VENTURI - immersiVe ENhancemenT of User-woRld Interactions

Alce, Günter; Chippendale, Paul; Prestele, Benjamin; Buhrig, Daniel; Eisert, Peter; BenHimane, Selim; Tomaselli, Valeria; Jonsson, Håkan; Lasorsa, Yohan; de Ponti, Mauro; Porthier, Olivier

2012

VENTURI – immersiVe ENhancemenT of User-woRld Interactions

Paul Chippendale[1], Benjamin Prestele[2], Daniel Buhrig[2], Peter Eisert[2], Selim BenHimane[3], Valeria Tomaselli[4], Håkan Jonsson[5], Günter Alce[5], Yohan Lasorsa[6], Mauro de Ponti[7], Olivier Pothier[7]

[1]Fondazione Bruno Kessler, Trento, Italy; [2]Fraunhofer HHI, Berlin, Germany; metaio GmbH, [3]Munich, Germany; [4]STMicroelectronics, Italy; [5]Sony Mobile Communications, Lund, Sweden; [6]INRIA, Grenoble, France; [7]ST-Ericsson, France

[1]chippendale@fbk.eu, [2](benjamin.prestele, daniel.buhrig, peter.eisert)@hhi.fraunhofer.de, [3]selim.benhimane@metaio.com, [4]valeria.tomaselli@st.com, [5](gunter.alce, akan1.jonsson)@sonymobile.com, [6]yohan.lasorsa@inria.fr, [7](mauro.de-ponti, olivier.pothier)@stericsson.com

Abstract: **The Augmented Reality (AR) concept has excited visionaries and researchers for decades, and finally its enabling technologies are beginning to materialize on consumer-grade mobile devices. The availability of powerful mobile computing platforms combined with fast mobile Internet access, opens the door to 'solid' visual/auditory augmentation of the world with geo-spatial knowledge, distilled from our personal digital lifestyles in social networks and the cloud. However, convincing AR has to date only been demonstrated on small mock-ups in controlled spaces. We have not yet seen key conditions being met to make AR a booming technology: seamless persistence and pervasiveness. This paper introduces the FP7 funded VENTURI project and its technologies, which aims to create a pervasive AR paradigm. The goal is to create an AR experience that is always present whilst never obstructing, delivering pertinent information in a 'user' rather than a 'device' centric way. This requires a change in how we think of and develop user interaction interfaces, making 'context' the starting point of all interactions. VENTURI addresses such issues, creating a user appropriate, contextually aware AR system, through a seamless integration of core technologies and applications on a state-of-the-art mobile platform.**

Keywords: Augmented Reality, Smartphone, Android, Machine Vision, Computer Graphics, Mobile Platform design

1 INTRODUCTION

Embedded mobile computing platforms are undergoing a fast evolution (powerful graphics processing unit, hi-res and high-quality cameras, wide variety of sensors, powerful multi-core CPUs and efficient power management policies), thus permitting increasingly complex and resource-demanding applications to be hosted on mobile devices. Miniaturised positional sensors and the availability of affordable "always-on" high-speed wireless data connections, are only now making it possible to implement AR scenarios beyond lab prototypes, targeting the consumer market on a larger scale. Accordingly, a complex ecosystem of AR technologies and content providers, software developers and industrial OEM mobile platform manufactures are currently co-evolving, striving to deliver all the bits and pieces needed to implement compelling AR scenarios.

In spite of this, the current AR experience still tends to be rather cumbersome and obtrusive to the user, typically requiring a lot of interaction to download and install applications, to print-out and position special markers for tracking, and to tap through on-screen menus to find the relevant information to be presented. Visual augmentation is still at an early state on mobile devices, and especially in outdoor environments it is often the case that AR overlays are far from integrating seamlessly into the scene perceived by the user. In this respect, AR today is just scratching the surface of its potential, requiring the user to setup a pre-defined context, rather than integrating naturally into the user's state. To make AR an unobtrusive and pervasive user experience, a more integrated approach is needed, which hides the complexity of the technology and puts the user in the focus of AR design.

