CORE
Metadata, citation and similar papers at core.ac.uk
Provided by International SERIES on Information Systems and Management in Creative eMedia (CreMedia)

Combining Intelligent Recommendation and Mixed Reality in Itineraries for Urban Exploration

Giulio Jacucci[1,3], Salvatore Andolina[1,2], Denis Kalkhofen[5], Dieter Schmalstieg[5], Antti Nurminen[2], Anna Spagnolli[4], Luciano Gamberini[4], Tuukka Ruotsalo[1,3]

[1]Helsinki Institute for Information Technology HIIT

[2]Department of Computer Science, Aalto University, Finland
name.surnanme@aalto.fi

[3]Department of Computer Science, University of Helsinki, Finland
name.surnanme@helsinki.fi

[4]Human Inspired Technologies Research Centre, University of Padova, Italy
name.surnanme@unipd.it

[5]Institute of Computer Graphics and Vision (ICG) Graz University of Technology, Austria
name.surname@icg.tugraz.at

ABSTRACT

Exploration of points of interest (POI) in urban environments is challenging for the large amount of items near or reachable by the user and for the modality hindrances due to reduced manual flexibility and competing visual attention. We propose to combine different modalities, VR, AR, haptics-audio interfaces, with intelligent recommendation based on a computational method combining different data graph overlays: social, personal and search-time user input. We integrate such features in flexible itineraries that aid different phases and aspects of exploration.

CCS CONCEPTS

• CCS → Human-centered computing → Human computer interaction (HCI) → HCI theory, concepts and models

KEYWORDS

Intelligent recommendation, Augmented Reality, Urban Exploration , Multimodal Interaction

1 INTRODUCTION

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. This offers great opportunities for personalised exploration but also the challenges. Firstly considering the choice of interaction modality aggravated by limited manual flexibility and competing visual attention for safety. Moreover, the large amount of POI near the user or potentially reachable, suggests personalized recommendations which can make use of the extensive social data available on the web, e.g., tags, ratings and reviews. We propose to combine multimodal AR and intelligent recommendations for a new generation of urban exploration systems (Figure 1). The approach has been prototyped in an actual system integrating different modalities, such as virtual reality (VR), augmented reality (AR), and haptics-audio interfaces, as well as advanced features. Intelligent recommendations use a computational method combining different data graph overlays: social, personal and search-time user input. The integration of multimodality, MR and intelligent recommendations is synthesized in flexible itineraries that aid planning, serendipitous discovery and wayfinding.

2 REQUIREMENTS FOR THE URBAN EXPLORER

2.1 Facilitate serendipity

Most touristic applications are based on finding the route from visitor position to a particular Point of Interest (POI), e.g., a museum, restaurant, or church. Such an approach usually relies on geographical visualization of the environment (such as a 2D map) complemented with a route and POI list. This information is suitable for wayfinding, but less for the broader issues of exploration, like getting an overview or choosing a worthwhile destination. Facilitating serendipity comes 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 by delivering cues, either proactively or by responding to explicit user requests.

2.2 Connect the user to the surroundings

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

While exploring a place, the user's focus of attention should be on the surroundings. Virtual information (e.g., 2D maps or POI lists) redirect the user's attention to the device screen and spare little cognitive resource in exploring the surroundings. To avoid safety concerns and a possible negative impact on the quality of experience, technology for urban exploration needs to redirect the attention of the visitor from the device to the surroundings. Multimodal displays provide the necessary redundancy so that the user can remain connected to the surroundings, while receiving visual, haptic, or aural cues from the system.

