Graphics and Media Technologies for Operators in Industry 4.0

Challenges and examples

Jorge Posada
GraphicsMedia.net

Mikel Zorrilla, Ana Domínguez, Bruno Simoes
Vicomtech

Peter Eisert
Fraunhofer HHI

Didier Stricker, Jason Rambach
DFKI

Jürgen Döllner
HPI

Miguel Guevara
CCG

Editor:
Andre Stork
andre.stork@igd.fraunhofer.de

Visual Computing technologies have an important role in manufacturing and production, particularly in new Industry 4.0 scenarios with intelligent machines, human-robot collaboration and learning factories. In this manuscript, we explore challenges and examples on how the fusion of Graphics, Vision and Media Technologies can enhance the role of Operators in this new context.

Recent developments in ICT and the latest Internet-related technologies are opening up revolutionary possibilities for manufacturing and production. The technology surrounding Industry 4.0 and Smart Factories is mainly focused in certain fields, such as Internet of Things (IoT), Industrial Big Data, Industrial Automation, Cybersecurity or Intelligent Robotics. Nevertheless, it is important to emphasize the relevance of Computer Graphics and Media technologies, under the umbrella of Visual Computing [1], as key enabling technologies in Smart Factories of the future [2].

In particular, this paper focuses on the role of the Operator in the Smart Factory, which plays a crucial role within the *Vertical integration dimension* in Industry 4.0. Smart Factories change the role of the operators, making possible new types of interaction between them and the machines. In this context, human-centricity becomes a key concept, fostering the cooperation of machines and humans towards a more efficient and effective human-automation symbiosis [3]. This concept is related to the Operator 4.0 term mentioned in [4].

In this context, it is interesting to notice that many authors explicitly identify Augmented Reality as one of the main Industry 4.0 technologies to empower the Operator [5]. However, being Augmented Reality an important aspect with its own challenges and possibilities, we believe that a broader combination of Computer Graphics, Vision and Media technologies can be very important to support Operators and to learn from their actions in new Smart Factory scenarios. In addition, the support from knowledge-based and intelligent systems is critical in this context [6].

We will present a set of challenges related with the Operator in the context of Smart Factories, and some current examples of relevant related R&D activities in this field.

IDENTIFYING THE CHALLENGES FOR OPERATOR

Figure 1 depicts a diagram of the role of the operators in Smart Factories, where three major challenges can be identified:

Figure 1. A diagram that represents the challenges of the operators within Smart Factories

The first challenge is to **support operators to perform the tasks** within a process / workflow. In this context Augmented Reality plays a crucial role with improved experiences, projecting instructions with the next steps of the work flow onto objects and tools (see No. 1 in Fig. 1), or enabling projection-based gestural interaction (No. 2 in Fig. 1). Moreover, AR technologies and devices can overcome the limitations of the operators with disabilities or special needs (No. 3 in Fig. 1), as well as Virtual Reality is a relevant enabling technology for training activities.

The second challenge is to **support operators' understanding and decision-making** in all the activities regarding the whole production process, to provide an adequate ecosystem to perform their tasks in an optimal way. More versatile and enhanced Human-Machine Interfaces (HMI) are needed, following a human-centred approach and facilitating the analytical reasoning through interactive multi-screen interfaces with Visual Analytics. It needs to address the different screens being used by operators, such as big shared screens in the premises (No. 4 in Fig. 1), specific HMIs of machines (No. 5 in Fig. 1) or personal devices (like smartphones or tablets). All these screens must provide a consistent overview of data and information in form of graphs or media presentations (No. 6 in Fig. 1), personalising the visualisation for the role of each one of the operators and enabling ubiquity, addressing the mobility of the operators from one set of screens to another.

The third challenge is to **learn from the activity of the operators** in order to be able to predict specific situations, optimise the process and better organise the Smart Factory. In this context, video managing capabilities (recording, streaming, etc.) are fully required with a complete integration with the Manufacturing Execution System (MES) (No. 7 in Fig. 1). On the one hand, the video management will enable real-time supervision of the tasks both from inside or outside the premises, while providing an enriched reporting of the production process. On the other hand, artificial intelligence techniques such as pattern recognition and machine (deep) learning make possible to learn from the tasks of the operators and influence in the process (No. 8 in Fig. 1).

