Personalised Human Device Interaction through Context aware Augmented Reality

Madhawa Perera
madhawa.perera@anu.edu.au
Research School of Computer Science, Australian National University
Canberra, ACT, Australia

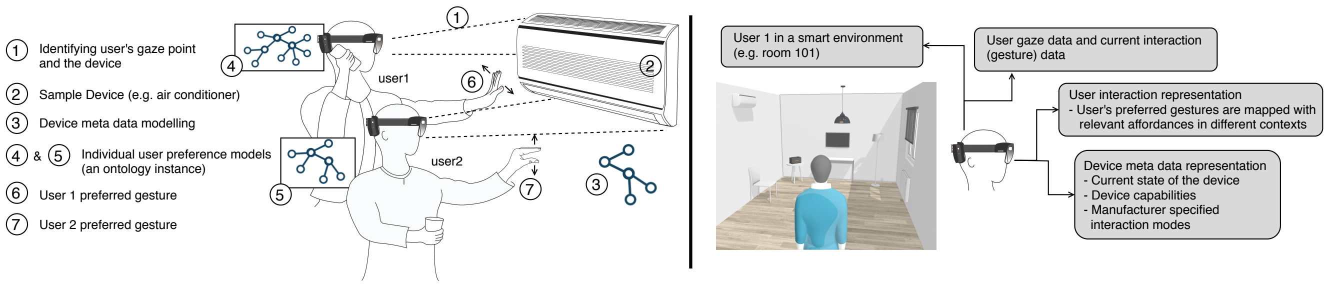

Figure 1: Left: Overview of the suggested personalised human device interaction framework. Right: AR application that consider user's preferred gestures and act as semantic middleware between device and the user.

ABSTRACT

Human-device interactions in smart environments are shifting prominently towards naturalistic user interactions such as gaze and gesture. However, ambiguities arise when users have to switch interactions as contexts change. This could confuse users who are accustomed to a set of conventional controls leading to system inefficiencies. My research explores how to reduce interaction ambiguity by semantically modelling user specific interactions with context, enabling personalised interactions through AR. Sensory data captured from an AR device is utilised to interpret user interactions and context which is then modeled in an extendable knowledge graph along with user's interaction preference using semantic web standards. These representations are utilized to bring semantics to AR applications about user's intent to interact with a particular device affordance. Therefore, this research aims to bring semantical modeling of personalised gesture interactions in AR/VR applications for smart/immersive environments.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Gestural input**; **User centered design**; • **Information systems → Semantic web description languages**; **Personalization**.

ICMI '20, October 25–29, 2020, Virtual event, Netherlands

ACM ISBN 978-1-4503-7581-8/20/10…$15.00
https://doi.org/10.1145/3382507.3421157

KEYWORDS

Gesture Interaction, Augmented Reality, Semantic Web, Gaze Detection, User Interface, Adaptive Systems

ACM Reference Format:
Madhawa Perera. 2020. Personalised Human Device Interaction through Context aware Augmented Reality. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20), October 25–29, 2020, Virtual event, Netherlands.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3382507.3421157

1 INTRODUCTION

Naturalistic user interfaces such as gaze and gesture, figure prominently in smart environments such as smart homes, automobiles, digital farming, smart hotel rooms, and future cities [13, 23]. Especially interactions in Augmented Reality (AR)/Virtual Reality (VR) environments and applications, gestural interactions have become prevalent due to their adaptability in such settings and accessibility to everyone. Most of such interactions consist of physical movements of the face, limbs, or body [32] and allow users to express their intentions to engage by sending out a corresponding signal [14] to a device or a system. However, most of the gesture vocabularies are designed based on a manufacturer's, designer's, producer's or vendor's decisions resulting in a greater variation in ways to interact with devices, even for similar referents (or similar affordances of a device). Users' preferred gestures to interact with device affordances also vary due to factors such as legacy bias [15], gender, ethnicity, culture, personal preference, ability/disability and context [28, 31]. As such, this could confuse the users accustomed to a set of conventional controls and users often interact with smart devices which they feel they are not familiar with. This inevitably leads to poor user experience (UX). Having to interact with unfamiliar smart devices or those forgotten how to interact with, users are compelled to refer to alternate information sources such as user manuals, tutorials or ask an expert which causes inefficiencies in system usage. Thus, it is vital to identify noninvasive interaction between humans and smart devices, without causing ambiguities, thereby preserving user preferences in order to provide a better UX in proliferating smart infrastructure/ecosystems.

