Software Architecture Design for Spatially-Indexed Media in Smart Environments

Ovidiu-Andrei SCHIPOR^{1,2}, Wenjun WU³, Wei-Tek TSAI³, Radu-Daniel VATAVU^{1,2} ¹University Stefan cel Mare of Suceava, 720229, Romania ²MintViz Lab / MANSiD Research Center, Suceava 720229, Romania ³State Key Laboratory of Software Development Environment, Beihang University, Beijing, China schipor@eed.usv.ro, wwj@nlsde.buaa.edu.cn, wtsai@asu.edu, vatavu@eed.usv.ro

Abstract—We introduce in this work a new software architecture design, based on well-established web communication protocols and scripting languages, for implementing spatially-indexed media in smart environments. We based our approach on specific design guidelines. Our concept of spatially-indexed media enables users to readily instantiate mappings between digital content and specific regions of the physical space. We present an implementation of the architecture using a motion capture system, a large visualization display, and several smart devices. We also present an experimental evaluation of our new software architecture by reporting response times function of changes in the complexity of physical-digital environment.

Index Terms—software architecture, multimedia communication, ambient intelligence, augmented reality, smart homes.

I. INTRODUCTION

Consuming interactive multimedia content has become part of our everyday life, and a large variety of devices and interaction techniques as well as a wide range of mobile apps on smart devices create a deep, intertwined connection between our physical and digital reality [1-3]. In the most typical scenario, multimedia content is located on some device, e.g., an mp3 player, or is streamed on demand from a content server, such as YouTube. However, recent research efforts from tangible computing and augmented reality have unveiled new promising alternatives for enriched multimedia consumption in smart spaces. For instance, "tangible bits" [4] map digital content to physical objects with the goal to create meaningful shortcuts between the physical and digital realities by "taking advantage of human abilities to grasp and manipulate physical objects and materials" [5]. "Radical atoms" [6] implement this further by working with computationallyvision reconfigurable smart materials. Yet another example is given by virtual environments that wrap virtual and physical objects together [7] or by recent augmented reality apps with wide mass adoption, such as Pokemon Go, that enrich the physical environment with digital content by implementing smart context adaptive strategies [8].

This work was supported from the project Interact-Cloud, "Interaction Techniques with Massive Data Clouds in Smart Environments", project no. 47BM/2016, financed by UEFISCDI, Romania under the PNIII framework. Work was carried out in the MintViz Lab of the MANSiD Research Center, for which the research infrastructure was partially supported from the project "Integrated Center for research, development and innovation in Advanced Materials, Nanotechnologies, and Distributed Systems for fabrication and control", contract no. 671/09.04.2015, Sectoral Operational Program for Increase of the Economic Competitiveness co-funded from the European Regional Development Fund.



Figure 1. An example of an interactive scenario implementing spatiallyindexed media: digital content associated to a specific location (a) can be visualized on public (b) or personal displays (c).

By relying on the perspective enabled by augmented reality app design, we propose in this work smart environments that implement *spatially-indexed multimedia*, *i.e.*, interactive spaces in which digital media content is linked to specific regions of the physical space. Users of this space access, visualize, and manipulate digital content with smart mobile devices and ambient visualization surfaces. For instance, consider John, who wishes to create two video messages for his family. John uses his smartphone to record the two messages for his son and his wife. Then, he picks each message with his fingers from the smartphone's touch screen (see Figure 1a), carries them through mid-air, and "attaches" the messages to a region near the fridge (a place where his family has been sharing messages in the form of physical notes). When John's son arrives at home, he explores that region with his hand and the video message starts playing (see Figure 1b). John's wife could do the same, but she prefers to make a grabbing gesture in the direction of that region and toward her smartphone (Figure 1c). For John and his family, attaching digital content to physical regions of their home space has become second nature. Over the years, they have created several associations between the physical space of their home and digital media: next to the cooking table, there are many

Digital Object Identifier 10.4316/AECE.2017.02003

shortcuts to recipes and cooking videos; the regions next to the sofa contain links to music files and motion pictures, etc.

In this article, we examine viable software architecture options to implement spatially-indexed media in smart environments. Our main contributions are (1) the design of a multi-level software infrastructure for spatially-indexed media in smart environments with specialized dataflows, which we build on top of standard web communication protocols and scripting languages and (2) an experimental evaluation of the technical performance of our new software infrastructure.