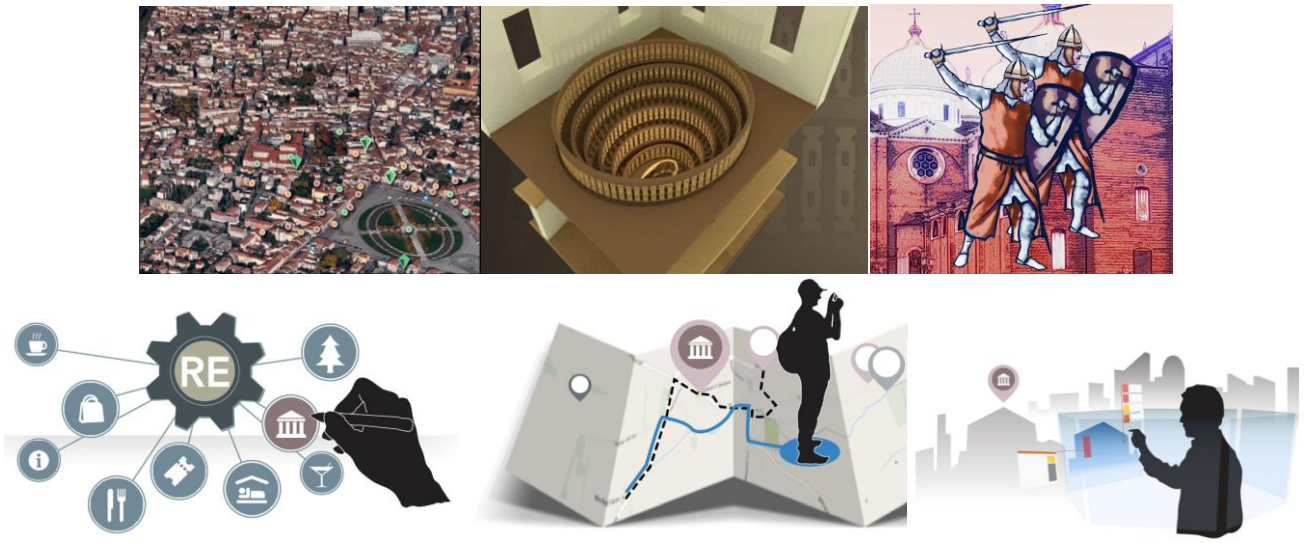

Figure 1: **Above example of cultural content in Padova, Italy, including POIs of different categories copiously distributed around the city with several dense areas, models, narratives of interiors or exterior artefacts such as the famous anatomical theater, or the history rich Abbey of Santa Giustina. Below: core support components responding to requirements of the explorer of urban environments with dense POIs: personalized recommendation, itinerary support, multimodal interaction responding to mobile and situational hindrances.**

2.3 Personalize Recommendations and Search

A rich urban environment can contain a large number of POIs. Yet, identifying relevant information is challenging with common search tools, resulting in an impoverished experience. For example, the same top-rated POI would be pushed to all nearby users. It is crucial to combine search engines with personalized recommendations based on the user profile 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 POI available in a particular place. This requires a model of users and their context, select the most appropriate content, and deliver it in the most suitable way (Ardissono et al., 2012). We also see a need to acknowledge variability in the visitors' profile. Users should be allowed to transfer from one profile to another one, rather than being confined to a static (or not easily changeable) profile.

2.4 Respect perceptual requirements

Multimodal interaction methods aim at improving human-computer communications by utilizing all available modalities of human input and output, in a natural manner (Turk 2014). Multimodality can free cognitive resources from the device toward the environment. This is especially important in a mobile setting (Lemmelä et al. 2008). Depending on the situation, users can choose the modality providing the most suitable level of safety, obtrusiveness, social acceptance, level of detail, and so on. For example, the visual modality may provide more detailed information, but requires more cognitive load. Multimodal access also makes information more accessible for disabled people. However, a user augmented with an interface supplying information through multiple modalities may easily be overwhelmed by information overload, or may fail to acknowledge information which is presented in a way that does not meet perceptual requirements. Occlusions on the screen, or clutter in general (including auditive and haptic jamming) defeat the objective of augmentation. Perception management must include strategies that avoid such situations. For visual displays, view management is required to ensure visibility of both real objects to relocate POIs to inconspicuous regions of the display (Bell at al., 2001). Multimodal interfaces need to generalize across multiple display dimensions (visual, audio, haptics) and decide where and how information items should best be displayed. These presentation management tasks require continuous optimization and can be computationally demanding.

