International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.6 (2016), pp.43-60
http://dx.doi.org/10.14257/ijmue.2016.11.6.05

Immersive Networking-A Framework for Virtual Environments with Augmented Reality in Human Decision-Making

Theo Kanter, Uno Fors and Rahim Rahmani

Department of Computer and Systems Sciences
Stockholm University
Sweden
{kanter,uno,rahim}@dsv.su.se

Abstract

In this publication we present Immersive Networking as a novel framework for connecting people and places and things in virtual environments with augmented reality to be used in eg. Virtual training environments. The research is mandated by technology advances in internetworking underpinned by 5G networks and Internet-of-Things. These advances present new possibilities and challenges to integrate people, places and things in virtual environments. Existing frameworks such as MPEG-V possess representational capabilities but have insufficient support for integrating entities from the real world via heterogeneous infrastructure. MPEG-V for instance makes no statements about distributed control. Seamless experiences in virtual environment require self-organization of connectivity between people, places and things via heterogeneous 5G and Internet-of-Things infrastructures. A second important aspect of the quality of our experience is the immediacy of responses. Both aspects of seamless and self-organizing connections between entities require that we push control to the end-devices co-located with the entities themselves. These end-devices may incorporate sensing gateways and interaction devices, which include both local and non-local information from the virtual environment in the interaction. Thus delegation of control to end-devices requires means for the organizing or relations and clustering by relevance. This capability is particularly important as the projected number of devices and sensors to be connected via the Internet-of-Things is projected to be in the order of 50 billion by 2020. Immersive Networking supported by MediaSense [1] constitutes a scalable self-organizing means for connecting people and places and things in virtual environments with augmented reality. MediaSense moves control to the edge enabling immediacy in experiences based on seamless self-organization through clustering of entities in relations organized by relevance. We conclude by validating our approach in several scenarios evaluating the relevance and application in human decision-making.

Keywords: *Internet-of-Things, Distributed Immersive Networking, Serious Gaming, Human Decision-Making*

1. Introduction

In many areas, traditional training and learning methods cannot be used to prepare learners for the situations they will meet in real life. One especially complex capacity to train is human decision making under stress, time restriction and where the decisions are depending on interaction with one or more individuals that are less willing or otherwise hindered to cooperate. Examples of such areas are Political or Peace negotiation, Healthcare education, Business negotiations, Second language training, Psychological consultations, Conflict resolution, Teacher training, Social and welfare care training, Management training, Legal interrogation, Prison and probation situations, Emergency/Fire brigade crisis management, as well as for Police, Military and similar

ISSN: 1975-0080 IJMUE

applications. Such encounters are inherently complex and difficult to train in real life, and require the understanding and management of human emotions and reactions, features real-time actions, decisions and consequences, requires cognitive skills as well as background information and involves encounters with one, or groups of, humans.

These training needs are mostly applicable to higher education and/or on the professional level, but may also be important in other settings.

In many of these examples, learners cannot be efficiently trained attending lectures, reading textbooks or similar models. Moreover, due to the complexity and the effects of such decisions, teachers cannot allow learners to simply walk out in the real world and interact in emergency scenes or other critical situations to train.

Hence, numerous learners often lack possibilities to train and re-train their skills, and are facing situations in real life they are not prepared for. The result of this limited training possibility is that human decisions in complex situations often are not optimal and result in more or less bad decisions that affect many families, patients, clients and other people. To enhance the exploring the level of training possibility toward a realistic human decision we propose a feasible solution to seamlessly manage virtual spatiality and augmented reality in a technology independent manner. In this context our proposed framework for shared spatiality will make able to sharing context information in real-time for virtual interaction. In order to make enable massive and scalable sharing of context information across heterogeneous network infrastructure we go beyond Internet-of-Things to make possible information sharing in real-time to make enable timely and intelligent decisions in different user scenarios. Here, we share our vision of such a crowd-sensing technology that allows users to share context information to practice multiple social-emergency scenarios electronically and obtain immediate, objective feedback in training response towards a relational Internet-of-Things solution which we have identified as Immersive networking. Immersive networking (IN) is a new area within Distributed systems focused on future Internet challenges for connecting people, places and things, be it virtual or real and how it enables us to live and train skills as immersed in new ways in a networked society.

2. Background and Motivation

2.1. Immersive Networking Interfacing with Virtual Environments

As explained in the previous section immersive networking focusing of the future Internet challenges for connection multi-user online in virtual worlds as networked virtual environments in sense of the virtual world data representation for enriched content of context-aware computing system. Immersive networking as MPEG-V framework [2] deals only with virtual world object characteristics (metadata) which helps to define a base type of attributes and characteristics of objects which will be shared by both virtual agents and the generic virtual objects. Immersive Networking explore MPEG-V virtual worlds infrastructure with distributed controller for interfacing with virtual environments as depicted in Figure 1.

