$\mathcal{V}i$TALiSE: Virtual to Augmented Loop in Smart Environments

Engineering Multiagent Systems Track

Paper #529

ABSTRACT

Future workplaces will be smart environments providing human users with features and functionalities augmenting their capabilities while lowering their cognitive/physical efforts. This cannot disregard the synergistic interplay between the physical and digital reality. $\mathcal{V}i$TALiSE (as *Vi*rtual *t*o *A*ugmented *L*oop *i*n *S*mart *E*nvironments) is a vision of future smart environments integrating Human-Agent Collectives (HAC) with Digital Twins (DTs). HAC puts hw/sw agents on par with humans, and promotes seamless social interaction, but lacks thorough investigation of the deep relations between both human and software agents with their environment, which is simultaneously physical and virtual. DTs focuses on projecting physical things and processes into a virtual overlay, where humans can manipulate them with higher capabilities and lower cognitive effort, but limited attention is devoted to the fact that humans may work simultaneously in the virtual and physical layer. $\mathcal{V}i$TALiSE proposes to integrate the HAC and DTs perspectives to overcome their individual limitations.

1 INTRODUCTION

Augmented Reality (AR), Virtual Reality (VR), pervasive computing / Internet of Things (IoT), intelligent distributed computing, and Decision Support Systems (DSS) are enabling paradigms/technologies that can be integrated together to envision novel kinds of workplaces as next generation *smart environments*, flexibly mixing the digital and physical, simulated and real world, in a scenario which can be defined as *Mixed Reality* (MR). These smart environments will be designed and exploited to *support human activities* by any possible means, to enhance the way in which people think about and practically conduct their work, *interact*, communicate, and cooperate, to augment their physical and cognitive capabilities, as well as to relieve the respective workload.

Examples in different application domains include: urban planning and architectural design of buildings; serious games for human behaviour modelling or for the training of users "in silico"—due to practical or ethical reasons (i.e. life-threatening scenarios, costly operations, etc.); empowerment of field operators and coordination of teams, in areas that go from healthcare (i.e. "augmented surgery room", electronic medical records, clinical decision support), to security, entertainment, and leisure activities (i.e. "augmented museums" or historical sites enriched with virtual missing buildings / decorations); manufacturing and plant automation control (i.e. for diagnosing failures or perform maintenance operations remotely).

The vision of *Digital Twins* (DTs) [16] can be interpreted as a first step towards this concept of smart environment, although originally limited in scope to manufacturing systems. DTs, in fact, is a concept that projects physical things and processes onto a *virtual layer*, where humans can manipulate them with higher capabilities and lower cognitive effort, while also overcoming the limitations of the physical world (i.e. spatial and physical constraints). However, DTs focus is on the *uni-directional* coupling from physical to virtual, whereas the feedback link from virtual back to physical is less explored. There is also limited consideration of the *dynamic nature* of such interlinking, which is often statically established at few moments in time (i.e. beginning of product design and transitioning to manufacturing process). Furthermore, in DTs the fact that humans may work *simultaneously* in the virtual and physical layer, not mutually exclusively solely, is mostly neglected. Finally, another limitation concerns the possibility of a *multi-party* interaction in the virtual layer, where many people collaborate on the same product/process.

The latter limitation is instead the principal focus of the *Human-Agent Collectives* (HAC) vision [20], as stemming from the body of research at the crossroad between agent technologies and pervasive computing. HAC fosters next generation socio-technical systems in which software agents and humans *co-exist*, interact, communicate, and cooperate, hence promotes *seamless social interaction*. Accordingly, the main focus is about how to model the highly heterogeneous behaviour of software and humans (i.e. being the former fully predictable and reliable, whereas the latter is usually not) under a coherent conceptual and technical framework, and on how to enable, support, and promote seamless bi-directional interaction between the visibile (humans) and the usually invisible (software). A consequence of such a strong orientation is the lack of consideration for *situated interaction*, that is, for the deep relations between both human and software agents – hence between their actions and interactions, in turn – with their *environment*, which is simultaneously physical and digital.

To some extent then, HAC and DTs are specialised, complementary visions of next generation smart environments where the physical and digital realms are inextricably interwoven and whose respective boundaries are inevitably blurred. Hereby, we propose $\mathcal{V}i$TALiSE[1] as a foundational framework – conceptual, architectural, technological, and methodological – to integrate these visions under the perspective of Agent-oriented Software Engineering (AOSE) so as to overcome their individual limitations:

- to HAC, $\mathcal{V}i$TALiSE adds the physical environment to the picture, hence the dimension of situated interaction (besides the social one considered in HAC);

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 2019, Montreal, Canada

https://doi.org/doi

[1] As *Vi*rtual *t*o *A*ugmented *L*oop *i*n *S*mart *E*nvironments.