To address this awareness, at the heart of VENTURI is an e-sensing philosophy, gathering as much information about a person's context as possible from collocated sensors and monitored geo-social activities. The platform will adapt intelligently in harmony with user context changes through the interpretation of not only the current surroundings and conditions, but also by monitoring past events that have left residual activity traces in the neighbourhood. This mobile context sensing task is very challenging and complex, stretching the current technology boundaries and the way machines perceive user context.

VENTURI brings together researchers, AR technology providers, mobile application developers, as well as Mobile Platform suppliers and mobile device manufacturers to take on this challenge and create an integrated platform for implementing such contextually aware AR scenarios. At the end of each project year, increasingly complex use-case prototypes will be demonstrated, carefully selected to evaluate all elements of VENTURI research and equally importantly to address the everyday needs of the public. The first year demonstrator encompasses an AR indoor gaming scenario, focussing on the initial integration of available project technologies; for the second year, an indoor personal-assistant and navigator for visually impaired people is planned, showcasing how context awareness in AR can really enrich peoples' everyday lives; the third year prototype will demonstrate the latest achievements in mobile AR algorithms and next-generation mobile Platforms, presenting an outdoor tourist guide and museum.

Corresponding author: Paul Chippendale, FBK, via Sommarive 18, Trento, Italy, +39 0461 314512, chippendale@fbk.eu

2 RELATED WORK

Recent AR software development in frameworks like (http://www.wikitude.org), (http://www.layar.com) and (http://www.junaio.com) have introduced a more interactive interface between users and the environment, offering people information relevant to their current location, using GPS and compass data to overlay points of interests (POI) into a live camera stream. However, this is still far from the seamless immersive user experience we desire, one in which collective knowledge becomes part of the real environment. In this regard, context-awareness needs to be more than simply location and time; it must take into account what users are sensing from their environment at a given moment.

Context-awareness, in terms of activities and geo-location, relies on a diverse repertoire of pervasive, state-of-the-art perceptual technologies (e.g. scene understanding, people detection, activity recognition, POI recognition, and object recognition). Instances of tracking and context recognition technologies in controlled indoor environments are nowadays reaching a robust and mature state, although in the outdoor environment many challenges still need to be overcome due to the complexity of unconstrained natural scenes (e.g. strong illumination variations and heavy occlusions).

Context can be identified using many different perceptual cues, and by utilising a wider variety (much in the same way as human beings do) a better understanding of the environment can be inferred, inevitably leading to AR deployment in everyday scenarios.

Scene understanding can be achieved from both the audio [1] and the video [2] domains. Inside the 'junaio' framework (realised by VENTURI partner metaio) marker-less 3D tracking of the environment has been implemented, exploiting feature detection approaches like SIFT (Scale invariant feature transform) [3] and gravity-aligned feature descriptors for improved stability [4]. In addition to enabling accurate and perspectively corrected overlays of 3D models (or videos), landmark POIs can also be better located through location cues, e.g. GPS. Bringing natural feature tracking and recognition technologies into the mix takes context-awareness to another level. User activity can be estimated through an analysis of device sensor data, e.g. from gyroscopes and accelerometers [5]. Context sensing approaches combine many of these sensing techniques and usually entail a feature extraction step followed by a classification, based on machine learning methods; hence performance is heavily influenced by the size and the data variability of databases used during learning.

The particular form factor restrictions of potential mobile AR devices has also been studied; Billinghurst et al. [6] described several advanced interaction techniques, such as tangible AR interfaces, multimodal interfaces and mobile AR interfaces, and more recently Xu et al. [7] described patterns for designing embodied interaction in handheld AR games to explore AR usability.

3 AUGMENTED REALITY IN VENTURI

To address the target goals of the next generation of AR, and to satisfy the needs of all of the stakeholders related to the VENTURI system, research and development within the project focuses on the following technological research objectives:

1. Enable developers and mobile device OEMs to create immersive and interactive environments by means of a mobile AR platform adapted to the users' devices. This will be based on the definition of an innovative system and software framework architecture, centred upon the upcoming generations of smartphone platforms.

2. Create a new paradigm for prosumer-friendly, immersive and multimodal media representation, targeting a mix of real and virtual worlds with improved interaction capabilities and exploring new sensing interfaces.