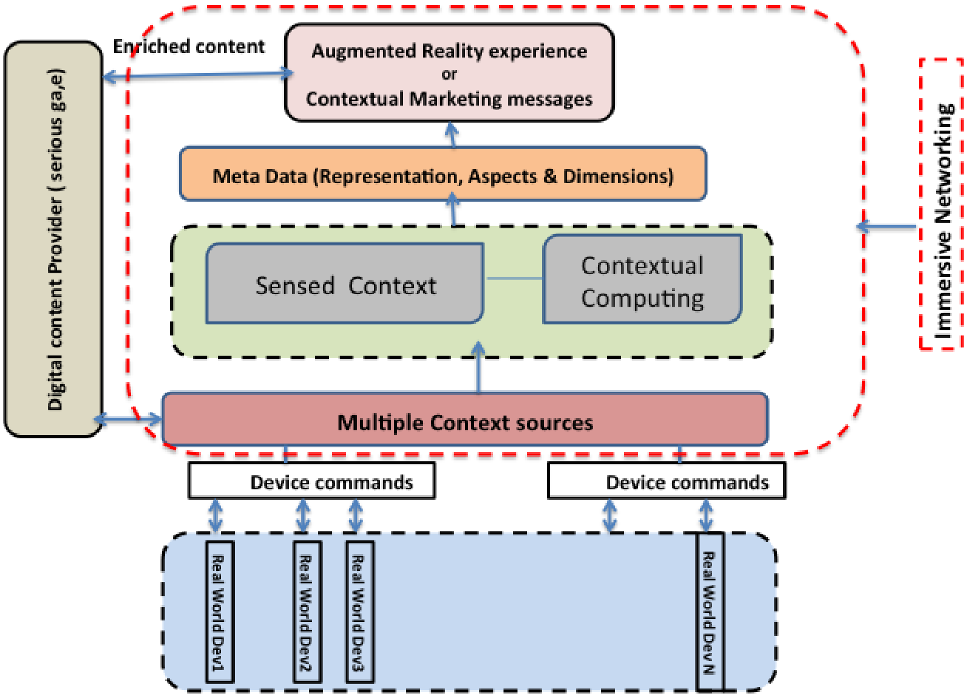

Figure 1. Immersive Networking Interfacing Virtual Environments

Immersive Networking specifies context and semantics required to provide interoperability in distributed controlling of virtual agents as well of generic virtual objects. Immersive Networking will be able to replace the adaptation engine (RV or VR) in MPEG-V [1] and make possible seamless experiences in virtual environments for self-organization of connectivity between people, places and things. Section 3 provides further description of system architecture. Next section covers serious games as digital content provider in real time or non real time, of various nature raging from an on line virtual world and simulation environment.

2.2. Serious Games as Digital Content Provider for Learning

There are numerous definitions of Serious Games (SGs) as well as a number of terms for it. The term Serious Gaming was coined by Ben Sawyer in the white paper Serious Games Initiative [3]. The terms used in this area include Serious Games, but also Gamification, Game-based learning, Educational games, Simulation for learning, Educational simulation *etc.* Definitions vary from “Computer games that are intended to not only entertain users, but have additional purposes such as education and training”, over “A serious game is usually a simulation which has the look and feel of a game, but is actually a simulation of real-world events or processes” to “A serious computer game is any kind of interactive application - for example a computer simulation or a micro-world - that is designed and implemented according to gameplay principles” [4]. One broad and probably one of the best definitions say that SGs are “The application of gaming technology, process, and design to the solution of problems faced by businesses and other organizations.

Serious games have become both a growing market in the games industry [5] and a field of academic research [6]. Serious Games combine the elements of gameplay like interactivity and immersion with modern teaching methods and imply the opportunity of a high grade of motivation, active decision-making and sophisticated game environments. A further development in this field is the Intelligent Tutoring System (ITS). Later, these systems were also incorporated with embodied animated tutors called Interactive Animated Pedagogical Agents (IAPA) such as Steve [7] and Au-toTutor [8] in order to increase the engagement by the means of adding social interaction capabilities. Eva was another agent that helped students learns about music [9] while its robotic counterpart [10] taught photography to students. Such agents are equipped with a conversational and behavioral system and may have personality and emotions. They can also motivate the students by encouraging them and by asking questions, or referring to past examples. In some cases, they also perform the domain task themselves and show how to do it. Such a learning framework offer the opportunity to implement technology from intelligent tutoring, pedagogical agents and virtual worlds and combine those with game techniques since games attract people with challenge and curiosity factors as well as bring the concept of "learning by doing". Purdy [11] mentions that games work in education because they have the potential to assist deep learning. Players make decisions that have consequences and they actively participate in the game or virtual environment. They can also safely try out multiple solutions, test their own limitations and assume different roles.

When it comes to advantages of this new type of learning framework, they are found to promote learning [12-14]. Regarding positive impacts, games can also support the development of a number of different skills, as discussed by [15], analytical and spatial skills, strategic skills and insight, learning and recollection capabilities, psychomotor skills, visual selective attention, *etc.*, and even violent games can be beneficial in that they provide an outlet to alleviate frustration. Wouters *et al* [16] have found SGs to improve the acquisition of knowledge and cognitive skills. It may also be concluded that the use of serious games is promising for the acquisition of fine-grid motor skills and to accomplish attitudinal change [17].

2.2.1. Affective Learning

Emotion skills, like coping behaviours, are central to human communication. Actually, many of the cognitive processes are interdependent and they rarely occur in isolation [18]. Emotions are also known to affect judgment, especially of a social kind and be used in persuasion. One can thus convey positive feelings by smiling, frowning or winking. Emotions certainly influence learning and both negative and positive emotions may have large effects on problem solving [19-20]. Therefore, in a simulation environment like the current presented framework, where decision making in complex situations and where encounters with cultural, social and/or mentally diverse humans should be trained, it will be necessary to elicit and monitor various emotional reactions and responses [21]. However, knowing that positive emotions have been observed to promote learning, our planned learning system will consequently elicit and provoke both positive and negative emotional responses. Positive responses from the virtual tutor might also be crucial in some situations.

Emotional mediated experiences have been reported as having a positive impact on cognitive learning outcomes [22-23]. By getting rid of internal and external dis-tractions, they do enhance concentration and attention. Although they are not directly responsible for knowledge or skill acquisition, they still contribute to efficient and reproducible reasoning patterns. Nevertheless, it is important to be aware of the effect of multi-sensory integration on emotional engagement (because of human’s selective attention in hearing, vision and touch). There-fore, presenting information in several sensory modalities (rather than one) leads to a more efficient use of memory resources [24-25]. We can thus easily conceive the potential educational benefits of multimedia technology using multi-sensory presentation of graphics audio and visual materials. The same goes for the importance of naturalness of visual and auditory interaction for the sense of mediated presence and interaction with virtual character that combine several modalities [26].

2.2.2. An Innovative Training Model

Our research group has during the recent years focused on how to design an IN based model for training decision making skills in complex situations where interaction between people, things and environments are crucial. This includes scenarios where individuals are more or less unwilling or incapable of cooperation and where also stress and time restrictions further complicated the decision process.

In such training environments continuous interpretation of multimodal physiological and emotional inputs would allow the interactive learning environment to be more receptive to the intended actions, as opposed to the traditional tracking of actual performed actions. This would in turn lead to a more personalized and accurate summative feedback (by providing objective measurements of emotional and social states of each user) and hopefully as a result induce behavior change/self-regulation on users. Combining traditional summative feedback with our framework's special debriefing feature where the virtual persons themselves can give personalized feedback to the learner in terms of spoken and animated dialogue, may also enhance deep learning and learners' metacognitive insights.

Such interactive socio-educational training models would definitively con-tribute to improved encounters with unwilling, less capable or otherwise affected humans like refugees, psychiatric patients or criminal care clients and enhance confidence in trust building, challenge coping behaviors and emotional responsiveness (*e.g.* relation skills like empathic communication). This will in its turn improve learner abilities of decision making in complex situations where human interpersonal communication is a key. Participants would thus gain a sense of mastery and self-confidence in decision-making that would rely on the synergy of cognitive and affective processes. Moreover, decision-making is a cognitive process involving reflection. This enables thinking about what to do, what the options are and thinking ahead considering what the consequences might be of carrying out certain actions before selecting strategies for carrying out tasks. Awareness mechanisms can also be "trained" and fine-tuned by purposely and inadvertently manipulating virtual actors in the physical and social context and/or overcharging their nonverbal behaviors. Collaborative environments for decision-making should not impede awareness in order to convey precious information about peoples' intentions and beliefs.