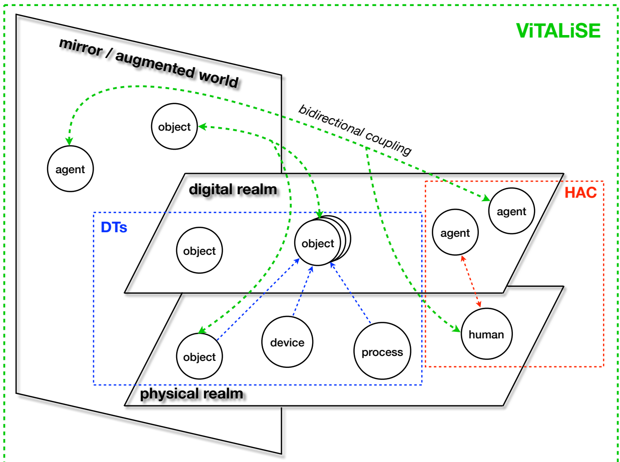

Figure 1: In *V*iTAL*i*SE, objects, devices, processes, and agents may have any combination[2] of digital and physical representations associated with them: digital only (as in VR), physical-to-digital (as in AR), and both physical-to-digital as well as digital-to-physical (as in MW/AW).

- to DTs, *V*iTAL*i*SE makes the uni-directional link going from physical to virtual become a *bi-directional* one, actually enabling a stronger, dynamic coupling between the physical and the digital.

*V*iTAL*i*SE allows system designers to consider multi-party cooperation (in the spirit of HAC) and simultaneous interaction with the physical and digital realms (through the enhanced, bi-directional coupling), in a setting which resembles that of *Mirror Worlds* (MW) [15, 32] or *Augmented World* (AW) [9, 10] – an extension of the MW concept – as graphically represented in Figure 1.

For doing so, particular focus is devoted to the engineering loop enabling *dynamic and adaptive coupling* of the digital and physical worlds that we call *V*iTA (as Virtual to Augmented). In there, humans and software agents *co-construct* a MR world, and *co-evolve* with it, meaning that, while agents *shape* the world and *adapt* to it by learning from the outcomes of their actions, the world too re-shapes itself and adapts to the inhabitant agents to better serve their needs—i.e. further augmenting their capabilities or further relieving their cognitive workload.

The remainder of the paper is organised as follows: Section 2 describes the overall *V*iTAL*i*SE vision, emphasising the main novelties w.r.t. HAC and DTs; Section 3 elaborates on the *V*iTA loop, as the core enabler of the kind of co-evolution fostered in *V*iTAL*i*SE; Section 4 identifies those technologies likely to play a key role as enablers of the *V*iTAL*i*SE vision, and discusses the main challenges still to be dealt with in order to bring *V*iTAL*i*SE to its full realisation; Section 5 concludes the paper.

[2]Physical only is obviously not of interest for the paper.

2 THE VISION OF *V*ITAL*I*SE

*V*iTAL*i*SE is a vision of future *smart environments* supporting humans activities being a *digital-physical mashup*, inextricably interwoven to form a Mixed Reality (MR) continuum, where digital and physical objects are *dynamically coupled* during operation, and *embodiment* of the digital into the physical is explicitly managed.

Let us delve deeper into each key term in the definition above to better understand the implications:

- we define as smart environments those workplaces where human activities are *enhanced* through computational technology by either augmenting their physical or cognitive capabilities, or by relieving the associated workload—as in the case of personal assistant agents [26] and decision support systems [38];
- the mashup between the physical and digital worlds is defined in the spirit of *Mirror Worlds* (MW) [15, 32], where physical objects have a digital counterpart continuously mirroring their structure, properties, and dynamics, and considering the full spectrum of Milgram and Kishino's Reality-Virtuality continuum [28] depicted in Figure 2, where objects existing in the virtual environment may have or not an image in the physical one, with varying degrees of coupling;
- we let objects in the digital and physical worlds have their *degree of coupling* dynamically established, depending on the goals of their users (or the application at hand), along two dimensions:
  - a digital object may mirror only the structure, properties, dynamics, or any different combination thereof of a physical one at any given time;
  - such a mirroring may have a uni-directional semantics (physical-to-virtual or virtual-to-physical) or a bi-directional one.

  For instance, a user may decide to "switch off" virtual-to-physical coupling of objects structure and dynamics when performing "what-if" analyses in a simulated environment—while still perceiving changes in their properties;
- we let digital objects be *dynamically embodied* into physical ones upon need, during operation, depending on the goals of their users (or the application at hand). As for the degree of coupling, such an embodiment may articulate along two directions:
  - digital-to-virtual, when digital objects are embodied into a visible representation in the virtual world (i.e. the case of a personal assistant agent associated to an avatar);
  - digital-to-real, when digital objects are embodied into a visible representation in the real world (i.e. a software robot controller deployed on a physical robot).

  Notice, for physical objects to be embodied as digital ones, they first need a digital representation, hence falling in the case digital-to-virtual—for instance, a human user moving in a MR scenario will necessarily have some kind of software agent associated, even in the form of a simple interactive UI, so as to model its behaviour in the digital realm.

It is worth noting here that the latter two items described above naturally fit the MiRA conceptual framework [18] for MR agents: there, the *interactive capacity* of agents is defined as their ability to sense and act on the environment (hence, resembles our notion of "coupling"), while their *corporeal presence* is defined as the property of having a visible representation in either the physical or virtual realm.