2 MOTIVATION AND PROBLEM STATEMENT

Our objective is to reduce the burden on users having to adjust to different system-defined gestures in order to preserve self-efficacy of system usage. This is aimed to achieve utilizing context-aware AR rather than system-defined gesture interfaces that may not be intuitive or are contrary to user expectations and their prior experience.

Multiple research and gesture elicitation studies attempt to define the most suitable gesture vocabularies for a selected set of referents. Nevertheless, the problem is, fixed gesture sets suppress some users' preferences as most of the elicitation studies consider 'gesture agreement rate' [25, 27, 30] or the gesture with the highest frequency of usage as the suitable winning gesture. In this research we explore the implications of enabling interactions to remain rooted in an individual user's preferences, while adapting seamlessly, even as the context changes. Context changes incorporate aspects such as user indoor/outdoor location, date and time, device type, device manufacturer, and user activities. We investigate how users can maintain a personalised gesture vocabulary when interacting with the same affordances. This poses challenges; 1) How to capture and model a user's preferred interactions with an affordance and dynamically map them in changing contexts? 2) How to maintain interoperability between interaction modalities? i.e. how to use consistent gesture interactions (despite what a device offers) and allow users who are accustomed to a set of personalised gestures to use them across different devices.

Presently, AR technology has shown remarkable progress with its availability in consumer-level hardware with an increased usage. [1, 5, 10]. AR's potential to blend real and virtual world, as a powerful visualization medium [20] that can assist people real-time during gesture interactions [19], has opened up new opportunities for building interactive and engaging applications in multiple application domains [21]. On top of that, when using an AR Head Mount Display (HMD) device, we get access to data such as user's hand movements, gesture recognition, user's spatial information, gaze points, which provide a promising direction of using AR HMD as a semantic middleware to address the stated challenges. Yet, capturing user interactions (gaze and gesture) and semantic modelling of captured gaze and gesture interactions and user context remain underexplored in AR. This is a necessity to bring interoperability between interaction modalities and to develop user specific interactions for similar referents (affordances) in changing contexts. It warrants answers to additional problems: how can we capture and model human gaze and gesture interaction in AR applications to provide personalised interactions to user?

Example use cases

1) Consider a user who uses a certain gesture to turn on their personal air conditioner. If this user visits a hotel room with an air conditioner of a different model that carries different interactions, how can the system accommodate the user's preferred gesture and let the user continue, as in their own room, without having to read instructions. In the domain of tourism and leisure where research and surveys are conducted to leverage the discomfort of being away from home [11, 16], personalised interaction offers a better UX. 2) Users switching vehicles or those who rent different cars need to interact with the same affordance they had in the previous model. The new model may use new/different gesture interactions, e.g. BMW's iDrive infotainment system expects users to point a finger to 'accept a call' whereas Mercedes-Benz' User Experience (MBUX) multimedia infotainment system uses the same gesture to select an icon on their touch screen. A user, however, may prefer a different gesture to "accept a call". Generally, functionality of a vehicle's interactive features has been designed based on proprietary frameworks that may not fit all. Can personalizing interaction could address this problem?