To instantiate the conceptual model of our learning framework, we have designed a "natural" interface giving the illusion of behaving and looking like real-world counterparts as well as providing a distributive learning environment. The interface can present a complex situation and be able to measure in real time the actions and reactions of the trainees. The monitored data will be continuously fed back into the system and the virtual agents will act accordingly (by means of dynamic scripts and predefined personalities and adaptive moods). Special emphasis will be put on exploratory interaction in an intercultural training environment supporting reflective cognition and affective involvement.

In order to offer a trustworthy and believable learning environment, the virtual subjects/actors will be defined with different personalities (extrovert, dominant, worried *etc.*) and endowed with different roles (like lawyer, mental patient, refugee, immigrant, client, *etc.*). These virtual humans will be independent, able to evolve autonomously in the virtual world. They will also en-gage in meaningful social interaction with participants. Hence, three main kinds of social mechanism will be represented in the training system: conversation, coordination, and awareness. The conversational interface will rely on a semantic engine (language interpretation) for semantic comprehension. The virtual humans will thus handle a dialogue with the trainees, for example by asking follow-up questions to better understand what a user wants to know. By using a robust speech understanding, faulty grammar, misrecognized words (in spoken interaction) and misspelled words (in written interaction) will not hinder the interaction.

However, to offer a certain level of adaptively, users' physiological data and behavioral reactions will influence the behaviors of the virtual agents. These virtual humans will be able to take several strategies to respond to user's stress level, emotional state, disengagement or fatigue. The learning sys-tem will thus have the ability to detect physiological variables like stress level or cognitive fatigue on trainees in order to affect the agents willingness to share information or establish trust with trainees. In the next subsection we will explore the level of training in the learning system via manage virtual spatiality for virtual agents in virtual environments by context-oriented association such as user modes of location and device position to imply knowledge assumptions and social environment scripts that give meaning to a situation.

2.3. The Exploring the Level of Training toward Seamlessly Manage Virtual Spatiality

The innovative on rethinking context to find a feasible solution to seamlessly manage virtual spatiality and augmented reality in a technology independent manner. In this way context-oriented association technologies to explore logical object-oriented technology to enable a system-initializing interaction mechanism for addressing social environment scripts and experiences with respect to system service logic. To establish and maintain the logical associations between time-critical situational awareness and social environment interactions are based upon establishing and information sharing via a crowd-sensing technology on that allows users to share context information based on situational awareness. To meet user accuracy requirements as confident con-text sensing, we use a system model as shown in Figure 3 for description of context sensing diversity and capability differences among heterogeneous and homogeneous context sources in a Human-supported learning system deployment. By exploring context sensing diversity to provide a conceptual construction of an alternative reality for virtual environments as a context for training in the learning system might have something to teach us about the construction of mixed realities for training purpose, Both seem to be about conveying meaning by constructing spaces, which the viewer/user is invited to explore and interpret. The concept will be able to create 'realities' which provide a seamless integration of the different realities, while at the same time keeping the strengths of each constituent reality. The next section provides further description on the concept of context-aware computing system for serious gaming technologies.

3. Conceptual Framework

We have designed a novel type of advanced immersive networking as con-text-aware computing system for human decision-making associated with human encounters. The new framework supports a wide range of complex scenarios applicable in formal and informal education, workplace training, life-long learning, policy making and in social and public processes. The features and components will include *e.g.* virtual human-like characters, interaction systems, novel human-machine interfaces, 3D-models, textures, simulations, game design, emotional models, emergent narratives *etc.* as well as an intuitive authoring environment and interfacing with existing game engines. A wide variety of training solutions will be supported, including one to one, one to many and local/distributed learners.

3.1. Architectural Overview

The framework is based on reusable, open components with concept of context-aware computing system for serious gaming technologies which exhibit novel capabilities for knowledge acquisition and training of skills associated with decision making in human encounters. This section describes the components and their organization within functional units within the architecture of the framework. Figure 2 provides the overview and highlights (in light blue) key components.

The concept supports a wide variety of user input and output modalities including audio, visual information, as well as sensing and actuation of physical parameters. The system is open ended and configurable in terms of modalities in accordance with the interaction described in the previous section and some key examples are given in the “Sensing” and “Multimodal Interface” functional units as depicted in Figure 2. The machine-learning and data-mining components in the “Context Sharing” functional unit carry out clustering and self-organization of information. From these heterogeneous sources sinks of input and output process, machine-learning and data-mining components fuse, extract and share presence- and context information.

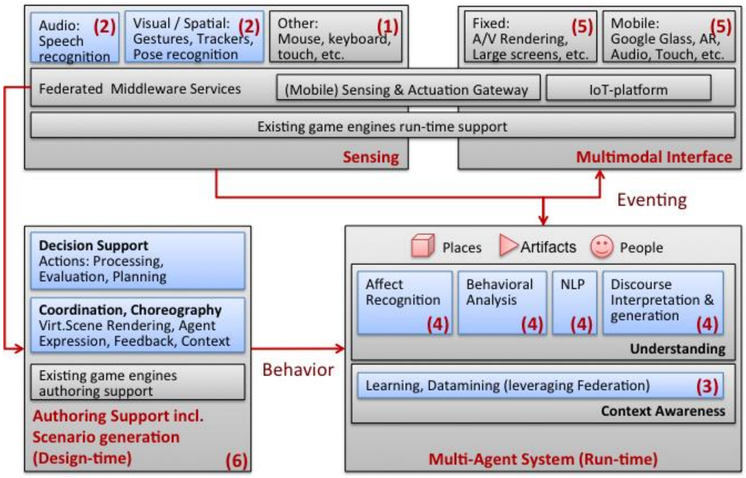

Figure 2. Architectural Overview

The components of the “Understanding” functional unit use the self-organizing capabilities and information from the “Context-Sharing” function-al unit to enable the entities modelled in the “Multi-Agent System” functional unit modify and react to the input. In a similar fashion, the components of the “Decision-Making” functional unit use the responses from the agents in the “Multi-Agent System” to provide input to the “Coordination, Choreography” functional units via the “Context-Sharing” functional unit. Thus, the “Coordination, Choreography” functional units can leverage the self-organizing capabilities and information from the “Context-Sharing” functional unit to adapt and present output in the form of Virtual characters, speech, scene back-ground and more via the “Multi-Modal Interface” functional unit.