3. Exploit a more complete set of sensor information to provide a comprehensive understanding of a user's context (on the move, at home or at work) from multiple perspectives.

4. Create high-quality 3D-content and enable an efficient distribution of immersive media over the Internet, relevant and adapted to the user's estimated context.

5. Evaluate and optimize Quality of Experience for the end-user by improving an AR application's usage of mobile platform resources.

VENTURI aims to fulfil a dual purpose, as it aims to both provide and improve the capabilities of creating (for developers) and experiencing (for the end user) a fully integrated, immersive, easy to use, truly mobile and high performing AR-based environment. Accordingly, a unified approach among core audio, imaging, video and graphics technologies, embedded mobile platforms and man-machine interfacing is instrumental to the success of the project's vision. Technical challenges, such as advanced real-time image recognition, high-speed networking, content extraction and generation, massive multimedia processing, as well as extensions to the mobile platform software and hardware structure need to be solved within the project, and will be integrated incrementally into the VENTURI platform. This unified architecture will provide a more E2E-immersive technology for the creation and delivery of AR content, striving towards a more seamless and user-centric mobile AR experience.

4 TECHNOLOGIES INSIDE THE VENTURI PLATFORM

The research and ideas encapsulated within the VENTURI project fall within six distinct and intersecting areas:

- Mobile Hardware/Software Platforms
- Context Sensing
- Content Creation
- Mobile Content Delivery Modalities
- Ensuring Quality of Experience for the User
- Gathering & Fusing Content Appropriate for AR delivery

4.1. Mobile Hardware/Software Platforms

All VENTURI research activities and developments belonging to the aforementioned areas will be deployed, used, embedded and evaluated in a mobile demonstrator platform. Since the user-centric approach is key, it is of utmost importance to use platforms that are actually targeting a device that the user would use, in terms of form-factor, HW/SW components and maturity. VENTURI has adopted, from its partner ST-Ericsson, platforms that include dual-core ARM Cortex CPU, high-performance GPU, hi-resolution camera, a wide variety of sensors, touchscreen display, embedded power management infrastructure [22] and runs the latest Android OS and services. Android was selected as the OS of choice because of its outstanding technical and diffusion momentum.

The platform architecture is being designed in such a way that 1) it remains as much as possible compatible with standard Android releases in order to speed-up the subsequent product integration phase, while 2) permitting a flexible and smooth integration of VENTURI research outcomes for testing, evaluation and benchmarking purposes.

The AR software framework relies on the metaio Mobile SDK which enables the straightforward creation of AR applications on mobile devices. In the case of vision-based real-time AR, video grabbing, vision analysis, camera pose estimation and tracking, rendering needs to be done in less than 33 milliseconds to sustain a smooth 30 fps delivery. This illustrates the crucial need for global optimization of the whole AR processing chain, placing strain on all of the platform's resources.

The services provided by the HW, multimedia and the Android framework have been cross-referenced against AR-specific requirements. Modifications identified during the first six months of the project are being implemented, to globally optimize the image processing and the positioning sensor pipes to enable a better integration of the AR framework on the platform. Of these, a key requirement focusses on the accurate synchronization of video frames with sensor data. This addition will furnish the AR framework with video frames of the exact format and resolution which are temporally aligned with hardware pose sensors to enable hybrid sensor/vision pose estimation.

VENTURI SW developments are founded on portable open standards like OpenGL and OpenSL, and AR-specific algorithms are evaluated with regards to vectorization and parallelization, to take full advantage of the multi-core processing units. Open standards such as OpenCL should benefit from the high-performance computing capabilities of the mobile embedded GPU (GPGPU), whilst remaining platform-independent.

Overcoming the challenge to optimally partition and embed AR Algorithms, Middleware and Applications into such a structured heterogeneous multi-processing platform is instrumental in realising a smooth, seamless, and user-centric AR paradigm for mobile devices.

4.2. Context Sensing

At a high level, the context sensing model in VENTURI consists of information concerning both user and the nearby environment. Visual/audio scene classification, visible/audio object recognition, scene text/logo spotting, planar surface detection, user activity classification, user gesture and pose estimation all fall into this modelling task. The scene model is further enriched with semantic information extracted from web resources, e.g. POI databases, image databases and online social networks. The user model is similarly enriched with user information from online social networks and other user related sources.