Leveraging on the synergies between all the above described features, $\mathcal{V}$*i*TAL*i*SE enables and supports the following concept of *co-evolution* within augmented / mirror worlds:

- digital and physical agents (humans) *together* build the virtual portion of the world, possibly (partially) coupled with the physical one, and act and interact with and within it to bring about their activities;
- in turn, the augmented world *reacts* to such activities in a number of ways, all meant to realise its function of better supporting agent behaviour, there including monitoring outcomes to *learn* from failures or inefficiencies, forwarding the effects of digital actions to the physical world, etc.;
- this causes a continuous *tension* between the virtual and augmented worlds, there including the inhabitant agents, in which on the one hand the agents and the digital objects continuously strive to improve their performance by learning from past experience, and on the other hand the overall digital representation of the world improves by learning from the continuous *feedback* of the agents living therein.

Responsible of the above "tension" between the virtual and the augmented is the $\mathcal{V}$*i*TA loop, described in next section.

3 THE $\mathcal{V}$*I*TA LOOP: FROM VIRTUAL TO AUGMENTED AND BACK

As the enabler of the co-evolution peculiarity of $\mathcal{V}$*i*TAL*i*SE, the $\mathcal{V}$*i*TA loop revolves around three main phases—depicted in Figure 3:

(1) *Virtualisation* of the physical environment $\mathcal{P}$ by means of proper sensor, computer vision, and data analysis technologies. This stage outputs a virtual model $\mathcal{V}$ coupled to the physical environment $\mathcal{P}$ in the sense described in Section 2. The complexity of the $\mathcal{V}$ model can vary dynamically along different dimensions, for instance:
   - *temporal coupling*—from a static model completely built at one point in time, and later used solely to retrieve and affect the properties therein tracked, to a dynamic model, possibly built incrementally and cooperatively in real-time, accounting also for the run-time actions and interactions of the agents (either humans or not);
   - *semantics*—from a purely geometric description of the elements composing the environment and the visible agents populating it, up to richer structured descriptions encompassing roles, permissions, sensing/actuating allowances, and so on;
   - *level of detail*—to be intended not only w.r.t. the quality of the visual rendering of digital objects in the virtual/augmented scene, but also regarding how fine-grained perception and action capabilities of agents are, to which extent the properties and structure of the virtual environment are amenable of inspection, and so on.

   The main concern here, besides establishing the initial mapping from $\mathcal{P}$ to $\mathcal{V}$, is the maintenance of their *coupling*: any significant change endogeneous to the physical realm should be reflected in its virtual counterpart. Furthermore, sometimes this also holds for the opposite direction, in the case of decisions carried out in $\mathcal{V}$ that call for enactment in the physical world $\mathcal{P}$.

(2) *Extension* of the virtual environment $\mathcal{V}$ by means of a VR-based system that allows humans to explore and work inside $\mathcal{V}$ through smart tools enabling them to annotate, adapt, act upon, and extend the virtual environment according to their needs. This could be done either by an individual user (as in DTs) or by a team of human users working and cooperating simultaneously inside the virtual world (as in HAC), potentially supported by smart applications and tools such as personal assistant agents and decision support systems. The core idea is to provide humans with the possibility of building and affecting an *Extended Virtual Environment* $\mathcal{E}$, to be considered an enrichment of $\mathcal{V}$—i.e. in the sense of augmented virtuality or mixed reality. The main concern here is about defining those means of *perception*, *action*, and *interaction* available to the users, and those necessary to be aware of events and changes taking place in and affecting $\mathcal{E}$: in a digital, augmented world, in fact, users can be empowered with novel forms of action and enhanced sensing capabilities.

(3) *Augmentation* of $\mathcal{P}$ with $\mathcal{E}$, by means of an AR/MR-based system that makes it possible to realise the $\mathcal{V}$*i*TAL*i*SE vision: a form of *bi-directional augmentation*, so that virtual objects (i.e. holograms) are used to enrich / augment the physical world, and physical things are meant to enrich / interact with the virtual world in turn. By designing MR-based systems as Augmented Worlds [9, 10] based on the extended virtual model $\mathcal{E}$, humans can play inside the smart environment where the physical reality and world are augmented with the digital objects and functionalities defined in $\mathcal{E}$, to be potentially further extended, manipulated, and adapted. The main concern here is realising a *seamless integration* between the digital and physical realms with respect to the augmentation of human perception capabilities, means of action, and the way they interact.

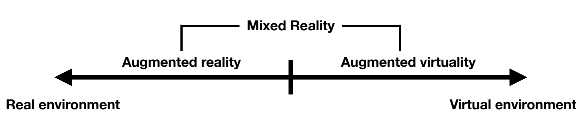

Figure 2: Milgram and Kishino's Reality-Virtuality continuum [28].

These three stages are repeated in loop – indeed, the name "$\mathcal{V}$*i*TA loop" – according to the need of the specific investigation / exploration / objectives of human users. This loop allows for repeatedly exploiting and applying *simulated action* in the virtual world and

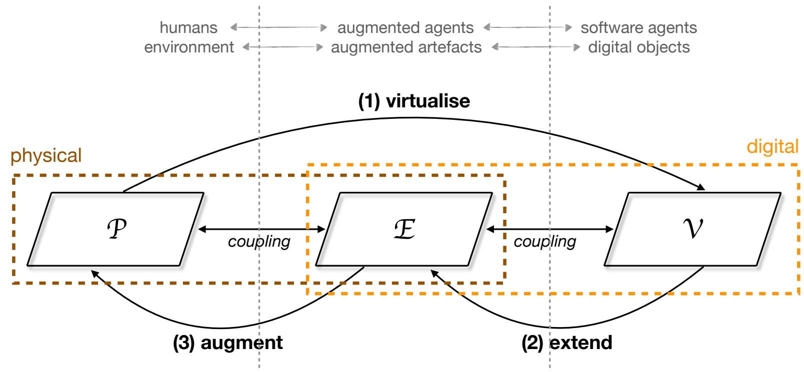

Figure 3: The $\mathcal{V}i$TA loop: by collecting data from the physical world, a virtual representation is built and later enriched by agents (either software or human) working therein. Then, the virtual environment is laid over the physical one, originating the augmentation which empowers users.

augmented action in the real world, enhanced to incrementally support the design and use of the smart environment, as well as the analysis of the work activities carried out by human users therein.