3 RESEARCH QUESTIONS

Driven from the motivation and problems stated above, this PhD project aims to answer the following research questions;

(1) **How to model relationships between user's preferred gestures, relevant user context and device affordances semantically, in an extensible way that can be processed by AR applications, through an automated reasoner?** This question explores ways to model captured user gesture interactions and the relationships that we identify through user studies, in a machine understandable format.
(2) **Can we use the model in question 1 for automatic composition of personalised AR scenes/interactions in selected contexts?** When there is a user preferred gesture, AR interfaces should be able to understand it in relation to a given context. Thus, here we investigate how to incorporate semantic models into an AR application development.
(3) **Can we use context aware AR to bring interoperability for gesture interactions?** This explores how to use AR as a medium that maps existing manufacture designed gestures with user-preferred gestures to provide personalised Human Device Interactions (HDI).

4 RELATED WORK

Here we review the prior work which use AR and semantic web technologies to interact with smart devices, model gestures and context.

Using AR to interact with devices

Zachariah et al. [33] propose a model which help users to interact with devices using Mobile AR (MAR). They provide developers a convenient platform to display custom interfaces for their device, yet not available for users. Jo et al. [8] presented ARIoT, an AR framework that has the capability to dynamically identify the target objects in the working space along with important features. Thereby interactive content is retrieved and communicated to the user's AR system. Even though their metadata modelling does not include user preferences, which is the major focus in my research, their approach signals towards the feasibility of using AR to interact with smart devices. Jo et al. [8] further highlight the importance of having standard data formats in the future for the purpose of adding smart object specific semantics to identify features that are required by an AR device with additional content and information about the device itself, such as data about control interface which will tell how to interact with the device. Additionally, Leepanen et al. [12] present an adaptable mobile agent composition that contains the data representation logic and mappings between AR applications and system resources. It is proved that AR can enhance access in smart home systems for physically disabled, autistic or elderly people when interacting with devices in Ullah et al. [26] and Hervás, et al. [7] studies. Sun et al.'s [24] use AR HMD as a visualization and interaction tool for users to control Internet of Things (IoT) devices with hand gestures. In their research they have implemented a standard 2D convolution neural network (CNN) for real-time hand-gesture recognition to interpret user intent. They do not use semantical mapping of device capabilities to user preferences, instead they use pre-defined gesture vocabulary. Above studies evince using AR to interact with devices is a feasible approach when creating personalised device interactions.

Blending Semantic Web and AR

There is a growing interest in using semantic web in AR application development. Rumiński et al.'s [21] suggest that semantic web techniques can be an efficient solution to search contextually described resources constituting interactive AR presentations. Further, Walczak et al. [29] propose a Contextual AR Environments (CARE) modelling approach which is utilised in Rumiński et al. [21] semantic model for distributed AR services. Chen L. et al.'s [2] show that embedding semantic understanding with Mixed Reality (MR) interfaces can greatly enhance the UX by helping to understand object-specific behaviours. Chen L. et al. [2] demonstrate a framework for a material-aware prototype system for generating context-aware physical interactions between the real and the virtual objects. However, the focus of their work is on material understanding and its semantic fusion with the virtual scene in a MR environment. Further, looking at context-awareness, Zhu J. et al. [34] have proposed a framework specifically designed for an assisted maintenance system in which they have incorporated context-aware AR with semantic web technologies to provide information that is more useful to the user [34] and presents a context-aware AR authoring tool. These studies along with Ruminski et al.'s [21] performance evaluation when using ontologies with AR applications evince the feasibility of using semantic web technology with AR to model user context.

Modeling Gesture Semantics

Next we review, mapping and modelling gestures. An approach adopted by researchers is to define taxonomies, enabling designers and manufacturers to use the standard definitions when defining gesture vocabularies. Following this path, Scoditti et al. [22] proposed a gestural interaction taxonomy in order to guide designers and researchers. Similarly, Choi et al. [3] developed a 3D hand gesture taxonomy and notation method. The results of this study can be used as a guideline to organize hand gestures for enhancing the usability of gesture-based interfaces. Moving beyond taxonomies, Osumer et al. [17] have modelled a gesture ontology based on a Microsoft Kinect-based skeleton which aims to describe mid-air gestures of human body. Khairunizam et al. [9] have conducted a study with a similar purpose. Flotyński et al. [4] research provides a promising direction for representing interactions in VR environments using semantic web technologies which we aim to follow closely. Overall, the attempts above establish that ontologies are a better approach for modelling user preferred gestures and features, attributes and relationships attached with it.