Making inferences about user and ambient context addresses a problem with inherent uncertainty, since sensors only see an error-prone snapshot of the world. As VENTURI is targeting immersiveness in both visual and aural modalities, camera and microphone sensors play an important role and will be supported by measurements from additional sensors, e.g. network info, location, accelerometer, gyros, compass etc., to better understand context. In VENTURI, we are extending the notion of sensors to include non-structured human generated content, e.g. text messages or online social network posts, as suggested in [8]. These data sources also come with errors and need to be interpreted and semantically anchored. Using multiple sources of contextual information reduces uncertainty, but increases complexity.

A visual analysis of the environment is performed for pose-tracking, scene classification and for refining geo-position. We use image classification for both middle-sized objects using a generic object image database, and for identifying cultural information associated to monuments using nearby geo-tagged images retrieved from http://www.flickr.com, or by matching sky/land profiles against geo-referenced synthetic 3D-models [9]. Moreover, visual scene classification is used

Figure 1: Turntable capture setup (left). Preliminary 3D reconstruction results of a miniature house model (middle/right)

to categorize the observed scene into a set of semantic classes (natural, manmade, home rooms, street, supermarket, etc.), which helps the application to identify context and activates specific actions accordingly.

Multiple approaches to physical surroundings modelling are being investigated, including single camera sparse 3D map reconstruction, stereo-camera dense reconstruction, RGB-D parallel tracking and meshing [10], and the tracking of deformable objects. Furthermore, the 3D-localisation of text [11] is being used to further understand context.

A Pedestrian Dead Reckoning module is also being developed in VENTURI to act as a sensor fusion component. On-board accelerometer and compass are exploited to provide an estimation of the relative position and orientation of a user in 'urban canyon' environments. Combined with geographical data, this component can also be exploited to estimate user activity patterns, such as walking or running. In conjunction with this sensor analysis, audio is also being explored to identify sound objects in the scenes using semantic data from Open Street Map. Combining these methods with machine-learning algorithms will enable VENTURI enabled devices to better determine context sources and help to generate a comprehensive context model of the user and their environment, with the aim of predicting future contexts.

4.3. Content Creation

The presentation of high-quality visual content strongly contributes to the overall appeal and user experience of an AR application. However, creating such content is complex and typically requires a considerable amount of manual work. The reproduction of real-world objects as highly realistic 3D-models is particularly involving, as it comprises many processing steps, including 3D-data acquisition, object reconstruction, mesh processing, and texture map creation. Research in VENTURI will help to ease this workflow by developing methods to acquire and reconstruct high-quality 3D-models from dense data sets, through the implementing of an incremental set of tools that require as little user intervention as possible.

However, at times it may be difficult or even impossible to capture consistent data-sets for the purpose of AR content creation. Imagine for example the challenge of creating an AR historic city guide: some buildings and landmarks may look quite different nowadays than they did a century ago; they may be partially destroyed, or even have completely vanished. In such cases, AR content will be gleaned from sparse and dissimilar data, such as old photographs or even paintings. Research will therefore target specialized techniques for the 3D-creation of AR content from this type of data, striving to go beyond the acquisition of merely tangible objects.