Let's discuss an exemplary scenario to clarify and substantiate what described above.

The Island of Stromboli, in Italy, is affected by most types of natural hazards, and experienced in 2002 severe damages along the entire coast as hit by tsunami waves. Despite the correct design of population emergency routes and meeting points being the only possibility to save lives, the evacuation plan conceived after the 2002 event has never been tested, neither virtually nor in reality.

Here, the possibility to share an extended VR environment among different users, inhabiting it at the same time to perform collaborative activities, represents an opportunity for "in-silico" training meant to support the specific activities of field operators, in a continuum between considerations and analyses carried out in a virtual augmented setting and their concretisation in the physical (augmented) world.

In such a scenario, the $\mathcal{V}i$TA loop may be actually instantiated as follows:

(1) data is collected from the physical environment $\mathcal{P}$, i.e. by using drones to capture images of the island to be later processed by suitable computer vision algorithms, with the aim of building the virtual representation $\mathcal{V}$
(2) field operators are trained in $\mathcal{V}$, that is, undergo learning sessions where they act in a *simulation* of the evacuation scenario, meant to purse a twofold goal—besides in-silico training in itself: on the one hand, collect data about *user behaviour* informing personal assistant agents and/or decision support systems, on the other hand, let users refine and extend $\mathcal{V}$ so as to build $\mathcal{E}$
(3) $\mathcal{E}$ is deployed as the augmentation layer of later "in-vivo" training sessions and simulations, where field operators act in the resulting Mixed Reality scenario, where guidance from personal assistant agents and decision support systems built during step §2 is available

By repeating these steps in loop, the co-evolution process enabled and supported by $\mathcal{V}i$TALiSE comes to life to promote a continuous improvement of the quality of simulated training sessions, on the one hand by incrementally refining both $\mathcal{V}$ and $\mathcal{E}$, and on the other hand by letting trained users learn to adapt to the ever changing contingencies of the actual environment.

4 ENABLING TECHNOLOGIES & CHALLENGES

The $\mathcal{V}i$TALiSE vision relies on a handful of technologies as key enablers for its realisation, many of which bring along challenges and issues to be still addressed. They are depicted in Figure 4, along with their role in the $\mathcal{V}i$TA loop, and described in the following paragraphs.

4.1 Computer Vision, VR, AR

Given the definition of AR in [2], one crucial step in AR is the ability to register virtual content in 3D with the underlying physical environment. To accomplish this, the pose (position and orientation) of the viewer with respect to some "anchor" in the real world must be determined. The real-time tracking of pose has been performed with different technologies: magnetic tracking, inertial tracking, GPS tracking and, recently, vision-based tracking. Although some robust solutions have been developed using markers (QR or other types of codes), other approaches are more innovative and fascinating, such as *natural feature tracking* (using unique visual features and discriminative "descriptors" such as SIFT [25], SURF [3], ORB [33], or FREAK [1] to match points in successive frames), *model-based tracking* [8] (where the tracking of real-world objects is achieved by using a known 3D structure, such as a CAD model), or *3D structure tracking* techniques (exploiting some sensors – such as MS Kinect – for estimating the 3D structure of the scene), as in the case of KinectFusion system [29].

Apart from the tracking task, computer vision techniques can be used within an AR application also for the *semantic instance segmentation* or *instance-aware semantic segmentation*. This task aims at recognizing a specific instance of an object (i.e., "chair_1" and "chair_2" not simply "chair") in the scene. This technique can bring a significant additional knowledge to the AR system and allows a