3 COMPONENTS FOR URBAN EXPLORATION

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

3.1 Personalized Social Discovery

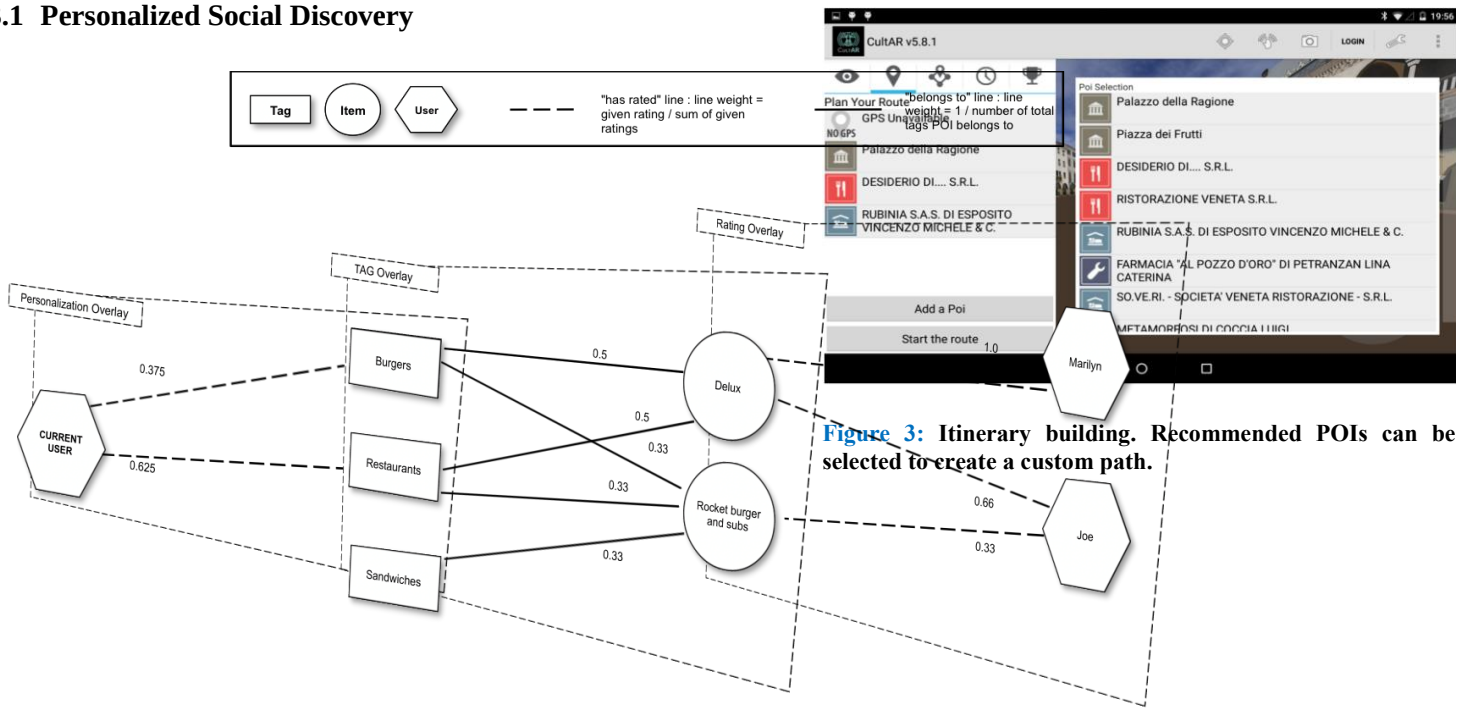

Figure 2: Social web data are modeled as a set of overlaid graphs composed of social tags, social ratings, users' personal preferences, and search-time preferences observed from the interaction with the system. Each of these data are connected to the retrievable items on a graph overlay.

We used two criteria for generating recommendations, one based on social rating and the other based on the visitors' profile. Social rating, as used in Tripadvisor or Booking.com, is a Web 2.0 technology, which analyzes the experience of other visitors to single out highly rated POI. Its credibility relies on two aspects: First, the information is neutral, since the ratings come from peer visitors. Second, reliability increases with the number of collected ratings, because unfair ratings are diluted with increasing sample size. Alas, the reason for certain destinations becoming highly popular are obscure and may be owed to reasons that do not concern the inquiring visitor, especially, if visitor deviates from the evaluation group.