Based on the requirements of the first year gaming demonstrator, visual content creation is initially focussing on 3D reconstruction from dense and consistent data-sets, targeting small-sized objects captured in a controlled environment, namely: hand-crafted miniature building models. Different reconstruction methods are being investigated to evaluate their advantages and drawbacks with particular classes of objects. Early studies are revealing that robust feature-based reconstruction approaches, such as [12], fail to reliably recover dense multi-view correspondences needed for 3D reconstruction. This is mainly due to the homogeneity and high repetitiveness of the surface structures of the model houses. Active structured light methods, similar to [13], are proving to be equally sensitive due to the high reflectiveness of plastic models. Additionally, such active approaches tend to produce strong outliers in areas of refraction, such as in the semi-translucent windows of the model houses. The preliminary results shown in the middle and right images of Figure 1, used a shape-from-silhouette [14] approach, which is very robust to illumination variations and is invariant with respect to object surface characteristics. For the acquisition of these dense data-sets, the miniature houses were placed on a PC-controlled turntable (see Figure 1 left) and silhouettes were record in 5° steps, from multiple perspectives. The images were semi-automatically segmented from their backgrounds to produce accurate silhouette masks. As the silhouette in the image-plane forms a 3D viewing cone under a given camera calibration, the intersection of all recorded viewing cones can be used to estimate an upper bound of an object's volume. The accuracy of this type of reconstruction is largely affected by the precision of the camera calibration. We utilized a reconstruction method described in [15,16], that simultaneously optimizes for silhouette reconstruction error and camera calibration, exploiting the circular trajectory of the turntable setup as a constraint. This enables an efficient reconstruction of an object's volume with high accuracy and resolution. The resulting voxel model is converted into a manifold polygon mesh by an enhanced marching cubes' method [17]. Finally, to create the consistent texture map for realistic visualization, we employed a rendering-based synthesis approach, exploiting the camera calibration estimated during the 3D reconstruction.

Multiple extensions and refinements to the model creation process are under development. Current work is concentrating on fusing the volume-based reconstruction method with a pixel-dense image registration approach in order to handle concave object surfaces, which are not recoverable by a purely silhouette-based approach. Additionally, methods for a more accurate reconstruction are being investigated, incorporating prior knowledge of building geometry.

For mobile AR applications in particular, hardware constraints, such as display resolution, CPU/GPU performance, and wireless bandwidth, need to be considered. Therefore, trade-offs are often needed, e.g. between model quality and model size in order to guarantee a fluent rendering on the mobile device and snappy updates from the content servers. On junaio, for example, developers are advised not to exceed 15000 polygons per model. Accordingly, different content creation tools are under development to support the deployment workflow, such as a tool for 3D-model simplification.

The results from all efforts detailed above, along with the evolution of mobile computing platforms, will provide the basis for tackling the even more challenging 3D content creation tasks of reconstructing outdoor objects, and the creation of 3D-content from sparse and dissimilar data, such as from old photographs or paintings.

Humans strongly sense and interact with the world using sound, hence AR applications must also include spatially aware audio cues to accentuate immersion within the augmented environment. Generating interactive geo-referenceable audio content must take into account multimodal user and scene context to enable the adaptation of audio soundtracks in real-time to the situation. In VENTURI, we use interactive audio techniques to react to user input and/or changes in the application environment. Audio content is created using specialized authoring tools and frameworks that separate the audio design and generation processes, enabling the end-user to create and customize audio content.

We use a new event-based XML language derived from the works on the A2ML [18] format and built on top of it a sound renderer which permits a user-selectable prioritization of the audio information. This approach is well suited to the needs of highly demanding applications such as guidance systems or gaming. In this way, a user can receive the most relevant information at a given time, limiting sound superposition for better intelligibility. This language enables the 'sonification' of AR applications by permitting a mix of small audio-chunks with synthesized speech, which can be arranged in real-time based on application events. An event synchronization system has also been created, based on SMIL [19], an XML language tailored towards multimedia content synchronization. In this way, audio content is interchangeable in the form of audio style sheets (in similar way to CSS), enabling the user to experience a different audio immersion according to context.

4.4. Mobile Content Delivery Modalities

In VENTURI, context awareness plays a key role in AR content delivery. Factors such as: delivery channel bandwidth limitations, a device's display resolution or battery life, or a user's current activity, can all advice a VeDi device and/or AR media-object server about how best to deliver information.

Detailed information from 3D visual tracking, object reconstruction and scene classification, will have a strong impact on AR content delivery and presentation. The proximity and line-of-sight to an object, for example, will determine the amount of content to be pre-loaded/requested from servers and presented to the user. Similarly, content should dynamically scale in complexity whenever a user stands still and concentrates on a specific point of interest, switching from unobtrusive spatialized audio-only augmentation to hi-definition video overlays. Such dynamic content selection and presentation strategy require a strong coupling between context sensing and content delivery and will be explored in depth in the project.