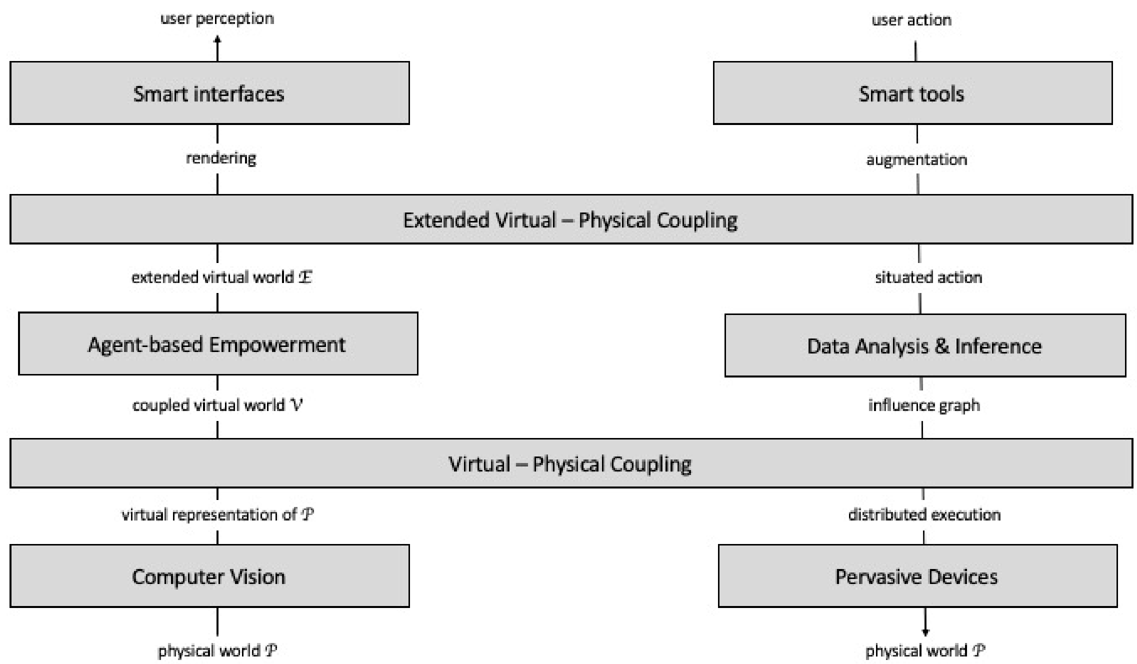

Figure 4: Key technologies and conceptual steps involved in the $\mathcal{V}$iTA loop: the physical world ($\mathcal{P}$) is progressively mapped and enriched up to an extended virtual representation ($\mathcal{E}$) thanks to Computer Vision and Agent-oriented technologies. There, human users may act and interact through smart interfaces (i.e. smart glasses) and tools (i.e. smart gloves), influencing both the digital realm (through automatic data analysis and inference processes) and the physical one (thanks to pervasive technologies suitably deployed).

finer interaction to the users. In [11] Multi-task Network Cascades consisting of three networks (respectively differentiating instances, estimating masks, and categorizing objects) are proposed. In [24] a fully Convolutional Neural Network (CNN) is designed for this purpose. Since semantic instance segmentation faces an inherent tension between semantics and location, deep feature hierarchies can handle encoding location and semantics in a nonlinear local-to-global pyramid.

4.2 Agents, MAS, and Decision Support

Crucial to the achievement of the kind of empowerment envisioned in $\mathcal{V}$iTAL*i*SE is the availability of proper models and technologies for engineering augmented smart environments at the right level of abstraction. One promising source of abstractions, methodologies, and technologies is the research area on Multi-Agent Systems (MAS) [12], which already proven to be suitable and particularly effective for a wide variety of goals and application scenarios, ranging from simulating pedestrians behaviour [39] to monitoring and control of power grids distribution [31].

Besides being fully aligned with the Human-Agent Collectives vision already described, an agent-based framework is particularly suited for the $\mathcal{V}$iTAL*i*SE vision because it captures perfectly the idea of an autonomous entity – the agent, indeed [19] – that has the ability to sense the environment and act upon it, and to communicate and cooperate with other agents, in order to fulfill its tasks and pursue its goals. The attention to the notions of action and perception is what made agent-based technology so successful in scenarios closely related to the ones here adopted as use cases, such as Smart Cities [7] and Internet of Things [13].

Agent-based models and technologies have also been proficiently applied for implementing (intelligent) personal assistant agents, then deployed in a number fo heterogeneous use cases [21, 34, 42]. Here, a pressing issue to be still fully dealt with concerns the concept of *adjustable autonomy*, that is, the capacity of software agents to dynamically vary their degree of autonomy, possibly (and temporarily) delegating *decision making* control to others, typically humans, when the current situation simply can't be handled properly—i.e. because of ethical issues, lack of information, malfunctioning, etc.

The issue of decision making calls for Decision Support Systems (DSS), a technology already extensively used in both the academia and the industry since the early 70s—in the form of expert systems. Encompassing a wide range of applications and actually implemented in many different forms – ranging from OLAP in database technologies to machine learning in AI – they are anyway meant to support humans by either augmenting their cognitive capabilities or lowering their cognitive workload—in the very same spirit of both personal assistant agents and smart environments. Recent research [38] in the field highlighted that moving from an individual setting to a team-oriented one, where decision support must consider possibly conflicting interests and also the communication dimension, poses significant challenges, such as recognising that when teams are distributed in an environment the communication paradigm changes, shifting to a more practical, implicit, task-oriented one.

4.3 Smart interfaces and tools

Humans working in a $\mathcal{V}$*i*TAL*i*SE smart environment are likely to send and receive commands exploiting natural interfaces, such as voice commands or gestures. Thus, natural language processing, speech recognition, and the aforementioned computer vision are essential to let users experience a more direct and *transparent* interaction with the digital environment [23]. Recent advances in machine learning are steadily improving conversational interfaces in the form of *chatbots*. Accordingly, many companies are specific IoT devices interacting via chatbots, such as Amazon Echo and Google Home. Currently, all is about single-party interaction, but in the near future multi-party interaction will become mandatory.