II. RELATED WORK

In this section, we discuss related work on the design and development of software infrastructure for multimedia consumption in smart spaces. We review current practices in service-oriented software architecture and cloud computing and highlight the practicability and effectiveness of httpbased services over web; we discuss previous systems that implemented multimedia delivery in smart spaces; and we connect to recent results from augmented reality implementing visualization and interaction techniques with multimedia content on smart mobile devices.

The concept of a smart space was derived from the broader paradigm of ubiquitous computing [9]. Smart spaces refer to parts of the physical world where smart devices share data, information, and knowledge and collaborate to improve the life quality of the inhabitants of that space [10]. Computational objects form an ecosystem that constantly acquires data, derives information, and creates and applies knowledge in order to actively adapt to its users and improve their interactive experience [11]. A key feature of such an environment is the ability to mix the physical and the digital in natural and unobstructed ways [12].

Smart spaces can improve learning by allowing people to be immersed into environments that educate [13]. They can also help people with disabilities to perform everyday tasks [14, 15]. The Magic K-Room [16] and P3S [17] are two frameworks that enable children to interact with smart objects and immersive multimedia content. Smart spaces can also improve speech therapy with automatic speech and emotion recognition [18]. Elderly people can use flexible interaction techniques to execute tasks in a smart environment [19, 20]. Moreover, such an ecosystem can become not only a mediator, but also an active actor that boosts social interaction [21] and collaborative work [22].

Different implementations of smart spaces have led to various strategies to deliver multimedia content. One common feature is providing content according to contextual information, such as the user's physical location in that space or the user's profile and their interaction history [23]. Device properties, such as pixel resolution, display size, and input capabilities, are leveraged to provide an enriched experience for the users of such environments [24, 25]. In such spaces, users interact with content directly by gestures [26, 27] or through a smart mobile device [28]. New interactive techniques for smart spaces are "Smart-Pockets," a continuously emerging, such as technique that links pockets with digital content for efficient retrieval of personal digital content and visualization on public ambient displays [31].

Researchers have also focused on specific software architecture designs to implement new interactions in smart spaces [32-34]. For instance, Gesture Profile for Web Services (GPWS) represented the first software architecture on the web, based on events, which delivered gesture recognition services to developers of smart environments [29]. A follow-up recent work introduced Gesture Services for Cyber-Physical Environments (GS-CPE) that extended GPWS with personalized gesture recognition and user identification.

III. SOFTWARE ARCHITECTURE DESIGN FOR SPATIALLY-INDEXED MULTIMEDIA

A. Spatially-indexed multimedia

In this article, we work with the concept of *spatially-indexed multimedia*, which we define as the layer connecting digital and physical realities. From the user's perspective, spatially-indexed media denotes digital content that has been associated *a priori* with a specific region of the physical space, *e.g.*, in the middle of the room, next to the TV set, or above the table. A digital space indexed in the physical world is defined in this work as a 3-D geometry of the physical space with 6 DOF localization properties.

Our configuration for a mixed-reality physical-digital space implementing spatially-indexed multimedia is shown in Figure 2: a motion capture system is used to locate and track objects in the physical space, creating thus a digital-physical indexed space. Regions of this space connecting digital content with physical locations are illustrated with cuboids in Figure 2. Smart devices and a large display instantiate both *personal* and *public* visualization surfaces for media indexed in this space. Users access content via smart mobile devices, *e.g.*, smartphones, tablets, etc.; see Figures 3b and 3c.

To implement spatially-indexed media, we employ the concept of *content maps*, which represent layers of various content types superimposed on top of the physical space. We define the following types of maps:

- (a) The *physical map* contains the coordinates, geometric shapes, and volumes of the active regions in this space.
- (b) The *user map* contains information about the dynamic localization of users, their profiles, and interaction history in this space.
- (c) The *device map* contains information about the dynamic localization of devices and their properties, such as display size, pixel resolution, and input capabilities.
- (d) The *multimedia map* represents an association between digital content, the space map, and the device map.
- (e) The *activity map* is a dynamic snapshot of users' interactions in the physical-digital space.