Blue sections in Figure 2 indicated components that will be developed (or re-designed) within the suggested framework. Also other inputs and output sources may be used (grey) to develop a Human-supported learning system. The numbers in brackets corresponds to data stream/analysis/outputs indicated below.

The architecture is thus subdivided into two main areas:

Design-time support entailing authoring support including scenario generation, capable of interworking with existing game engine authoring support.

Run-time support entailing sensing (input) & multimodal interface (output) capable of interworking with existing game engine authoring support and connected to a multi-agent system harboring the handling of complex behavior pertaining to scenarios.

The devices and modalities involved for in- and output may range from, *e.g.*, tablets, smartphones, smartTV, *etc.* with various means of network connectivity, such as WiFi, Bluetooth, and mobile broadband. Hence, we presume users, places and artefacts involved in all types of interaction to be connected, reachable on the Internet and able to share information via federated middleware services. Thus, the architecture includes the following functional areas and capabilities with inputs, data streams, analyses and outputs listed in Table 1 below:

Table 1. Examples of Inputs, Data Streams, Analyses and Outputs for the Architecture

Functional Areas		Capabilities	Sources, Devices
1.	Input	Text, pointing	Keyboard, Mouse
		Gesture, pose	Touch screen interface Trackers
		Location, direction, movement	GPS, mobile positioning, accelerometers, *etc.*
		Ambient physical parameters	Moisture, light, *etc.*
		(Spatial) audio, video	Microphone(s), Camera(s), Kinect , *etc.*
2.	Sensing related to input	Visual capture & analysis (facial expressions, gaze, gesture, pose, *etc.*)	Face expression recognition
		Audio capture & analysis	Speech recognition, NLP
		Social proximity	Determined from analysis of profiles and preferences of actors.
		Spatial proximity	Determined from analysis of spatial and/or physical parameters of actors.
		Existing game engine events	
3.	Context Information	Fusion and Self-Organization	
		Acquisition and Dissemination of Context Information	
4.	Modelling and analysis	Objects, Relations, and Behaviour	Learners reactions
		Affect recognition	Interpretation of learner affect state
		Graphical Representations	Rendering of ECAs and scenes
		Dialogs, Storylines, Scripts	Scripts and editing of dialogue
5.	Multimodal Interface incl. output devices	In situ, Free Movement: Augmented Reality	Google Glass, *etc.* Mobile devices, Smartphone, Tablets, *etc.* Remote with Free Movement Remote with Limited Movement Large Screen, *e.g.*, Smart TV, *etc.* Computer with Gaming Auxiliary Devices, *etc.*
		Traditional computers, monitors	
6.	Authoring support	Human authoring support	Novel basic authoring support for instructors and basic users
		Existing game engine authoring support	Advanced authoring by specialists

3.2. Diversity of Context Information, Objects and Model

To meet user accuracy requirements as confident context sensing we proposed a new approach as shown in Figure 3 below. This model describes context sensing diversity and capability differences among heterogeneous and homogeneous context sources in a Human-supported learning system deployment.

We explore the impact of multiple context sources on context values collaboration and exploit diversity for confident context sensing coverage to support the dynamic nature of diversity of context information as well as the heterogeneous array of applications and service requirements. Users interact or even immerse themselves in virtual reality plays if such plays stirred their experiences intensely and broadly. In achieving this, multiple sources of information must be considered when observing the interaction between users with their environment. This provides the information required to build relational con-text models that can identify and establish relationships between avatars. The-se relationships in turn are the springboard for creating situations where an avatar's perceived relationships are elevated and through a diverse set of unintended actions to immerse enriched and novel creative experiences. In this respect context data is the key for the creative systems bringing about such communication patterns and experiences. The proposed model is a collection of objects with defined relationships. As an object-oriented approach, it inherently gains the properties of class inheritance and as such can be extended with new sub-concepts of Entity and Context Information-Source. This is made possible by adaptive software techniques such as a combination of computational refection, automated compilation and dynamic class loading. Agents, applications and services reside above and use the meta-model as a source of data and deriving diversity context information. The Contextual computing part (in the model) is the use of software and hardware to automatically collect and analyze data about a device's surroundings in order to pre-sent relevant, actionable context information to the end user via Google Glass. This kind of devices can further the experience, providing the end user with an augmented reality experience.

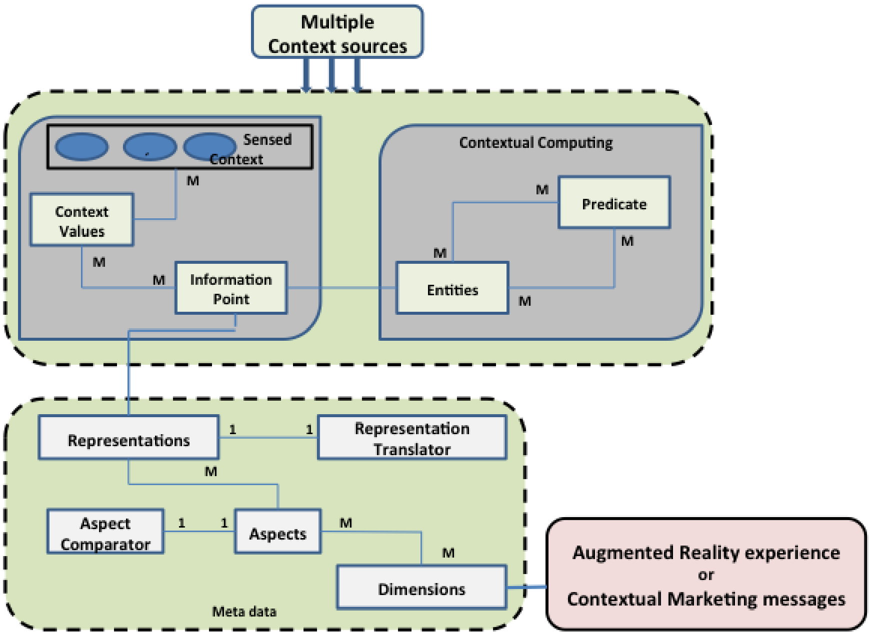

Figure 3 Diversity of Context Information Model

4. Design and Development Methodology

The choice of design and development methodology can be understood in two ways: first a methodology of ontology for framework development and second, a methodology for generating specific architectures. Please refer to Figure 4 below concerning how to derive all aspects of the model as well as how to drive the best practices.