5 RESEARCH HYPOTHESES

Research hypotheses, formed in this context, are: H1) Users' preferred gestures (captured through AR wearables) can be modelled semantically with their relationship to the user context, device context and device capabilities (affordances that device provides) for the purpose of personalizing user interactions; H2) Compared to non-adaptive systems, where users have to adapt to a given set of gestures, personalised human device gesture interactions can provide a better or improved user experience; H3) AR can supplement to personalised gestural interactions by providing ability to access information and communicate that information with a user appropriately at any time in any location (with real time feedback).

6 RESEARCH PLAN AND METHODOLOGY

This research is planned to conduct in four main stages within three and a half years (42 months).

Stage 1: Literature Review and Feasibility Analysis

The first twelve to fifteen months are dedicated to 1) Conducting a comprehensive literature review covering the use of semantic web technology in AR and context modelling. 2) A qualitative user study to understand users gesture interactions, types of user preferences and relevant context which are necessary to model. 3) Build an ontology to represent relationships between user preferred interaction (gesture information), user context, and device affordances from the above user study.

Stage 2: Data Collection and Analysis

The second phase of the research aims to capture user interactions and to classify user contexts. To supplement this, we will incorporate other sensor data such as an accelerometer, a gyroscope, and a magnetometer in an AR HMD device. 1) A user experiment will be designed to capture sensor data, then will decode the data to different contexts through classification algorithms. It is desirable to apply classifiers that need little to no training in advance and that are able to perform the classification as fast as possible. 2) Mapping data using ontology schema that models the relationship between identified contexts, user interaction preferences and affordances after the experiment. This stage will be run through the sixteenth to twenty sixth month of the PhD.

Stage 3: Design and Implementation

Next step is designing the framework to build dynamic AR interfaces by incorporating modelled data from stage 2. We will implement a prototype AR application using the intended framework and test/evaluate in this stage. This will be run through the twenty seventh month to thirty second month of the PhD.

Stage 4: Evaluation and Optimization

Finally, a usability study will be conducted with the users who participated in the initial user study to evaluate whether the system meets their expectations as described in the initial study. Additionally, we will run another user study with a completely new set of users to evaluate UX with AR assisted personalised gestures when interacting with smart devices. Details will be defined further as the designs and implementations are finalized. This is the final stage and will be run through the thirty third month to forty second month of the PhD.

7 COMPLETED AND PLANNED RESEARCH

As the initial step, a comprehensive literature review was conducted to understand the state of the art of the identified problem. This gave a detailed overview of the latest work, the existing gap and the technologies and suitable directions to develop personalised and context aware gestures for users to interact with smart devices. We identified three major points that need to be addressed/modelled when personalizing interactions. 1) Perception of the situation; is understanding or being aware of the device and user context. 2) Ubiquitous access; is the availability of information and ability to communicate that information with the user appropriately at any time in any location. 3) Natural interaction; is to enable users to communicate with a device non-invasively to fulfil a certain intent. AR blended with ontologies provides a promising direction to map the relationship between each of those three points and conduct automated reasoning to identify personalised interactions. Further, literature show possible ontologies such as Semantic Sensor Network (SSN) [6] to fulfil part of conceptual requirements (specially to represent device meta data). Yet, user preferences and hand gestures modelling need to be defined. A roadmap of preliminary work is published in [18]. We designed the first version of the ontology which describes a gesture semantically with their relationships to device affordances. Ontology schema can be found via www.w3id.org/hdgi. Currently, brainstorming is in progress for further revision of the ontology schema. The next version of the ontology schema will be released after conducting the first user study, which has been designed and received ethical clearance. Currently we are looking at analysing hand tracking, eye tracking and other sensor data of the selected AR HMD.