We did not talk extensively about the availability of VR / AR headsets: on the one hand, this is deliberate, due to the nature of a research paper. On the other hand, it is clear that there is a "technology pull" phenomenon, associated to the growing availability of cheap VR headsets (i.e. Oculus Go, or the announced Oculus Quest), that is not matched by something similar in the AR area. The presence of affordable devices of high technological quality is surely and enabling factor, but not necessarily a sufficient one: as recently stated by one of the co-founders of Oculus, "No existing or imminent VR hardware is good enough to go truly mainstream, even at a price of \$0.00."[3], suggesting that technology still needs improvements, but also that current applications and approaches probably do not represent such a compelling push towards adoption.

The current situation concerning the ad-hoc interfaces for VR devices, such as Oculus Rift, Go, and the foreseen Quest, considers the presence of ah-hoc controllers, connected wirelessly with the processing unit. These devices enable the usage of gestures, as an alternative or complement to vocal command. Human-computer interaction and usability research on this topic should, on one hand, lead to updated results in the evaluation of the effectiveness of this kind of input techniques [6], on the other new UI metaphors and radically new approaches could still be devised. It is still unclear if it is possible to design a UI in VR/AR systems able to enable users performing complicated tasks, e.g., of virtual / real environment annotation, including the positioning of markers or virtual objects that require some form of configuration.

4.4 Data analysis

Machine learning techniques are clearly involved in the pipeline that elaborates input data coming from either the physical world while building its virtual representation, as from the augmented world while analysing users' behaviours and outcomes of their actions. Steps such as data aggregation and fusion, feature extraction, clustering, and classification, are commonly employed in almost any pervasive computing system [30].

Even in the actuation stage, where actions from the augmented world are to be translated into the physical one, machine learning techniques may prove invaluable. For instance, robots may be given just the goal to achieve (i.e., reach a destination, or perform an action) without explicitly telling how to do so, and then let to *learn* how to act and *coordinate* with others so as to accomplish the task. Swarm robotics [5] may greatly help in this context.

[3]https://goo.gl/g9NZai

4.5 Serious Gaming

The level of maturity of VR technologies makes it possible to implement and deploy the $\mathcal{V}$*i*TAL*i*SE system with relatively contained costs and human efforts for the implementation of synthetic environments in which individuals are trained in an eLearning setting (see, i.e., the survey of [14]), participate serious games for being technically trained (i.e. within different contexts, from fire service [41] to healthcare [40]) or sensibilized with reference to potentially dangerous behaviours (i.e. children safety in road crossing [36]). The enabling technologies are growingly affordable, to the point that even modern personal devices like smartphones can represent perfectly adequate terminals for this kind of applications [37].

ARGO3D[4] and 3DTeLC[5] projects, for instance, successfully started to use VR technologies in the "real world" developing a series of virtual reality models for different natural hazard environments in both onshore and offshore environments and generating teaching tool kits using these data. In particular, users can navigate terrains via an Oculus or a VR mobile (i.e. Gear VR) headsets and are able to fly and walk over the terrain, allowing interaction with their surroundings. AR technologies suffered a setback due to the failure of the Google Glasses initiative and they are not as diffused in terms of consumer technologies, but they are consolidating as mature supports to professional field activities in the most diverse areas, from civil engineering [4], to neurosurgery [27], and technologies like head-up displays (HUD) are in the process of transitioning from the military sector to more mundane civil applications [22].

4.6 Human-Agent Collectives

A common element of the current type of applications is the fact that they are mostly focused on the activities of an individual user, and mostly they do not consider the potential scenario of applications in which different users are simultaneously connected to the same VR/AR or MR system. This kind of situation is relatively similar to a moment of the development of pervasive computing technologies in which devices were becoming affordable, the number of potentially interested users was growing, and the expected return-on-investment justified facing the increased complexity implied by the design and development of systems going beyond the level of personal application.

Within this framework, one significant challenge for $\mathcal{V}$*i*TAL*i*SE is to produce a model and a framework intrinsically targeting this class of social applications, in which VR and AR technologies are actually enabling the access to smart environments that now have more believable means of implementation and deployment.

Also, the nature, structure, role, and interactions means of software agents have to be discussed, as according to the $\mathcal{V}$*i*TAL*i*SE vision they are demanded to act in a mixed world where digital objects may dynamically acquire a physical embodiment, as well as agents' themselves may be dynamically deployed on a physical object (a robot) providing access to a whole array of sensor

[4]http://argo3d.unimib.it/

[5]https://sites.google.com/port.ac.uk/3dtelc/

and actuator devices that the agent should promptly learn to exploit. A starting point towards this research line could be that of re-thinking the sense-plan-act cycle of rational agents in the context of the 𝒱iTA loop, where an agent that was acting in a virtual only environment moments ago could be suddenly deployed in the physical world (either as an augmentation overlay or actually embodied in a robot device). Under this perspective, the agent should be able to seamlessly sense, act, and interact in the digital as well as in the physical realm.

4.7 User modelling

A common element of most MAS applications is that they still struggle to deal with "humans-in-the-loop" [17]: usually either they focus on software-only platforms where humans, if present, are merely external users of the system, playing no role whatsoever in it besides consuming services and exploiting functionalities, or they assign software agents as representatives of humans, constraining their admissible interactions with the system to those that software agents are capable of.