Simply dissecting them into design steps and processes is not enough - one needs to know how to achieve each step. In the table below we summarize how architecture methodologies can be used for the development of the framework on the higher abstract level.

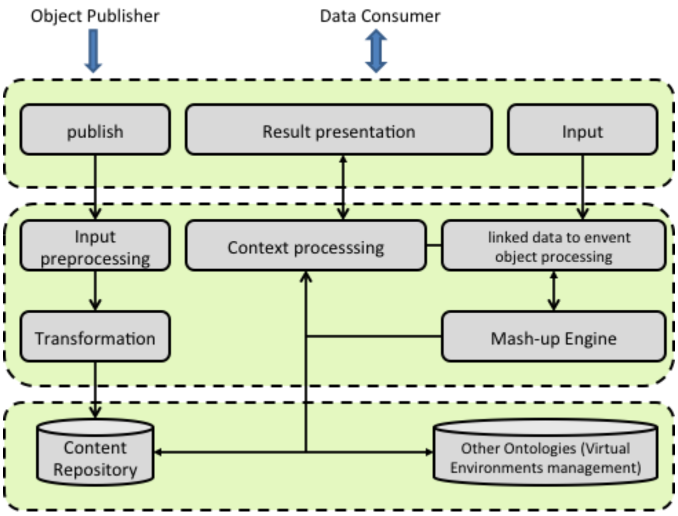

Figure 4. Main Component in the Framework for Architecture Methodology

Table 2. Usage of Architecture Methodologies for the Development of the Framework

Methodology	Aspect adopted in framework.
Aspect oriented programming	Delineation of functionalities by aspects. This is embodied in the concept of functionality groups.
Model-Driven Engineering	General concept of transformation from a generic to a more specific model. We use this concept for describing and developing our best practice.
Pattern-Based Design	We will test the efficacy of this method upon deriving a concrete architecture as the best practice's test case.
View and perspectives	We adopt the concept of view and perspectives for the derivation of the crowd-sensing platform reference architecture; we arrange all aspects of our reference architecture according to views and perspectives. The same is done for the unified requirements.

5. Results and Evaluation

To evaluate the effectiveness of framework in terms of creating massive, scalable participatory and immersive experiences in live cultural events, we developed a prototype implementation of the components and performed experiments using this prototype. All data recorded from the runs of the application case studies described in Sections 5.1 to 5.3 and in all experimentations scenarios by using audiences in the pervasive and ubiquitous service domain, directed through logical sensors and connected role cluster in different locations. The results demonstrate that we are capable of realizing collaborative experiences that are built on relevant existing and stored content. These experiences are enhanced through decisions that leverage context interoperability accuracy based clustering of objects, people and places engaged in performing art productions in suburban environments.

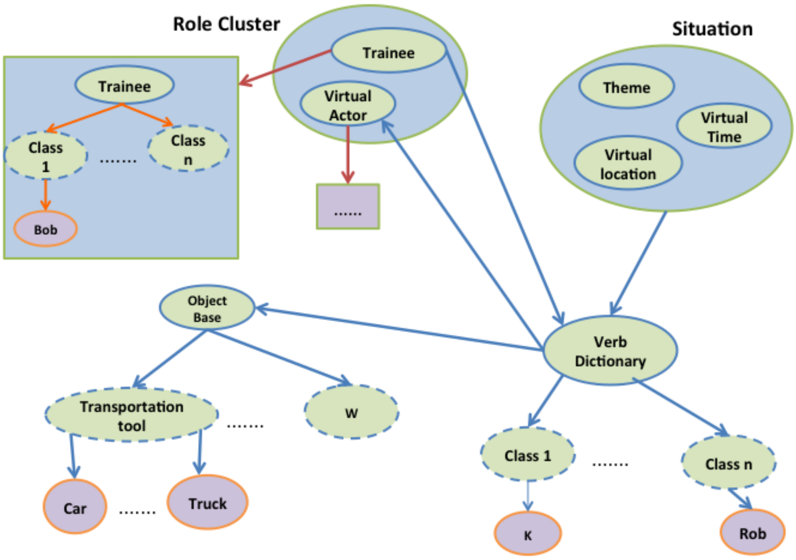

Figure 5. The Ontology Based Prototype as an Example

To validate the framework, we defined an ontology shown in Figure 5. Each white ellipse is a sub context domain, while each arrow shows the entity properties for each entity. Entity property is the way to mark the relations among entities, where each entity can be regarded as a property for its peers. The blue ellipse describes the constructions for the message in each context domain. According to a case scenario abstracted from the system, Position, Trainee Status, Geo-information, situation-information and Actions are extracted out as the context domains, where each message can be marked with category under the construction format. A proof-of-concept prototype is written with Jess as shown in [27] which helps establish the entity proper-ties. An OWL/XML file is generated by a short Jess code, as illustrated in Figure 6.

```
<rdf:RDF xmlns="http://www.owl-ontologies.com/Ontology1410186992.owl#
xml:base="http://www.owl-ontologies.com/Ontology1410186992.owl"
<owl:Ontology rdf:about=""/>
  <unnamed_class1 rdf:ID="Bob"/>
  <Wapon rdf:ID="Bomb"/>
  <Transportation_tool rdf:ID="Car"/>
  <Theme rdf:ID="Disaster"/>
  <Wapon rdf:ID="Gun"/>
  <Virtual_Time rdf:ID="0:00"/>
  <Virtual_Time rdf:ID="1:00"/>
  <Virtual_Time rdf:ID="2:00"/>
  <Virtual_Time rdf:ID="3:00"/>
  <Verb_Dictionary rdf:ID="Kill"/>
  <owl:Class rdf:ID="Object_Base"/>
  <Verb_Dictionary rdf:ID="Rob"/>
  <Theme rdf:ID="Romantic"/>
  <unnamed_class1 rdf:ID="Sam"/>
  <owl:Class rdf:ID="Theme"/>
  <owl:Class rdf:ID="Trainee"/>
  <owl:Class rdf:ID="Transportation_tool">
    <rdfs:subClassOf rdf:resource="#Object_Base"/>
  </owl:Class>
  <Transportation_tool rdf:ID="Truck"/>
  <owl:Class rdf:ID="unnamed_class1">
    <rdfs:subClassOf rdf:resource="#Trainee"/>
  </owl:Class>
  <owl:Class rdf:ID="unnamed_class2">
    <rdfs:subClassOf rdf:resource="#Trainee"/>
  </owl:Class>
  <owl:Class rdf:ID="Verb_Dictionary"/>
  <owl:Class rdf:ID="Virtual_Actor"/>
  <owl:Class rdf:ID="Virtual_location"/>
  <owl:Class rdf:ID="Virtual_Time"/>
  <owl:Class rdf:ID="Wapon">
    <rdfs:subClassOf rdf:resource="#Object_Base"/>
  </owl:Class>
</rdf:RDF>
```


Figure 6. The OWL/XML File Generated Depicting the Context

5.1. Case Study1

The training framework will use the components mentioned in Sections 2 and 3 to facilitate educational scenarios that may look like the example shown in Figures below.