8 EXPECTED CONTRIBUTION

This research contributes to the field of human centered computing and AR, especially with the techniques to capture and model human gesture interactions semantically, along with a user context. As a methodological contribution, we intend to develop a framework which uses AR as a medium to capture and store user specific interactions (gestures and gaze) with relevant referents, aiming to offer personalised interaction experiences to users. Further, the research contributes in developing several artefacts such as functional prototypes (for context modelling using AR devices, for modelling personalised gesture interaction). As an empirical and philosophical contribution, we propose to design an ontology to map users' preferred interactions with dynamic context and device affordances which can be used for personalised interaction modelling. This would help in many industries such as automobile, leisure (hotel), gaming etc. to enhance their customer experience as they attempt to provide more personalised services to their clients. In addition, for users with disabilities, it extends accessibility to modern smart ecosystems by personalizing interactions. Further, all the artefacts including data sets of the research are and will be openly available for further development enabling other researchers to enhance the framework and conduct their own human interaction studies with AR HMDs.

ACKNOWLEDGMENTS

I would like to thank my supervisors Dr. Armin Haller and Dr. Matt Adcock for their continuous guidance and the helpful supervision of my work.

REFERENCES

[1] Hakan Altinpulluk. 2019. Determining the trends of using augmented reality in education between 2006-2016. *Education and Information Technologies* 24, 2 (2019), 1089–1114.
[2] Long Chen, Wen Tang, Nigel John, Tao Ruan Wan, and Jian Jun Zhang. 2018. Context-aware mixed reality: A framework for ubiquitous interaction. *arXiv preprint arXiv:1803.05541* (2018).
[3] Eunjung Choi, Heejin Kim, and Min K Chung. 2014. A taxonomy and notation method for three-dimensional hand gestures. *International Journal of Industrial Ergonomics* 44, 1 (2014), 171–188.
[4] Jakub Flotyński, Adrian Nowak, and Krzysztof Walczak. 2018. Explorable representation of interaction in VR/AR environments. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 589–609.
[5] Juan Garzón, Juan Pavón, and Silvia Baldiris. 2019. Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality* 23, 4 (2019), 447–459.
[6] Armin Haller, Krzysztof Janowicz, Simon J. D. Cox, Maxime Lefrançois, Kerry Taylor, Danh Le Phuoc, Joshua Lieberman, Raúl García-Castro, Rob Atkinson, and Claus Stadler. 2019. The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. *Semantic Web* 10, 1 (2019), 9–32. https://doi.org/10.3233/SW-180320.
[7] Ramón Hervás, Alberto Garcia-Lillo, and José Bravo. 2011. Mobile augmented reality based on the semantic web applied to ambient assisted living. In *International workshop on ambient assisted living*. Springer, 17–24.
[8] Dongsik Jo and Gerard Jounghyun Kim. 2016. In-situ AR manuals for IoT appliances. In *2016 IEEE international conference on consumer electronics (ICCE)*. IEEE, 409–410.
[9] Wan Khairunizam, K Ikram, SA Bakar, ZM Razlan, and I Zunaidi. 2018. Ontological Framework of Arm Gesture Information for the Human Upper Body. In *Intelligent Manufacturing & Mechatronics*. Springer, 507–515.
[10] Inta Kotane, Daina Znotina, and Serhii Hushko. 2019. Assessment of trends in the application of digital marketing. *Scientific Journal of Polonia University* 33, 2 (2019), 28–35.
[11] Asterios Leonidis, Maria Korozi, George Margetis, Dimitris Grammenos, and Constantine Stephanidis. 2013. An intelligent hotel room. In *International Joint Conference on Ambient Intelligence*. Springer, 241–246.
[12] Teemu Leppänen, Arto Heikkinen, Antti Karhu, Erkki Harjula, Jukka Riekki, and Timo Koskela. 2014. Augmented reality web applications with mobile agents in the internet of things. In *2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies*. IEEE, 54–59.
[13] James Manyika. 2015. *The Internet of Things: Mapping the value beyond the hype*. McKinsey Global Institute.
[14] Sushmita Mitra and Tinku Acharya. 2007. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37, 3 (2007), 311–324.
[15] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, MC Schraefel, and Jacob O Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *interactions* 21, 3 (2014), 40–45.
[16] Oracle. 2017. Hotel 2025 emerging technologies destined to reshape our business. https://www.oracle.com/webfolder/s/delivery_production/docs/FY16h1/doc31/Hotels-2025-v5a.pdf
[17] Mehdi Ousmer, Jean Vanderdonckt, and Sabin Buraga. 2019. An ontology for reasoning on body-based gestures. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. 1–6.