4.5. Ensuring Quality of Experience for the User

Beyond the visual augmentation experience, it is clear that future AR applications will also need new interaction and application models to facilitate new forms of communication and meet increasingly high user expectations [20]. This is a huge challenge since AR cannot rely on design guidelines for traditional user interfaces. New user interfaces permit interaction techniques that are often very different from standard WIMP (Windows, Icons, Menus, Pointer) based user interfaces [21]; WIMP interfaces share basic common properties, whilst AR interfaces can be much more diverse. Digital augmentation can include different senses such as sight, hearing and touch, hence they must be realized with an array of different types of input and output modalities such as gestures, eye-tracking and speech.

To make AR non-obtrusive and pervasive, user experience understanding is essential. Aspects such as: finding out how a user performs tasks in different contexts; letting the user interact with natural interfaces; and hiding the complexity of the technology; are central to ensure good quality applications and a good user experience. Quality of Experience will thus be guaranteed by iteratively performing user studies. In the initial qualitative user study, we will try to understand how weaknesses in technical stability influence user experience, e.g. if the recognition of a 'marker' is lost what would be the correct way to place overlaid graphics?, thus reducing the impact of spatial instability and camera lag. This will be followed up with further user studies in indoor and outdoors scenario, to get insights into social acceptance.

4.6. Gathering & Fusing Content Appropriate for AR delivery

To enable the re-deployment of existing content in AR scenarios, research efforts will be directed towards context sensitive delivery and the fusion of different data sources.

Various content retrieval methods are under investigation that will support example-based queries from multi-model databases, ranging from OCR-ed text to visual fragments.

Methods for the robust registration of 3D-models with 2D-images are being explored. These will help to realise novel scenarios that aim, for example, to drape historical paintings or 3D-texture maps (generated from user provided photos) into reality from arbitrary view points.

For an immersive and believable experience, such content will need to be wrapped and rendered into rich multimedia objects that integrate naturally into the real scene. Information from the diverse hardware sensors, user gesture analysis, pose estimation, and the context sensing tasks will need to be fused in order to manipulate content in response to user interactions, handle occlusions and collisions between virtual and real objects and smoothly adapt the illumination of virtual overlays.

5 VALIDATION SCENARIO

To validate the first-year principle elements of VENTURI, a hardware/software demonstrator (based on the STE NovaThor U9500 platform [22] and nicknamed VeDi 1.0) will be built, realising a table-top AR game. The game will take place in a real 70x90cm city model, which mimics an imaginary city block, with AR characters, objects and scenes being superimposed, interacting with the user. The player (or players in the case of multi-player mode) will be able to interact with objects in the city model, detected through the marker-less, visual 3D-tracking of the real model. In the game, players will enjoy an experience unobtainable using traditional methods. By grabbing hold of a VeDi device and activating the application, player will be immersed in an engaging virtual world, navigating virtual vehicles inside a real physical world. Users will be faced with different missions (e.g. fire-fighter mission) in which they will face time/pressure challenges. Moreover, to make the game more exciting, in the multi-player mode, other users will be able to place virtual or real 'obstacles' to hamper the other person's efforts.

The primary objectives of VeDi 1.0 are to bench-test existing technologies and show how the integration of different algorithms (e.g. 3D marker-less tracking, 3D audio placement, superposition of virtual models on real objects) can create a real sensation of blurring reality. VeDi 1.0 will demonstrate a solid and engaging AR experience thanks to its state-of-the-art platform and sensing advantages, giving developers a taste of what VENTURI is striving to achieve in the next three years of research.

Figure 2: VeDi 1.0 shown at Mobile World Congress 2012

6 CONCLUSION AND UPCOMING WORK

The VENTURI project introduced in this paper aims to create a pervasive AR paradigm built around mobile platforms and an extensive e-sensing philosophy. By exploiting the computational power and the mix of sensors available in current and next-generation mobile platforms, as well as sophisticated algorithm for audio-visual scene analysis and large-scaled social data mining, we believe that future AR applications can be driven by user context and will adapt to user needs, thus creating a more seamless AR experience. To empower this vision, a wide spectrum of challenges is addressed within the project, tackling areas such as: mobile AR platform optimization, audio-visual scene analysis, context sensing, gathering/creating/fusing/delivery of AR content, and mobile human-machine interactions. To this end, the project brings together researchers, AR technology providers, mobile application developers, as well as Mobile Platform and mobile device manufacturers, to create an integrated hardware and software platform that is capable of implementing the VENTURI vision.