B. Software architecture for spatially-indexed multimedia

To implement the concept of spatially-indexed media, we designed and developed a multi-level software infrastructure by following several design guidelines:

(a) *Identification services*. The software infrastructure automatically assigns a new ID for each device and user that enters physical-digital space.



Figure 2. Spatially-indexed multimedia implemented with a motion capture system. Physical regions of the space (cuboids in the figure), smart devices and large displays, and users represent active components of this space.



Figure 3. A user consuming digital content linked to a physical region (cuboid) of the smart space. Cuboids are visualized on a large public display (a) and on personal devices (b, c).

- (b) *Localization services*. The software infrastructure reports in real time the location and orientation for all devices and users in the physical-digital space.
- (c) *Portability*. Heterogeneous devices can interact with each other using simple message-based protocols.
- (d) *Scalability*. New devices and users can be registered into the system in real-time;
- (e) *Simple integration and simple access to data.* Access to data provided by the software infrastructure should be straightforward to use and integrate in new applications designed for the physical-digital space.

We designed the software architecture on four distinct levels (see Figure 4), as follows:

(a) The sensor layer is responsible with location and motion data acquisition in the physical space. We implemented this layer using a Vicon[™] Motion Capture System composed of six Bonita[™] infrared cameras working at 100 fps interconnected through a 100 Mbps Ethernet local network. Each data frame contains the positions of infrared markers placed in the scene, which can be detected with an accuracy of 0.5 mm. Patterns of multiple markers are automatically recognized and tracked, making possible identification and localization of multiple devices and users. The administrator of the space interacts with this layer through the Vicon Nexus Interface (dataflow no. 8) in order to define marker patterns. The information regarding these patterns is accessed by the next layer through the Vicon Datastream[™] SDK (dataflow no. 1).



Figure 4. Multi-layer software architecture for the physical-digital space implementing spatially-indexed multimedia.

(b) The *data processing layer* is responsible for the following tasks: ensuring the persistence of physical entities (data is transmitted to the next layer even if the markers disappear shortly from the view of the system), computing and converting the position and orientation, and registration of devices and users. Data is then converted to an open format (JSON) by the Data Access Point. Devices and users maps are loaded at this point with localization information. The configuration

interface allows the space administrator to interact with this layer in order to visualize and debug the dataflow, to register entities and to load additional information.

- (c) The web & cloud layer consists of a web server, a JavaScript engine and several web pages and web services. We chose JavaScript to implement the logic and interface of our software infrastructure because of its portability on virtually every device with a web browser. The Media Consumption Engine is a web page generator that allows users to access and interact with multimedia. Other third party applications can also access data through the Web Services Module. The regions of the indexed space (contained in the space map) are defined by the space administrator through the Space Configuration Module (dataflow no. 5 and 6 in Figure 4). The administrator can also associate multimedia files with specific users, devices and space regions (through the Media Map).
- (d) The *clients layer* contains applications that access the information delivered by the previous layers of our software infrastructure (Space Map, User Map, Device Map, Media Map, and Activity Map) through web protocols. For example, the following two steps are needed for a complete integration of a device that entered the physical-digital space (Figure 5): *registration* (by attaching a marker pattern to the smartphone) and *content association* (by requesting a web page from the smartphone's browser). From this moment, the device is able to interact with content in the physical-digital space. Due to the universal portability of web pages, new devices can be integrated in our scenarios instantly.



Figure 5. Markers attached to the smart device (a) are tracked in real time by the ViconTM system (b) making the smart device visible to our system implementing the physical-digital indexed space (c). When the device enters a cuboid, content is accessible to that device, such as a video file (a).

IV. EXPERIMENTATION AND SIMULATION

We measured the technical performance of our software infrastructure in several scenarios, involving multiple subjects in the physical-digital space. Each subject is represented by a 5-marker pattern. We evaluated dataflows (1), (2), and (3) for our architecture (Figure 4). For each benchmark evaluation, we ran 100 repetitions for our tests. Since the video cameras work at 100 fps, we also examined whether our software infrastructure was able to match this time resolution as well. We ran our evaluation benchmarks on a Intel® CoreTM i7-4790 @3.60GHz 64-bit machine with 8GB RAM running Windows 7.