Therefore, recommendations in our system are always considering the user's profile. Weighting by both crowd-sourced ratings and by personal preferences better facilitates serendipity, since recommendations can be based on either. Users may also turn off either ratings or profiles (or both), allowing unfiltered access to all POI information.

Our technique for personalized search and recommendations is based on three main parts (Orso et al. 2017): a data model that defines the representation of multiple data sources (content, social, and personal) as a set of overlaid graphs (see Figure 2), the relevance-estimation model that performs random walks with restarts on the graph overlays and computes a relevance score for information items, and the user interface (Figure 2).

3.2 Flexible Itineraries

Figure 3: Itinerary building. Recommended POIs can be selected to create a custom path.

Recommended POIs as delivered by the component in the previous section can be flexibly managed in planning and during visits providing support for situational plan changes and wayfinding (see Figure 3 for the UI where the user can choose recommended POI for an itinerary). In order to do some planning and optimize the visit, the tourist plans a path connecting the different POIs that are retrieved by the recommendation engine, creating a customized itinerary. The itinerary is built with the help of a routing service (e.g., Nokia Here). However, regular way-finding is replaced by a joint space+time optimization concerning the recommended POIs: Which set of POIs could be visited this morning? Which area of the city should be visited today? Alternatively, itineraries curated by professional tour guides can be retrieved, which highlight a topic of the urban environment (e.g., religious places, art museums, famous nightlife places or restaurants) or even tell a story (e.g., life of Galilei). Both types of itineraries do not force the visitor to follow the suggested path. Instead, the itinerary is merely an initial suggestions, 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.

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

3.3 Wearable and haptic guidance

Deploying aural and tactile channels as a complement to a visual channel into the system would benefit the users in terms of safety, while enhancing awareness of the surroundings. Touch based visual interfaces have presented high responsiveness and enabled integrated input and output concentration on the same surface. However, this comes at the cost of drawing much visual attention of users onto the interface and would result in ignoring surrounding environment, which could raise concerns in the context of designing for urban exploration. Eyes-free interaction could be an alternative to visual interfaces, releasing the visual attention back to the environment and increasing awareness.

The advance of wearable technologies enables increasing awareness of the surrounding as well as a more engaged experience through interaction design. For example, a vibrotactile vest could offer a hands and eyes free navigation, leveraging both location on the body and distinctive vibration patterns to indicate cues such as direction, degree of turn, speed or user error. This literally enables embodied interaction with the environment. The wearers are steered towards their chosen POIs, prompted in a 'natural-enough', pleasant and easily understood manner. Similarly, a headset capable of 3D audio could not only provide essential textual information, but also spatial information of the environment, such as distance, increasing or decreasing proximity and direction of a POI as well as category of POI (e.g., cultural, shop or service). Moreover, a sensor and actuator equipped glove provides hand gesture recognition for direct interaction with the real environment, such as selection via pointing. While a POI is distant and untouchable, it could be represented through the vibration on the hand. The non-visual multimodal interaction is complemented visually by the real scene seen by the user. Data from a comparative study with a context-aware mobile app has shown that, while experiencing similar performance of different evaluation metrics, smartphone users spent in average 70% of time looking at the screen when exploring the urban area, while users wearing our haptic glove (Figure 4) were able to have a good exploration experience while leaving their visual attention on the surroundings (Jylhä et al., 2015).

3.4 Augmented Reality of Urban Artifacts

Mobile AR browsers have recently become very popular (Grubert et al., 2011). 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, which is able to precisely estimate the camera pose of the user's AR device. Our implementation includes a localization system based on an analysis of visual features and an incremental tracker. The goal of our system is to localize a query image (the current video frame) within a known 3D world. We represent the world using a 3D point cloud and we accurately identify the 6-degrees-of-freedom pose (i.e. position and orientation) of the acquired camera image with respect to the 3D world model by searching for corresponding points within the 3D model and the 2D camera image.

Figure 4: Left the haptic glove; and right the audio-haptic exploration, in which a user points at a POI with the haptic glove and listens the associated audio description.

On top of the localization system, we have developed a fast incremental 3-degrees-of-freedom tracker in order to update the user's pose after its initial localization. For AR in urban environments, typical scenarios use 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.