[18] Madhawa Perera, Armin Haller, and Matt Adcock. 2019. A Roadmap for Semantically-Enabled Human Device Interactions.. In *SAWSemStats@ ISWC*.

[19] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction*. Springer, 282–299.

[20] Dariusz Rumiński and Krzysztof Walczak. 2017. Semantic model for distributed augmented reality services. In *Proceedings of the 22nd International Conference on 3D Web Technology*. 1–9.

[21] Dariusz Rumiński and Krzysztof Walczak. 2020. Large-scale distributed semantic augmented reality services–A performance evaluation. *Graphical Models* 107 (2020), 101027.

[22] Adriano Scoditti, Renaud Blanch, and Joëlle Coutaz. 2011. A novel taxonomy for gestural interaction techniques based on accelerometers. In *Proceedings of the 16th international conference on Intelligent user interfaces*. 63–72.

[23] Constantine Stephanidis, Gavriel Salvendy, Margherita Antona, Jessie YC Chen, Jianming Dong, Vincent G Duffy, Xiaowen Fang, Cali Fidopiastis, Gino Fragomeni, Limin Paul Fu, et al. 2019. Seven HCI grand challenges. *International Journal of Human–Computer Interaction* 35, 14 (2019), 1229–1269.

[24] Yongbin Sun, Alexandre Armengol-Urpi, Sai Nithin Reddy Kantareddy, Joshua Siegel, and Sanjay Sarma. 2019. Magichand: Interact with iot devices in augmented reality environment. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1738–1743.

[25] Theophanis Tsandilas. 2018. Fallacies of agreement: A critical review of consensus assessment methods for gesture elicitation. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 3 (2018), 1–49.

[26] Ahmed Mohmmad Ullah, Md Rashedul Islam, Sayeda Farzana Aktar, and SK Alamgir Hossain. 2012. Remote-touch: Augmented reality based marker tracking for smart home control. In *2012 15th International conference on computer and information technology (ICCIT)*. IEEE, 473–477.

[27] Radu-Daniel Vatavu and Jacob O Wobbrock. 2015. Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1325–1334.

[28] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob A Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies. In *Proceedings of ACM Int. Conf. on Designing Interactive Systems (DIS'20)*.

[29] Krzysztof Walczak, Dariusz Rumiński, and Jakub Flotyński. 2014. Building contextual augmented reality environments with semantics. In *2014 International Conference on Virtual Systems & Multimedia (VSMM)*. IEEE, 353–361.

[30] Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. 2005. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems*. 1869–1872.

[31] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1083–1092.

[32] LI Yang, Jin HUANG, TIAN Feng, WANG Hong-An, and DAI Guo-Zhong. 2019. Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware* 1, 1 (2019), 84–112.

[33] Thomas Zachariah and Prabal Dutta. 2019. Browsing the web of things in mobile augmented reality. In *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications*. 129–134.

[34] J Zhu, Soh-Khim Ong, and Andrew YC Nee. 2015. A context-aware augmented reality assisted maintenance system. *International Journal of Computer Integrated Manufacturing* 28, 2 (2015), 213–225.