The required time for a data frame to travel from the Sensor layer to the Data Processing layer depends on the number of subjects that occupy the physical-digital space. Each subject is identified by one or multiple markers, which are tracked by the Vicon system. We found that the average time follows an approximately linear growth $(R^2 = 0.979, v = 0.013x + 0.011)$ with the number of subjects; see Figure 6. Although the maximum time (i.e., the worst-case scenario for our software infrastructure) growth $(R^2 = 0.978)$. presents а quadratic $y = 0.0088x^2 - 0.0261x + 0.0726$), it is nevertheless under 2.5% of the 10 ms limit (i.e., the critical time for a frame to be fully processed) and stays under 0.25 ms for the maximum times measured during our evaluation. A Friedman ANOVA test revealed a significant effect of the scene complexity on average execution times $(\chi^2_{(5,N=100)} = 460.726, p < .001)$ and follow-up Wilcoxon signed-rank tests showed significant differences (Bonferroni corrected at p = .01/5 = .002) for all pairs of consecutive experimental conditions (all p < .001) with medium to large Cohen effect sizes (r from .373 to .590).



Figure 6. Relationship between scene complexity and execution time needed by the Processing layer to collect subjects' locations.

A similar relation between execution time and scene complexity is obtained for dataflow 2, from the Data Processing layer to the Web and Cloud layer. We found that the average time follows an approximately linear growth ($R^2 = 0.9958$, y = 0.0519x + 0.285) with the number of subjects; see Figure 6. The maximum time (i.e., the worst-case scenario for our software infrastructure) also presents a

linear growth ($R^2 = 0.9568$, y = 0.0814x + 0.285). It is nevertheless under 10% of the 10 ms limit which represents the critical time for a frame to be processed.

A Friedman ANOVA test revealed a significant effect of the scene complexity on average execution times ($\chi^2_{(5,N=100)} = 444.350, p < .001$) and follow-up Wilcoxon signed-rank tests showed significant differences (Bonferroni corrected at p = .01/5 = .002) for all pairs of consecutive experimental conditions (all p < .001) with medium to large Cohen effect sizes (r from .437 to .569).



Figure 7. Relationship between scene complexity and execution time needed by the Processing layer to generate a JSON description of the scene.

To study the performance of the software infrastructure for dataflow 3 from the Web and Cloud layer to the Clients layer, we measured the average time a client application needs to wait until it receives data. We ran several tests for various number of multimedia consumers (i.e., devices asking data). Results seem to indicate no relationship between execution time and scene complexity or the number of multimedia consumers. A possible explanation might be that the performance of dataflow number 3 is given majoritarily by the performance of the network communications. The random variation of network parameters (e.g., throughput and latency) seems to be much more important than any variation caused by the scene complexity or by the number of client devices making data requests. The average execution time corresponding to workflow 3 is below the 10 ms limit.

V. CONCLUSION

We presented in this work a new software architecture design for implementing consumption of spatially-indexed media in smart environments. We evaluated the technical performance of an implementation of our software architecture design for a smart space indexed by a motion capture system, in which users access and visualize spatially-indexed media on their smart mobile devices or on a large ambient display. Future work will focus on designing interaction techniques for this space as well as on updating our architecture design to accommodate for a wide range of mobile interactive devices and wearable sensors, such as smart watches, rings, and gesture-sensing gadgets.