One significant challenge in 𝒱iTALiSE is thus that of considering humans as seamlessly integrated with the software substrate of the system – possibly aided by personal assistant agents meant to support decision-making – physically immersed in the Mirror World created by augmentation of the physical reality, thus capable of perceiving and affecting and realms: the virtual and the physical one. Similar things have been done, for instance, in the field of Web Services and Internet-scale collaboration [35].

5 CONCLUSIONS

The current state of development of virtual and augmented reality and of innovative user interfaces pushed by wearables devices allows for dreaming big about next generation smart environments supporting humans workload. To properly engineer such a sort of mixed reality workplaces, where a cohesive collective of human and software agents need to cooperate towards achievement of shared goals calls for a principled approach to the design of the infrastructure devoted at maintaining the coupling between the physical and the digital worlds. On purpose, in this paper we proposed the 𝒱iTALiSE vision and conceptual framework, featuring a structured loop – called 𝒱iTA – outlining the methodology to approach engineering of smart environments, focussed in particular on enabling co-evolution of agents – by learning from the simulated environment – and the environment itself—by adapting to agents' behaviours.

The multi- and trans-disciplinarity of the envisioned 𝒱iTALiSE approach calls for many different technologies to overcome their existing challenges and be seamlessly integrated together. Nevertheless, a unifying framework for the design of the system can and should be based on the agent abstraction, as it has been widely proven to be the most effective conceptual tooling to model both human behaviour and autonomous software agents, as well as to implement the latter.

With this vision paper, we hope to stimulate discussion on the important topic of next generation smart environments effectively mixing the digital and the physical, the virtual and the real, in a mixed reality continuum where humans are relieved of their cognitive efforts and augmented in their capabilities of carrying out their activities.

REFERENCES

[1] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. 2012. Freak: Fast retina keypoint. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. Ieee, 510–517.
[2] Ronald T Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* 6, 4 (1997), 355–385.
[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. 2006. Surf: Speeded up robust features. In *European conference on computer vision*. Springer, 404–417.
[4] Amir H Behzadan, Suyang Dong, and Vineet R Kamat. 2015. Augmented reality visualization: A review of civil infrastructure system applications. *Advanced Engineering Informatics* 29, 2 (2015), 252–267.
[5] Manuele Brambilla, Eliseo Ferrante, Mauro Birattari, and Marco Dorigo. 2013. Swarm robotics: a review from the swarm engineering perspective. *Swarm Intelligence* 7, 1 (2013), 1–41.
[6] Marcio C. Cabral, Carlos H. Morimoto, and Marcelo K. Zuffo. 2005. On the Usability of Gesture Interfaces in Virtual Reality Environments. In *Proceedings of the 2005 Latin American Conference on Human-computer Interaction (CLIHC '05)*. ACM, New York, NY, USA, 100–108. https://doi.org/10.1145/1111360.1111370
[7] Min Chen. 2013. Towards smart city: M2M communications with software agent intelligence. *Multimedia Tools and Applications* 67, 1 (2013), 167–178.
[8] Andrew I Comport, Éric Marchand, and François Chaumette. 2003. A real-time tracker for markerless augmented reality. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 36.
[9] Angelo Croatti and Alessandro Ricci. 2017. Mashing Up the Physical and Augmented Reality: The Web of Augmented Things Idea. In *Proceedings of the Eighth International Workshop on the Web of Things (WoT 2017)*. ACM, New York, NY, USA, 4–7. https://doi.org/10.1145/3199919.3199921
[10] Angelo Croatti and Alessandro Ricci. 2018. A Model and Platform for Building Agent-Based Pervasive Mixed Reality Systems. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Complexity: The PAAMS Collection*, Yves Demazeau, Bo An, Javier Bajo, and Antonio Fernández-Caballero (Eds.). Springer International Publishing, Cham, 127–139.
[11] Jifeng Dai, Kaiming He, and Jian Sun. 2016. Instance-aware semantic segmentation via multi-task network cascades. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3150–3158.
[12] Jacques Ferber and Gerhard Weiss. 1999. *Multi-agent systems: an introduction to distributed artificial intelligence*. Vol. 1. Addison-Wesley Reading.
[13] Giancarlo Fortino, Antonio Guerrieri, Wilma Russo, and Claudio Savaglio. 2014. Integration of agent-based and cloud computing for the smart objects-oriented IoT. In *Computer Supported Cooperative Work in Design (CSCWD), Proceedings of the 2014 IEEE 18th International Conference on*. IEEE, 493–498.
[14] Laura Freina and Michela Ott. 2015. A Literature Review on Immersive Virtual Reality in Education: State Of The Art and Perspectives. *eLearning & Software for Education* 1 (2015).
[15] David Gelernter. 1991. *Mirror Worlds or the Day Software Puts the Universe in a Shoebox: How Will It Happen and What It Will Mean*. Oxford University Press, Inc., New York, NY, USA.
[16] Michael Grieves. 2014. Digital twin: manufacturing excellence through virtual factory replication. *White paper* (2014).
[17] Jane Hillston, Jeremy Pitt, Martin Wirsing, and Franco Zambonelli. 2015. Collective adaptive systems: qualitative and quantitative modelling and analysis (Dagstuhl Seminar 14512). In *Dagstuhl Reports*, Vol. 4. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
[18] Thomas Holz, Abraham G. Campbell, Gregory M.P. O'Hare, John W. Stafford, Alan Martin, and Mauro Dragone. 2011. MiRA—Mixed Reality Agents. *International Journal of Human-Computer Studies* 69, 4 (2011), 251 – 268. https://doi.org/10.1016/j.ijhcs.2010.10.001
[19] Nicholas R Jennings. 2000. On agent-based software engineering. *Artificial intelligence* 117, 2 (2000), 277–296.
[20] N. R. Jennings, L. Moreau, D. Nicholson, S. Ramchurn, S. Roberts, T. Rodden, and A. Rogers. 2014. Human-agent Collectives. *Commun. ACM* 57, 12 (Nov. 2014), 80–88. https://doi.org/10.1145/2629559
[21] A. Jones, C. Moulin, J. Barthès, D. Lenne, A. Kendira, and T. Gidel. 2012. Personal assistant agents and multi-agent middleware for CSCW. In *Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. 72–79. https://doi.org/10.1109/CSCWD.2012.6221800
[22] Hyungil Kim, Joseph L Gabbard, Alexandre Miranda Anon, and Teruhisa Misu. 2018. Driver Behavior and Performance with Augmented Reality Pedestrian Collision Warning: An Outdoor User Study. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1515–1524.
[23] Matthias Kranz, Paul Holleis, and Albrecht Schmidt. 2010. Embedded interaction: Interacting with the internet of things. *IEEE internet computing* 14, 2 (2010),