The possible training scenarios cover a wide variety of domains, but which may be illustrated like the example shown in Figure 7 below where a single learner is training interviewing skills for encountering a refugee.

Figure 7 show one learner who is interfacing with a single virtual actor via speech, keyboard input and while the system is monitoring the users emotion-al and physical reactions. The training model is constantly analyzing the user data and automatically (and/or via a scripted flow), reacts to the user input and monitored data to create output like dialogue in natural language, gaze postures, background and feedback to the learner.

Figure 7. Learner Interfacing

5.2. Case Study 2

However, also more complex scenarios can be supported, like when a group of physically distributed learners are training to work as a team to interact with a multitude of persons like as shown in Figure 8 below.

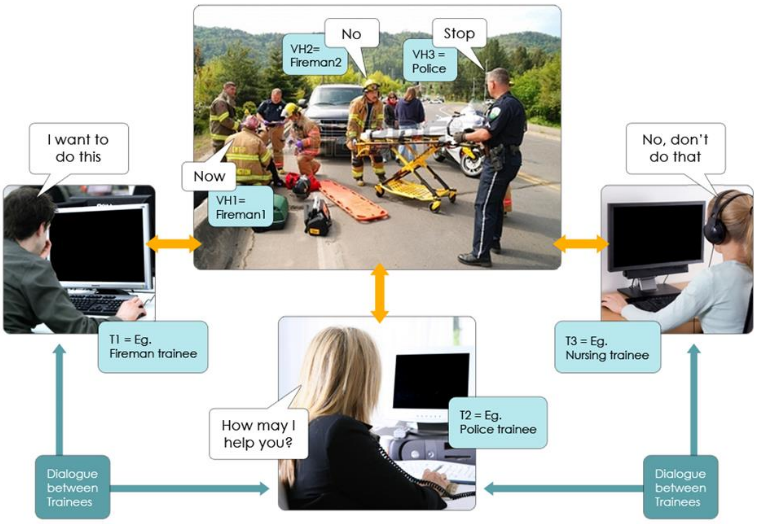

Figure 8. Physically Distributed Learners

Three physically distributed learners are interfacing with a number of virtual actors on a virtual scene. Also user inputs like speech, gaze, facial expressions *etc.* are monitored, analyzed and used to create the output to the virtual scene. The distributed learners also interact with each other and train to collaborate on complex tasks.

In this way, the new training concept can be applied to a vast number of skills acquisition scenarios where a combination of factual domain knowledge and results from human-human encounters are needed to solve complex problems and make correct decisions, in time and resource restricted situations.

5.3. Case Study 3: Exploring Augmented Reality & Contextual Computing

Contextual computing especially with Google Glass, which knows where you are, what you're doing, what you like, pretty much everything about you is about understanding consumers and providing the information or the ability to do something when they need it. Context-aware applications development with Google Glass are in rampage, which is a big contextual play. This innovative technology provides the combination of augmented reality and a wear-able device for a new way to capture and interact with context information. We explore the previous case study via exploring Google Glass technology as shown in Figure 9 for Virtual Humans and dynamic scene by first capture and explore context sensing diversity by the deployment model shown in Figure 3 with a flow sensor network deployment [28] for con-textual detection. We show that context sensing capabilities greatly differ among flow sensors in a real deployment and identify when flow sensor collaboration is needed. When collaboration is needed, we show that arbitrary typically sensor collaboration often fails to meet user accuracy requirements, not to mention a joint consideration of accuracy.

Figure 9. Using Google Glass for Virtual Humans & Dynamic Scene

We explore different machine learning techniques for on-demand collaboration and identify one appropriate for contextual computationally constrained flow sensor networks. These results provide key insights into protocol design for collaborative flow sensing. This case study provides an extension of the framework that will be able to lead to further growing of the framework in case of reality mixer. The platform combines future Internet technology with augmented reality, further the platform will focus on bringing the gaming experience out of the living room to the world around users by interconnecting the crowd-sensing [29] with other part of platform make an immersive platform. The extended framework will be able to make possible the intersection of ubiquitous connectivity, networked media and distributed sensor networks to share data and use services as mobile sensor-based service [30].

5.3.1. The Concept of Serious Gaming Technologies as Context-Aware Computing System

Context-aware computing systems aim to proactively help users by automatically sensing the context and adapting system behaviour to address users' needs and expectations. The context can comprise personal factors such as physical activity, social interactions, and the psychophysiological or affective state as well as environmental factors such as location and the surrounding infrastructure. We will try to make the concept to be simple, by following a seven-step process which shown in Figure 10.

Understanding the context (Attention), concept formation (Defining learning objectives), Reasoning (structuring the Experience), Intention (identifying Resources), Memory, Applying Gamification elements (such as Learning and Decision Making).

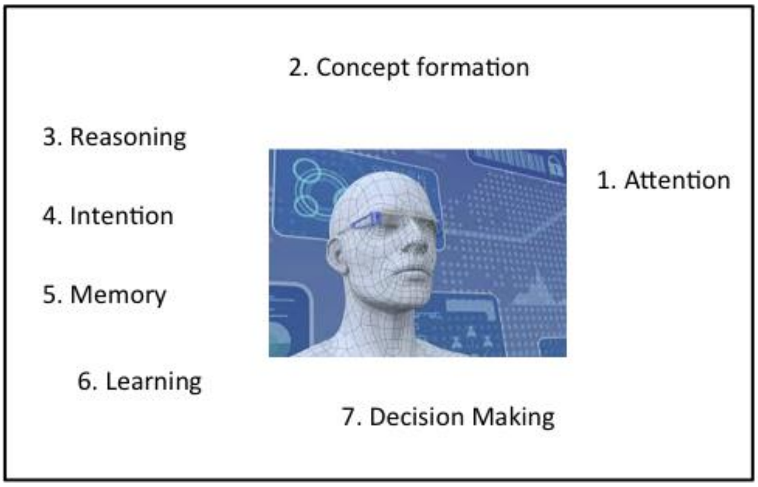

Figure 10. Context-Aware Computing in Serious Gaming Technologies

A key factor that determines the success of a training program, is a good un-der-standing (**Attention**) of who the trainee is. This combined with context which the program is being delivered, will help in designing a program that empowers the trainee to achieve the objective of the program.