The following table summarises the three major challenges, provides a detailed analysis on how the major challenges can be divided into smaller particular challenges, and shows how Graphics and Media Technologies can support overcoming those challenges.

Major Challenge	Key Graphics & Media Technology	Particular challenges	Enabling technologies
Train operators and support them to perform tasks	Improved Augmented Reality experiences for Operator	Access through heterogeneous devices	Multiplatform & adaptive authoring tools
		Complicated authoring for ad-hoc cases	Video learning from experience, automatic segmentation/edition
		Difficult interaction and device dependant (clicks on glasses, air-tap gestures, etc.)	Gesture-based touchless interaction in projection space
		Better integration in real-world operator workspaces	Projection based augmented reality
		Decoupling & misalignment of real/virtual models	Precise tracking (3D model- based, others)
		Delays & batteries	Video/graphics compression & streaming optimization for client-server AR
		Human robot interaction	Gesture interaction, augmented reality
Support operators understanding and decision-making	Advanced Visual Analytics and Information Visualisation	Industrial IoT Big Data Analysis from Sensors	Ultra-High Frequency Data Visualization & Analytics (e.g. 200 variables, each at 1kHz)
		Ubiquitous and adaptive multi-screen User Interfaces	Multiscreen video and content adaptation (from media broadcast advances)
	Digital Twins	Functional simulation of digital twin models. Real-time visualization of Digital Twins	Simulation results fusion with 3D CAD/CAM Model AR on real machine/workpiece HPC and GPGPU approaches for near real-time simulations Web 3D based interaction
Learn from the activity of operators	Video-based knowledge capture and management of Operators and Production ecosystem	Capture and manage operator tasks and contexts in video. Integration with Manufacturing Execution Systems (MES) is a must	Automatic video segmentation Bitstream adaptive video storage and transmission Semi-automatic annotation tools
		Learning Operator skills for producing intelligent training systems for operators	Video / graphics-based intelligent learning based on actual operation Vision based deep learning techniques for classification of repetitive tasks Motion tracking for 3D reconstruction of actions
		Traceability proof of production procedures	Connected and adaptable video annotation/segmentation for live or pre-processed video channel for customers

HOW GRAPHICS AND MEDIA TECHNOLOGIES SUPPORT OPERATORS IN INDUSTRY 4.0

Challenge 1: Train operators and support them to perform tasks

Augmented Reality Assistance for Human-Machine Interaction and Cooperation

Automation in manufacturing is continuously increasing. Robots with their fast and precise motion can support humans, especially for monotonous, tiring, or dangerous operations, while humans are still essential due to their flexibility, fine motor skills, and intelligence. Currently, most robots work spatially separated from workers, but it is expected that the collaboration of humans and robots in shared working spaces will increase productivity in many complex processes. However, such concepts require entirely new solutions for human-machine interaction and collaboration, as well as safety. In this context, Visual Computing provides many solutions for augmenting information (e.g. about robot intentions), working area surveillance, and machine control (e.g. touchless via gestures) in order to provide an assistance tool to the human.

Existing R&D activities are targeting the creation of new assistance tools for human-machine interaction and collaboration using 3D scene sensing, gesture control, and projection based augmented reality visualisation[1]. Application fields are for instance the manufacturing of cars and electronic parts collaboratively between robots and humans. A context sensitive spatial presentation of additional information and visualizations of three-dimensional data augmented into the real scene is supporting the user and allows the machine to communicate its status. Also, security areas depicting the robot's operating range and motion path are projected into the scene while human workers are tracked three-dimensionally to avoid potential collisions.

For data visualisation, AR-glasses and holographic HMD are used in many research projects (and some actual industrial implementations), combined with computer vision, in order to perform tasks such quality control and error detection of production, and collaborative scenarios to monitor and discuss the production plans and execution, as for example in the Intelligent4DMould project for the Portuguese Foundry sector.

In contrast with this approach, projection-based augmented reality [7] is exploited that provides seamless integration of information in the operator's place. In comparison to AR glasses, which still do not provide ergonomics and quality for a permanent use, the projection of graphical data into the real scene via projectors enables hand-free working without any additional worn devices. Especially for long use, discomfort due to motion sickness or vergence-accommodation conflicts is also eliminated. The additional information is projected directly in the operation space, avoiding these known disturbing effects in HMD and AR glasses. Due to the benefits, projection-based AR can significantly enhance interaction. This can be used for example to visualize welding seams that need to be checked by a worker, or to display virtual user interfaces for touchless machine control.