REFERENCES

- Z. Huang, W. Li, and P. Hui, "Ubii: Towards seamless interaction between digital and physical worlds," in Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 341–350. doi:10.1145/779359.779362
- [2] M. Goble, "Managing the gap between the physical and digital world through a balance between transparent and performative interaction," Thesis project, Malmö Högskola University, 2010.
- [3] M. Rohs, J. Schöning, M. Raubal, G. Essl, and A. Krüger, "Map navigation with mobile devices: virtual versus physical movement with and without visual context," in Proceedings of the 9th international conference on Multimodal interfaces, 2007, pp. 146– 153. doi: 10.1145/1322192.1322219
- [4] H. Ishii and B. Ullmer, "Tangible bits: towards seamless interfaces between people, bits and atoms," in Proceedings of the ACM SIGCHI Conference on Human factors in computing systems, 1997, pp. 234– 241. doi: 10.1145/258549.258715
- [5] H. Ishii, "Tangible bits: beyond pixels," in Proceedings of the 2nd international conference on Tangible and embedded interaction, 2008, pp. xv-xxv. doi: 10.1145/1347390.1347392
- [6] H. Ishii, D. Lakatos, L. Bonanni, and J.-B. Labrune, "Radical atoms: beyond tangible bits, toward transformable materials," Interactions, vol. 19, no. 1, pp. 38–51, 2012. doi: 10.1145/2065327.2065337
- [7] A. Elliott, B. Peiris, and C. Parnin, "Virtual reality in software engineering: Affordances, applications, and challenges," in Software Engineering (ICSE), 2015 IEEE/ACM 37th IEEE International Conference on, 2015, vol. 2, pp. 547–550. doi: 10.1109/ICSE.2015.191
- [8] M. Billinghurst, A. Clark, G. Lee, and others, "A survey of augmented reality," Foundations and Trends® Human–Computer Interaction, vol. 8, no. 2–3, pp. 73–272, 2015. doi: 10.1561/1100000049
- [9] M. Weiser, R. Gold, and J. S. Brown, "The origins of ubiquitous computing research at PARC in the late 1980s," IBM systems journal, vol. 38, no. 4, pp. 693–696, 1999. doi: 10.1147/sj.384.0693
- [10] D. J. Cook, "Prediction algorithms for smart environments," Smart environments: Technologies, protocols, and applications, pp. 175– 192, 2005.
- [11] S. K. Das, "Designing Smart Environments: Challenges, Solutions and Future Directions," in Proceedings of ruSMART conference, St. Petersburg, Russia, 2008. doi: 10.1002/047168659X
- [12] D. G. Korzun, S. I. Balandin, and A. V. Gurtov, "Deployment of Smart Spaces in Internet of Things: Overview of the design challenges," in Internet of Things, Smart Spaces, and Next Generation Networking, Springer, 2013, pp. 48–59. doi: 10.1007/978-3-642-40316-3_5
- [13] M. Gardner and J. Elliott, "The Immersive Education Laboratory: understanding affordances, structuring experiences, and creating constructivist, collaborative processes, in mixed-reality smart environments," EAI Endorsed Transactions on Future Intelligent Educational Environments, vol. 14, no. 1, p. e6, 2014. doi: 10.4108/fiee.1.1.e6
- [14] P. Carrington, A. Hurst, and S. K. Kane, "Wearables and chairables: inclusive design of mobile input and output techniques for power wheelchair users," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014, pp. 3103–3112. doi: 10.1145/2556288.2557237
- [15] S. K. Kane, B. Frey, and J. O. Wobbrock, "Access lens: a gesturebased screen reader for real-world documents," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013, pp. 347–350. doi: 10.1145/2470654.2470704
- [16] F. Garzotto, M. Gelsomini, A. Pappalardo, C. Sanna, E. Stella, and M. Zanella, "Monitoring and Adaptation in Smart Spaces for Disabled Children," in Proceedings of the International Working Conference on Advanced Visual Interfaces, 2016, pp. 224–227. doi: 10.1145/2909132.2909283
- [17] G. Agosta et al., "Playful Supervised Smart Spaces (P3S)–A Framework for Designing, Implementing and Deploying Multisensory Play Experiences for Children with Special Needs," in Digital System Design (DSD), 2015 Euromicro Conference on, 2015, pp. 158–164. doi: 10.1109/DSD.2015.61