Acknowledgements

This research is being funded by the European 7th Framework Program, under grant VENTURI (FP7-288238).

References

[1] Clarkson B., Sawhney N. and A. Pentland, "Auditory context awareness in wearable computing", Workshop on Perceptual User Interfaces, (1998)

[2] Battiato S., Farinella G.M., Gallo G. and Ravi' D., "Scene categorization using bag of textons on spatial hierarchy", in IEEE International Conference on Image Processing (ICIP-08), pp. 2536-2539, (2008)

[3] Lowe, D. G., "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, vol. 60, issue 2, pp. 91-110, (2004)

[4] D. Kurz and S. Benhimane. "Gravity-aware handheld augmented reality", IEEE International Symposium on Mixed and Augmented Reality, (2011)

[5] Avci A., Bosch S., Marin-Perianu M., Marin-Perianu R., Havinga P.J.M., "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: a survey", ARCS Workshops, pp. 167-176, (2010)

[6] Billinghurst, M., Hirokazu, K., Myojin, S., Advanced Interaction Techniques for Augmented Reality Applications, Springer, (2009)

[7] Xu, Y., Barba, E., Radu, I., Gandy, M., Shemaka, R., Schrank, B., MacIntyre B., "Pre-Patterns for Designing Embodied Interactions in Handheld Augmented Reality Games", IEEE International Symposium on Mixed and Augmented Reality, (2011)

[8] Aggarwal, C. C. and Abdelzaher, T. "Integrating sensors and social networks.", Social Network Data Analytics, Springer, Chapter 14, (2011)

[9] Chippendale P., Zanin. M and Andreatta C., "Spatial and Temporal Attractiveness Analysis through Geo-Referenced Photo Alignment", IEEE International Geoscience Remote Sensing Symposium, Boston, USA, (2008)

[10] Lieberknecht S., Huber A., Ilic S. and BenHimane S., "RGB-D camera-based parallel tracking and meshing", Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality, Basel, Switzerland, (2011)

[11] Messelodi S. and Modena C.M., "Scene Text Recognition and Tracking to Identify Athletes in Sport Videos", Multimedia Tools and Applications, Automated Information Extraction in Media Production, (2011)

[12] Snavely N., Seitz S. and Szeliski R., "Photo Tourism: Exploring image collections in 3D", in ACM Transactions on Graphics, SIGGRAPH, (2006)

[13] Fechteler P., Eisert P. and Rurainsky J., "Fast and High Resolution 3D Face Scanning", 14th International Conference on Image Processing, (2007)

[14] A. Laurentini, "The visual hull concept for silhouette-based image understanding", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no.2, pp. 150-162, (1994)

[15] Eisert P., "3-D Geometry Enhancement by Contour Optimization in Turntable Sequences", in Proc. IEEE International Conference on Image Processing (ICIP), Singapore, pp. 1947-1950, (2004)

[16] Hernandez C., Schmitt F. and Cipolla R., "Silhouette Coherence for Camera Calibration under Circular Motion", in IEEE Transactions on Pattern Analysis and Machine Intelligence, (2007)

[17] Lewiner T., Lopes H., Vieira A. and Tavares G., "Efficient implementation of Marching Cubes' cases with topological guarantees", in Journal of Graphics Tools, vol. 8, (2003)

[18] Lasorsa Y., Lemordant J., "An Interactive Audio System for Mobile", 127th AES Convention, (2009)

[19] Synchronized Multimedia Integration Language, W3C, http://www.w3.org/TR/SMIL2

[20] Barba, E., MacIntyre, B. and Mynatt, E. D., "Here we are! where are we? Locating mixed reality in the age of the smartphone", Proceedings of the IEEE, 100, pp. 929-936, (2012)

[21] Furht, B., "Evaluating Augmented Reality Systems", Chapter 13, pp. 289-307, (2011)

[22] ST-Ericsson, "NOVATHOR™ U9500", http://www.stericsson.com/products/u9500-novathor.jsp