While an analysis of the target context via Google Glass will help determine factors like diversity of context information for a propose group, learning abilities, current skill-set, *etc.*, analyzing the context can provide with details of the trainee group size, environment, sequencing of skills, and the time frame.

Concept formation: Specific Learning Goals which could include the trainee understanding a concept, being able to perform a task after the training, or completing the learning program.

Reasoning (structuring the Experience): Stages and milestones are powerful tools that enable trainer to sequence knowledge and quantify what the trainee need to learn and achieve by the end of each stage or milestone.

Intention (identifying Resources): Once the stages have been identified, the trainer can more easily judge which stages, can be gamified, and how. Questions an instructor should think about while considering gamification.

Applying Gamification elements (Memory, Learning and Decision Making): The gamification process in training comes down to the elements that are applied to the training program. Game mechanics can be classified as self-elements (competing with themselves and recognizing self-achievement) or social-elements (interactive competition or cooperation). In both game mechanics using a crowd-sensing platform via Google Glass to improve the identification of things which allow us to discover these, such as new building or event is then receiving its information from various sources even sharing the similarity of the context information.

The extended framework adds a distributed infrastructure with a distributed global collection of naming, localization, representation local & global and 3D organization in order to achieve fully immersive experience in real-time.

Several challenges exist in providing confident immersive experience in real-time by flow sensing through exploitation of Google Glass technology and sensing diversity:

Learning flow Sensing Diversity, Machine learning techniques many of which can be utilized to learn the activity discovery and activity-aware application will be designed for trainee to identify when an individual performs an activity as a function of learning.

On-demand flow sensor collaboration, We derive the guidelines for when flow sensor collaboration is needed and when it is not.

Distributed and online diversity exploitation for Augmented Reality Tracking & Reality Mixer make correlation among objects in the framework and the objects or sensing events obtained by means of external real-world sensing *e.g*., via a Google glass. Improve the identification of things by means of crowd-sensing mechanisms *e.g*., via a crowd-sensing centred application. This application is based on MediaSense which is an open-source platform offering scalable, seamless and real-time access among a global collection of entities (artefacts, people, places, information objects) via a heterogeneous network infrastructure [1]. This constitutes an Internet of Things services infrastructure negotiating a heterogeneous infrastructure of local gateways, mobile and fixed access networks. Mediasense which allow us to discover these, *e.g*., a new building is then receiving its information from various sources (people, *etc.*, contribute to this.). Co-visualize the context information of entities with the very same entities in real-time as the information changes. Immersive environment tracking in terms of correlation of different sources of position and locality from Google glass, smartphones, mobile positioning systems, WLAN.

6. Conclusions

This publication presented Immersive Networking (IN) as a novel frame-work for connecting people and places and things in virtual environments with augmented reality. IN has capabilities beyond those in existing frame-works such as MPEG-V to support the integrating with entities in the real world via heterogeneous infrastructure through distributed control. The MPEG-V framework [1] deals only with virtual world object characteristics (metadata) shared by both virtual agents and generic virtual objects. In contrast, underpinning IN as a novel framework, MediaSense [1] constitutes a scalable self-organizing means for connecting people and places and things in virtual environments with augmented reality. As a result IN enables seamless experiences in virtual environment through self-organization of connectivity between people, places and things on top of heterogeneous 5G and Internet-of-Things infrastructures.

A second important aspect of the quality of our experience is the immediacy of responses. IN supported by MediaSense enables seamless and self-organizing connections between entities to push control to the end-devices co-located with the entities themselves. These end-devices may incorporate sens-ing gateways and interaction devices, which include both local and non-local information from the virtual environment in the interaction. The IN frame-work allows delegation of control to end-devices though functionality in MediaSense for the organizing or relations and clustering by relevance. This capability is particularly important as the projected number of devices and sensors to be connected via the Internet-of-Things is projected to be in the order of 50 billion by 2020. MediaSense moves control to the edge enabling immediacy in experiences based on seamless self-organization through clustering of entities in relations organized by relevance.

Further, we presented a concrete architecture as an instantiation of the IN framework for the design of augmented reality experiences for human en-counters in virtual environments integrated with real-world objects. In comparison to related work, the architecture is context-based to drive adaptive behaviour of functional units in design-time and run-time support. The con-text-driven approach based on a diverse context information model puts objects (be it people, places or things) in the adaptive loop of the architecture. Through the distributed context sharing, the control is moved to the objects themselves.

We conclude by validating our approach in several scenarios evaluating the relevance and application in human decision-making.

Acknowledgements

The authors acknowledge researchers from Centre National de la Recherche Scientifique France, Idiap Research Institute Switzerland, University of Ge-neva, MIRAlab Switzerland, Noldus Information Technology B.V. Nether-lands and Serious Games Interactive Denmark for inputs discussions on the new learning model suggested.

References

[1] T. Kanter and J. Walters, "The MediaSense Framework, In Proceedings of 4th International Conference on Digital Telecommunications. IEEE, **(2009)**, pp. 144-147.

[2] J. H. A Gelissen, "Working Draft of ISO/IEC 23005 Architecture", "ISO/IEC JTC 1/SC 29/WG 11/N10616, Maui, USA, **(2009)**.

[3] B. Sawyer, "Serious games: improving public policy through game-based learning and simulation: Whitepaper for the Woodrow wilson international center for scholars, **(2002)**.

[4] L. P. Rieber, "Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. Educational Technology", Research and Development, vol. 44, no. 2, **(1996)**, pp. 43-58.

[5] T. Susi, M. Johannesson and P. Backlund, "Serious Games an overview", from http://www.his.se/upload/19354/HS-%20IKI%20-TR-07-001.pdf, **(2007)**.

[6] U. Ritterfeld, M. J. Cody and P. Vorderer, "Serious Games: Mechanisms and Effects: Routledge Chapman & Hall", **(2009)**.

[7] J. Rickel and W. L. Johnson, "Animated agents for procedural training in virtual reality: Perception, cognition, and motor control", Applied Artificial Intelligence, vol. 13, **(1999)**, pp. 343-382.