- [18] O.-A. Schipor, S.-G. Pentiuc, and M.-D. Schipor, "Toward Automatic Recognition of Children's Affective State Using Physiological Parameters and Fuzzy Model of Emotions," Advances in Electrical and Computer Engineering, vol. 12, no. 2, pp. 47–50, 2012. doi:10.4316/AECE.2012.02008
- [19] L. Liu, E. Stroulia, I. Nikolaidis, A. Miguel-Cruz, and A. R. Rincon, "Smart homes and home health monitoring technologies for older adults: A systematic review," International journal of medical informatics, vol. 91, pp. 44–59, 2016. doi: 10.1016/j.ijmedinf.2016.04.007
- [20] O. Geman et al., "Challenges and trends in Ambient Assisted Living and intelligent tools for disabled and elderly people," in Computational Intelligence for Multimedia Understanding (IWCIM), 2015 International Workshop on, 2015, pp. 1–5. doi: 10.1109/IWCIM.2015.7347088
- [21] E. Gilman, O. Davidyuk, X. Su, and J. Riekki, "Towards interactive smart spaces," Journal of Ambient Intelligence and Smart Environments, vol. 5, no. 1, pp. 5–22, 2013. doi: 10.3233/AIS-120189
- [22] D. Korzun, I. Galov, A. Kashevnik, and S. Balandin, "Virtual shared workspace for smart spaces and M3-based case study," in Open Innovations Association FRUCT, Proceedings of 15th Conference of, 2014, pp. 60–68. doi: 10.1109/FRUCT.2014.6872437
- [23] A. Amato, B. Di Martino, and S. Venticinque, "Semantic brokering of multimedia contents for smart delivery of ubiquitous services in pervasive enviroments," IJIMAI, vol. 1, no. 7, pp. 16–25, 2012. doi: 10.9781/ijimai.2012.172
- [24] J. W. Lee, S. Cho, S. Liu, K. Cho, and S. Helal, "Persim 3D: contextdriven simulation and modeling of human activities in smart spaces," IEEE Transactions on Automation Science and Engineering, vol. 12, no. 4, pp. 1243–1256, 2015. doi: 10.1109/TASE.2015.2467353
- [25] R.-D. Vatavu, "Point & click mediated interactions for large home entertainment displays," Multimedia Tools and Applications, vol. 59, no. 1, pp. 113–128, 2012. doi: 10.1007/s11042-010-0698-5
- [26] B. Kollee, S. Kratz, and A. Dunnigan, "Exploring gestural interaction in smart spaces using head mounted devices with ego-centric sensing," in Proceedings of the 2nd ACM symposium on Spatial user interaction, 2014, pp. 40–49. doi: 10.1145/2659766.2659781
- [27] A. Matassa and F. Cena, "Body experience in the ubiquitous era: towards a new gestural corpus for smart spaces," in Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, 2015, pp. 945–950. doi: 10.1145/2800835.2806205
- [28] R.-D. Vatavu, "A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments," Journal of Ambient Intelligence and Smart Environments, vol. 5, no. 2, pp. 187– 211, 2013. doi: 10.3233/AIS-130200
- [29] R.-D. Vatavu, C.-M. Chera, and W.-T. Tsai, "Gesture profile for web services: an event-driven architecture to support gestural interfaces for smart environments," in International Joint Conference on Ambient Intelligence, 2012, pp. 161–176. doi: 10.1007/978-3-642-34898-3_11
- [30] Y. Lou, W. Wu, R.-D. Vatavu, and W.-T. Tsai, "Personalized gesture interactions for cyber-physical smart-home environments," Science China Information Sciences, vol. 60, no. 7, p. 072104, 2017. doi: 10.1007/s11432-015-1014-7
- [31] R.-D. Vatavu, "Smart-Pockets: Body-deictic gestures for fast access to personal data during ambient interactions," International Journal of Human-Computer Studies, vol. 103, pp. 1–21, 2017. doi:10.1016/j.ijhcs.2017.01.005
- [32] G. Van Seghbroeck, S. Verstichel, F. De Turck, and B. Dhoedt, "WS-Gesture, a gesture-based state-aware control framework," in Service-Oriented Computing and Applications (SOCA), 2010 IEEE International Conference on, 2010, pp. 1–8. doi: 10.1109/SOCA.2010.5707162
- [33] A. Mingkhwan, P. Fergus, O. Abuelma'atti, M. Merabti, B. Askwith, and M. B. Hanneghan, "Dynamic service composition in home appliance networks," Multimedia Tools and Applications, vol. 29, no. 3, pp. 257–284, 2006. doi:10.1007/s11042-006-0018-2
- [34] K.-I. BENŢA and M.-F. Vaida, "Towards real-life facial expression recognition systems," Advances in Electrical and Computer Engineering, vol. 15, no. 2, pp. 93–102, 2015. doi:10.4316/AECE.2015.02012