S., Shaer O., Solovey E.T., Zigelbaum J.: Reality-based interaction: a framework for post-WIMP interfaces. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 201–210. Association for Computing Machinery, New York, NY, USA (2008)
26. Schmalstieg D., Reitmayr G.: The World as a User Interface: Augmented Reality for Ubiquitous Computing. In Gartner, G., Cartwright, W., and Peterson, M.P. (eds.) Location Based Services and TeleCartography. pp. 369–391. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
27. Schmalstieg D., Wagner D.: Experiences with Handheld Augmented Reality. 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. pp. 3–18 (2007)
28. Reitmayr G., Drummond T.W.: Going out: robust model-based tracking for outdoor augmented reality. 2006 IEEE/ACM International Symposium on Mixed and Augmented Reality. pp. 109–118 (2006)
29. Nurminen A.: Mobile 3D City Maps, IEEE Comput. Graph. Appl., 28, pp. 20–31 (2008)
30. Dünser A., Billinghurst M., Wen J., Lehtinen V., Nurminen A.: Exploring the use of handheld AR for outdoor navigation, Comput. Graph., 36, pp. 1084–1095 (2012)
31. Oulasvirta A., Estlander S., Nurminen A.: Embodied interaction with a 3D versus 2D mobile map, Pers. Ubiquit. Comput., 13, pp. 303–320 (2009)
32. Fröhlich P., Oulasvirta A., Baldauf M., Nurminen A.: On the move, wirelessly connected to the world, Commun. ACM, 54, pp. 132–138 (2011)
33. Baldauf M., Fröhlich P., Masuch K., Grechenig T.: Comparing viewing and filtering techniques for mobile urban exploration, Journal of Location Based Services, 5, pp. 38–57 (2011)
34. Vaittinen T., McGookin D.: Uncover: supporting city exploration with egocentric visualizations of location-based content, Pers. Ubiquit. Comput., 22, pp. 807–824 (2018)
35. McGookin D., Tahiroğlu K., Vaittinen T., Kytö M., Monastero B., Carlos Vasquez J.: Investigating tangential access for location-based digital cultural heritage applications, Int. J. Hum. Comput. Stud., 122, pp. 196–210 (2019)
36. Ventä-Olkkonen L., Posti M., Koskenranta O., Häkkilä J.: Investigating the balance between virtuality and reality in mobile mixed reality UI design: user perception of an augmented city. Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. pp. 137–146. Association for Computing Machinery, New York, NY, USA (2014)
37. Lemmelä S., Vetek A., Mäkelä K., Trendafilov D.: Designing and evaluating multimodal interaction for mobile contexts. Proceedings of the 10th international conference on Multimodal interfaces. pp. 265–272. Association for Computing Machinery, New York, NY, USA (2008)
38. Grubert J., Langlotz T., Grasset R.: Augmented reality browser survey Institute for computer graphics and vision, University of Technology Graz, technical report, 1101, pp. 37 (2011)
39. Arth, Pirchheim, Ventura, Schmalstieg, Lepetit: Instant Outdoor Localization and SLAM Initialization from 2.5D Maps, IEEE Trans. Vis. Comput. Graph., 21, pp. 1309–1318 (2015)
40. Hartmann K., Ali K., Strothotte T.: Floating Labels: Applying Dynamic Potential Fields for Label Layout. Smart Graphics. pp. 101–113. Springer Berlin Heidelberg (2004)
41. Kruijff E., Swan J.E., Feiner S.: Perceptual issues in augmented reality revisited. 2010 IEEE International Symposium on Mixed and Augmented Reality. pp. 3–12 (2010)