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 (Tatzgern et al., 2016) 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 (Figure 5).

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. In the case of Figure 4 the user can activate an animation on the . astronomical clock (Figure 6).

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

Figure 5: **Mixed reality cues and affordances through computer-vision-based AR provide visual augmentations of a POI.**

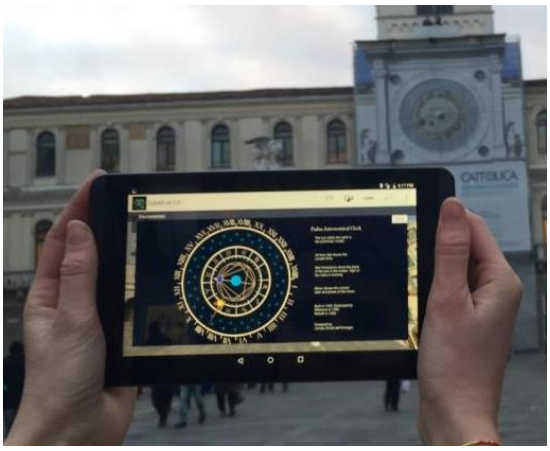

Figure 6: **The animated explanation of the astronomical clock activated from the AR interface in Figure 4.**

3.5 3D Engine and Virtual Reality.

As an alternative to AR we provide a virtual reality view of the environment. The AR view has the fundamental property of being egocentric. On one hand, it provides intuitivity due to the short cognitive distance between the real view and the view on the screen—they are always aligned (see section Augmented Reality of Urban Artifacts for AR Content Registration). However, 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. A label that is supposed to be on a building facade will hang in front of everything. This contradicts 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. 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 occluder. For the first case, we offer an alternative.

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 (Figure 7). 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 preprocessing, level-of-detail management, and temporal coherence for optimal resource usage (Nurminen, 2008).

Figure 7: **Virtual Exploration.**

4. CONCLUSIONS AND WORKSHOP INPUT

In the approach presented we demonstrated the combination of machine learning and Mixed Reality in exploring urban environments. In particular we show how a recent machine learning technique utilizing social media data can support personalization for recommendation and search in mobile VR and AR in urban settings. The approach has been evaluated only in partial experimentation with separate components , however it demonstrates important opportunities in benefiting from big data such as social media reviews, ratings and tags for application in MR. Recently important investment have been made in big data including ambitious promises of benefits. However only recently approaches have started to (e.g. cognitive big data Lugmayr et al. 2017).

According to estimates, our digital universe is growing at increasing speed. From 2013 to 2020, it is expected to grow from 4.4 trillion gigabytes to 44 trillion by a factor of 10 (EMC Digital Universe with Research & Analysis by IDC). Not only does it seem to more than double every two years, but the proportion of analyzed and used data, which is merely 20% now, will double to 40% by 2020.

Employing machine learning and making use of Big data includes a variety of challenges and opportunities for human-computer interaction. In particular it is interesting to investigate what interaction techniques are particularly interesting or should be researched in mining, analyzing, searching and exploring data? These might include, for example, implicit (Eugster et al 2013, Barral et al 2017, subliminal (Zhu et al 2016, Aranyi et al 2016) or

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal

gestural and multimodal interaction (Zhu et al. 2016, Jylhä et al 2015).

The approach taken has focused on using machine learning to personalize recommendations using social media thanks to the integrative role flexible itineraries accessible in VR and AR interfaces. We believe this is however only one possible instance of utilizing machine learning and big data in MR for urban exploration.

Other large cultural data sets including pictures, videos, 3D models and texts (Lugmayr et al. 2016) can be exploited to enhance cultural experiences through machine learning techniques.

ACKNOWLEDGMENTS

This work was partially funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No 601139 (CultAR).

REFERENCES

[1] Aranyi, G., Kouider, S., Lindsay, A., Prins, H., Ahmed, I., Jacucci, G., ... & Cavazza, M. (2014). Subliminal cueing of selection behavior in a virtual environment. PRESENCE: Teleoperators and Virtual Environments, 23(1), 33-50.