[8] A. C. Graesser, P. Chipman, B. C. Haynes and A. Olney, "An intelligent tutoring system with mixed-initiative dialogue", IEEE Transactions in Education, vol. 48, **(2005)**, pp. 612-618.

[9] M. B. Moussa, Z. Kasap, N. M. Thalmann, K. Chandramouli, S. N. H. Mirza, Q. Zhang and P. Daras, "Towards an expressive virtual tutor: an implementation of a virtual tutor based on an empirical study of non-verbal behaviour", Paper presented at the 2010 ACM workshop on Surreal media and virtual cloning (SMVC '10), New York, USA, **(2010)**.

[10] Z. Kasap and N. M. Thalmann, "Towards Episodic Memory Based Long-term Affective Interaction with a Human-like Robot", IEEE International Symposium on Robot and Human Interactive Communication, **(2010)**.

[11] J. A. Purdy, "Serious Games: Getting Serious About Digital Games in Learning", Corporate University Xchange, no. 1, **(2007)**.

[12] J. M. Randel, B. A. Morris, C. D. Wetzel and B. V. Whitehill, "The effectiveness of games for educational purposes: a review of recent research", Simulation and Gaming, vol. 23, no. 3, **(1992)**, pp. 261–276.

[13] M. Szczurek, "Meta-analysis of simulation games effectiveness for cognitive learning", (EdD), Indiana University, **(1982)**.

[14] V. Sickle, "A Quantative review of research on instructional simulation gaming: a twenty year perspective", Theory and Research in Social Education, vol. 14, no. 3, **(1986)**, pp. 245-264.

[15] A. Mitchell and C. S. Smith, "The use of computer and video games for learning", A review of the literature, **(2004)**.

[16] P. Wouters, E. V. D. Spek and H. V. Oostendorp, "Serious games for crisis management: What can we learn from research on animations", Retrieved from http://www.readanybook.com/wouters-et-al-serious-games-for-crisis-management-pdf-i742660, **(2014)**.

[17] P. Wouters, E. V. D. Spek and H. V. Oostendorp, "Current practices in serious game research: A review from a learning outcomes perspective Games-Based Learning Advancements for Multisensory Human Computer Interfaces: Techniques and Effective Practices: Hershey", **(2011)**.

[18] J. Gratch and S. Marsella, "The Architectural Role of Emotion in Cognitive Systems Integrated Models of Cognitive Systems: Oxford University Press", **(2007)**.

[19] R. A. Calvo and S. K. D'Mello, "New Perspectives on Affect and Learning Technologies. Explorations in the Learning Sciences", Instructional Systems and Performance Technologies, 3. doi: DOI: 10.1007/978-1-4419-9625-1, **(2011)**.

[20] S. K. D'Mello, B. A. Lehman and N. Person, "Monitoring Affect States During Effortful Problem Solving Activities", International Journal of Artificial Intelligence in Education, vol. 20, no. 4, **(2011)**, 361-389.
[21] S. K. D'Mello and A. C. Graesser, "Dynamics of Affective States during Complex Learning", Learning and Instruction, vol. 22, **(2012)**, pp. 145-157.
[22] C. Dede, P. Whitehouse and T. B. L'Bahy, "Designing and Studying Learning Experiences That Use Multiple Interactive Media to Bridge Distance and Time Current Perspectives on Applied Information Technologies", Distance Education. Greenwich, Conn: Information Age Press, vol. 1, **(2002)**, pp. 1-30.
[23] C. Sansone and C. Morgan, "Intrinsic motivation and education: Competence in context", Motivation and Emotion, vol. 16, **(1992)**, pp. 249-270.
[24] R. Moreno, "Does the modality principle hold for different media? A test of the method-affects-learning hypothesis", Journal of Computer Assisted Learning, vol. 22, no. 3, **(2006)**, pp. 149-158.
[25] R. Moreno, R. Mayer, H. Spires and J. Lester, "The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact With Animated Pedagogical Agents", Cognition and Instruction, vol. 19, no. 2, **(2001)**, pp. 177-213.
[26] S. Buisine and J. C. Martin, "The effects of speech-gesture co-operation in animated agents' behaviour in multimedia presentations", Interacting with Computers, vol. 19, **(2007)**, pp. 484-493.
[27] C. Harry, "Intelligent Agents Meet the Semantic Web in Smart Spaces", Internet Computing, IEEE, vol. 8, no. 6, **(2004)**, pp. 69-79.
[28] H. Rahman, R. Rahmani and T. Kanter, "Enabling Scalable Publish/Subscribe for Logical-Clustering in Crowdsourcing via MediaSense", IEEE Technically Co-Sponsored Science and Information Conference 2014, London UK, **(2014)**.
[29] T. Kanter, R. Rahmani, J. Walters and W. Sauter, "Distributed Immersive Participation As Crowd-sensing in Culture Events", Published in Journal of Virtual World Research, **(2014)**.
[30] T. Kanter, R. Rahmani and Y. Li, "MediaSense: A Scalable Testbed for IoT Applications", Accepted for publication on the EU-China FIRE website, **(2014)**.

Authors

Theo Kanter, earned a Ph.D. in computer communications, from the Royal Institute of Technology. Theo has held a number of leading positions in telecommunications research. From 1999 to 2007, he was a senior scientist at Ericsson Research. He is now a professor at the Department of Computer and System Sciences at Stockholm University, where he leads research in adaptive and context-aware mobile communication, service architectures and self- organizing application infrastructures.

Uno Fors, is professor in IT and Learning at the Department of Computer and System Sciences of Stockholm University with research focus on Simulation and Visualization for learning and assessment. He is especially focused on virtual patients and other virtual encounters for learning in eg. law, teacher education *etc.* Fors has published more than 170 research publications and has been leading several international and national research projects within these fields. Simulation systems like Web-SP created by Fors is used worldwide in different research projects and education.

Fors, is also active in research in healthcare informatics, including decision support, electronic health records and simulation tools for diagnosis and rehabilitation in psychology, psychiatry and domestic violence. Prof Fors is head of the Department of Computer and System Sciences of Stockholm University.

Rahim Rahmani, earned a Ph.D. in communications in heterogeneous networks and he is an Associate Professor of Computer Science at the Department of Computer and System Sciences of Stockholm University, where his research focuses on Collaborative ubiquitous services and security and Ubiquitous networks. He has served as reviewer and in technical committees of international conferences. He is a member of the editorial board of International Journal of Wireless Networking and Communications.