46–53.
[24] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.
[25] David G Lowe. 1999. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, Vol. 2. Ieee, 1150–1157.
[26] Pattie Maes. 1994. Agents That Reduce Work and Information Overload. *Commun. ACM* 37, 7 (July 1994), 30–40. https://doi.org/10.1145/176789.176792
[27] Antonio Meola, Fabrizio Cutolo, Marina Carbone, Federico Cagnazzo, Mauro Ferrari, and Vincenzo Ferrari. 2017. Augmented reality in neurosurgery: a systematic review. *Neurosurgical review* 40, 4 (2017), 537–548.
[28] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.
[29] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. 2011. KinectFusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*. IEEE, 127–136.
[30] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos. 2014. Context Aware Computing for The Internet of Things: A Survey. *Communications Surveys Tutorials, IEEE* 16, 1 (First 2014), 414–454.
[31] Manisa Pipattanasomporn, Hassan Feroze, and Saifur Rahman. 2009. Multi-agent systems in a distributed smart grid: Design and implementation. In *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*. IEEE, 1–8.
[32] A. Ricci, M. Piunti, L. Tummolini, and C. Castelfranchi. 2015. The Mirror World: Preparing for Mixed-Reality Living. *IEEE Pervasive Computing* 14, 2 (Apr 2015), 60–63. https://doi.org/10.1109/MPRV.2015.44
[33] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. ORB: An efficient alternative to SIFT or SURF. In *Computer Vision (ICCV), 2011 IEEE international conference on*. IEEE, 2564–2571.
[34] J. Santos, J. J. P. C. Rodrigues, J. Casal, K. Saleem, and V. Denisov. 2018. Intelligent Personal Assistants Based on Internet of Things Approaches. *IEEE Systems Journal* 12, 2 (June 2018), 1793–1802. https://doi.org/10.1109/JSYST.2016.2555292
[35] Daniel Schall, Hong-Linh Truong, and Schahram Dustdar. 2008. Unifying human and software services in web-scale collaborations. *IEEE Internet Computing* 12, 3 (2008).
[36] David C Schwebel, Tabitha Combs, Daniel Rodriguez, Joan Severson, and Virginia Sisiopiku. 2016. Community-based pedestrian safety training in virtual reality: A pragmatic trial. *Accident Analysis & Prevention* 86 (2016), 9–15.
[37] David C Schwebel, Joan Severson, and Yefei He. 2017. Using smartphone technology to deliver a virtual pedestrian environment: usability and validation. *Virtual reality* 21, 3 (2017), 145–152.
[38] J.P. Shim, Merrill Warkentin, James F. Courtney, Daniel J. Power, Ramesh Sharda, and Christer Carlsson. 2002. Past, present, and future of decision support technology. *Decision Support Systems* 33, 2 (2002), 111 – 126. https://doi.org/10.1016/S0167-9236(01)00139-7 Decision Support System: Directions for the Nest Decade.
[39] Giuseppe Vizzari, Lorenza Manenti, and Luca Crociani. 2013. Adaptive pedestrian behaviour for the preservation of group cohesion. *Complex Adaptive Systems Modeling* 1, 1 (2013), 7.
[40] Ryan Wang, Samuel DeMaria Jr, Andrew Goldberg, and Daniel Katz. 2016. A systematic review of serious games in training health care professionals. *Simulation in Healthcare* 11, 1 (2016), 41–51.
[41] F Michael Williams-Bell, B Kapralos, A Hogue, BM Murphy, and EJ Weckman. 2015. Using serious games and virtual simulation for training in the fire service: a review. *Fire Technology* 51, 3 (2015), 553–584.
[42] Shumin Wu, Hamada Ghenniwa, Yue Zhang, and Weiming Shen. 2006. Personal assistant agents for collaborative design environments. *Computers in Industry* 57, 8 (2006), 732 – 739. https://doi.org/10.1016/j.compind.2006.04.010 Collaborative Environments for Concurrent Engineering Special Issue.