[2] L. Ardissono, T. Kuflik, and D. Petrelli, "Personalization in cultural heritage: the road travelled and the one ahead," User Modeling and User-Adapted Interaction, Vol. 22, no. 1, 2012, pp. 73-99.

[3] Barral, O., Kosunen, I., Ruotsalo, T., Spapé, M. M., Eugster, M. J., Ravaja, N., ... & Jacucci, G. (2016). Extracting relevance and affect information from physiological text annotation. User Modeling and User-Adapted Interaction, 26(5), 493-520.

[4] B. Bell, S. Feiner, and T. Höllerer, "View management for virtual and augmented reality," Symposium on User Interface Software and Technology, 2001, pp. 101–110.

[5] Eugster, M. J., Ruotsalo, T., Spapé, M. M., Kosunen, I., Barral, O., Ravaja, N., ... & Kaski, S. Predicting Term-Relevance from Brain Signals.Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval (SIGIR '14). ACM, New

[6] York, NY, USA, 425-434.
J. Grubert, T. Langlotz, R. Grasset (2011). Augmented reality browser survey, Technical Report 1101, Institute for Computer Graphics and Vision,

[7] University of Technology Graz, 2011
Jylhä, A., Hsieh, Y. T., Orso, V., Andolina, S., Gamberini, L., & Jacucci, G. (2015, November). A Wearable Multimodal Interface for Exploring Urban Points of Interest. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (pp. 175-182). ACM.

[8] A. Lugmayr, B. Stockleben, C. Scheib, and M. Mailaparampil, "Cognitive Big Data. Survey and Review on Big Data Research and its Implications: What is Really 'New'? Cognitive Big Data!," Journal of Knowledge Management (JMM), vol. 21, no. 1, 2017.: http://www.emeraldgrouppublishing.com /products/journals/ call_for_papers.htm ?id=5855kÝ, Journal of Knowledge

[9] Management/Emerald
A. Lugmayr, A. Greenfeld, A. Woods, and P. Joseph, "Cultural Visualisation of a Cultural Photographic Collection in 3D Environments – Development of 'PAV 3D' (Photographic Archive Visualisation)," in Entertainment Computing - ICEC 2016: 15th IFIP TC 14 International Conference, Vienna, Austria, September 28-30, 2016, Proceedings, G. Wallner, S. Kriglstein, H. Hlavacs, R. Malaka, A. Lugmayr, and H.-S. Yang, Eds. Cham: Springer International Publishing, 2016, pp. 272–277

[10] Nurminen, A. 2008. Mobile 3D City Maps. IEEE Computer Graphics and Applications 28, 4 (2008), 20–31.

[11] Orso, V., Ruotsalo, T., Leino, J., Gamberini, L., & Jacucci, G. (2017). Overlaying social information: The effects on users' search and information-selection behavior. Information Processing & Management, 53(6), 1269-1286.

[12] S. Lemmelä, A. Vetek, K. Mäkelä, and D. Trendafilov, "Designing and evaluating multimodal interaction for mobile contexts," Proc. 10th international conference on Multimodal interfaces, (ICMI '08), 2008, pp. 265-272.

[13] M. Tatzgern, V. Orso, D. Kalkofen, G. Jacucci, L. Gamberini, and D. Schmalstieg, "Adaptive Information Density for Augmented Reality Displays," In Proc. IEEE Virtual Reality, November 2016.

[14] M. Turk, "Multimodal interaction: A review," Pattern Recognition Letters, Vol. 36, 2014, pp. 189-195.

[15] Zhu, K., Ma, X., Chen, H., and Liang, M. (2016). Tripartite Effects: Exploring Users' Mental Model of Mobile Gestures under the Influence of Operation, Handheld Posture, and Interaction Space. International Journal of Human-Computer Interaction.

In: Artur Lugmayr, Kening Zhu, Xiaojuan Ma (edts), Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence Meets Virtual and Augmented Worlds (AIVR), International Series for Information Systems and Management in Creative eMedia (CreMedia), International Ambient Media Association (iAMEA), n. 2017/2, ISSN 2341-5576, ISBN 978-952-7023-17-4, 2017, Available at: www.ambientmediaassociation.org/Journal