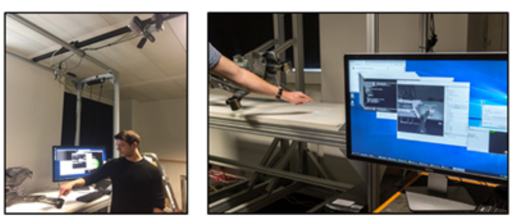

Figure 2. Projector-camera system for enhanced human machine interaction. Information and virtual user interfaces are precisely registered to the working place or even moving objects.

Projector-camera systems can be used for data augmentation. Cameras allow for an automatic calibration and 3D measurement of the working area. In addition, continuous re-calibration can be performed by analysing the known projected information from the slightly different viewing position of the camera [8]. Projection is supported for both static working places as well as parts or objects being moved by a robot. For a precise registration of projected information to the object's surface, 3D tracking of parts is essential, which is realised by model-based tracking [9] that registers CAD model data with the camera image. Since manufactured objects usually show untextured, highly reflective surfaces, additional information is extracted from object contours or the projected information itself, similarly to [8]. This enables precise visualization of information directly onto the object's surface simplifying the information exchange between humans and machines.

[1] EasyCohmo R&D project, funded by the German Federal Ministry of Education and Research, grant no. 03ZZ0443E

In order to provide feedback to the machine, e.g. to document working processes or to control the machine, the same system is used to project user interfaces into the working area as illustrated in Fig. 3. Gesture analysis [10] enables touchless selection of buttons or even fine motion control. Special care is taken in the definition of gestures in order to become robust against individual variations in the performance. Due to the projection of information and feedback interfaces directly into the operating area of the worker, context switching and body motion is reduced making manufacturing processes of the future more efficient.

In summary, what is happening is that the objects (pieces, machines, tools), and the operation environment, are now able to *communicate visually* in a direct way with the operator -and vice versa-, without the need of added screens and devices, in a seamless and natural way; enhancing and augmenting the production explicit knowledge, and allowing a smarter working space.

Figure 3. Interaction in a collaborative human-robot inspection by projector-based AR. Left: Touchless interaction with a projective interface. Right: Gesture based object interaction with a welding line marked by the projector.

Skills-driven Visual Interaction for Operators in Augmented Shop floors

Designing solutions that take advantage of human workers in their processes, empowered by collaborative augmented intelligent systems, reap benefits to the automation of manufacturing processes while maintaining the flexibility presented by their human components. However, these hybrid manufacturing solutions must address the inherent variability of their human element in terms of physical and cognitive capabilities, previous expertise, task adaptation and human aspects of technology acceptance. Not meeting this need will represent an important societal problem; the forthcoming cutting edge intelligent and flexible production may develop in such a way that it could be detrimental to individuals, organisations and societies.

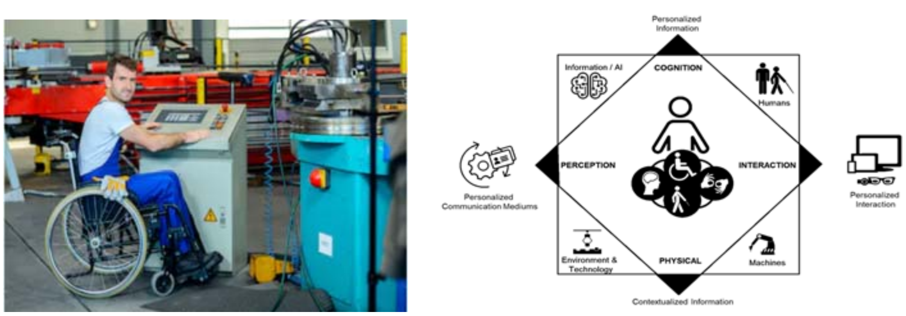

Figure 4. Augmented operators require a skill-driven visual interaction model.

For instance, there is a renewed interest in the use of new devices and technologies for Augmented and Virtual Reality in production and maintenance tasks. Or in new ways of gesture or voice interaction with the production system, or on visualization and interaction on mobile devices. In addition, human-robot interaction has novel important visual components, relevant to both the operator and the robot.

However, in many of these approaches there is an excessive simplification of the Operator profile and skills. Current research is addressing the *challenge to adapt the visual interaction technologies* of the augmented shop-floor to the *actual profile of the individual operator*, taking into account his/her physical and cognitive conditions, interaction preferences, previous experience and work attitudes. This adaptation has to consider also the *task automation level (human/computer)* as described in [11], ranging from manual control to full automation, with several levels in between (e.g. blended decision making). Thus, the new interaction models should be *skill-driven*, in the sense that these models take explicitly the skills of the operator to define the visual interaction paradigms.

In this context, it is critical to investigate new methods to harness this technological shift towards the creation of more flexible working environments, setting as objectives the enhancement of the human cognition, alongside with a boost in worker inclusiveness. In this research field, the augmentation of shop-floors (through AR, as in [12]) is explored in distinct task roles perspectives. Also, a current international research line is particularly addressing the role of designing new human-automation workspaces, i.e., how visual interaction is defined and information is communicated and assimilated by different workers, with an emphasis on Operators with disabilities or in advanced age groups. As a notorious example of this need, there is research work for augmented shop floors with associations of disabled operators that play a relevant role in certain assembly domains (e.g. electronics), as for instance Lantegi Batuak (2500 employees), and several others in Europe, in order to balance human and automation levels in the manufacturing workplaces of the future[2]. Augmented booths for assembly operations are defined in such a way that they allow the definition of user-friendly techniques to facilitate the adaptation of the worker experience to different interaction layouts.

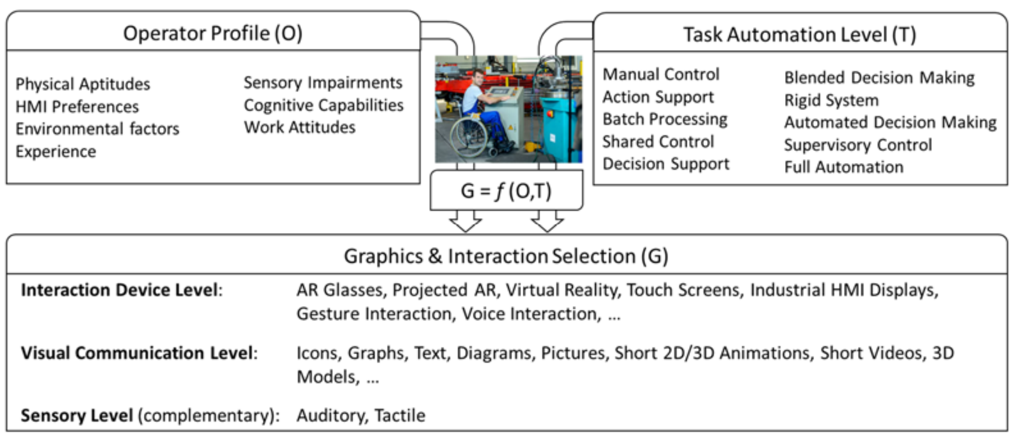

Figure 5. Skill-drive selection of Visual Interaction

In the view of this model, workers with partial blindness, for example, can have content and interaction automatically adapted towards aural senses with varying degrees depending on the properties of the workspace. Another advantage can be perceived in tasks where worker hands are used to hold/perform the assembly of specific components (i.e. hands are not available for interaction with visual interfaces) or are constrained by physical disabilities. The framework can provide interaction modalities like eye-gaze, voice control or interaction with projected content.

Challenge 2: Support operators understanding and decision-making

There are several Computer Graphics technologies that are useful to support operators understanding and decision making. We focus here in two lines of work not sufficiently explored so far in the production context, and that show promising research lines for the future: 3D Point Clouds as a source for Digital Twins supporting operators, and Dynamic ubiquitous multiscreen Dashboards.

---

2 MANUWORK project, with the participation of Vicomtech and CEA-List, has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no° 723711.

Digital Twins based on 3D Point Clouds

3D Point Clouds acquisition and processing is central to multiple applications in several sectors, as for instance in Geographic Information Systems, Cultural Heritage reconstruction, Quality Control for industrial inspection and so on. 3D point clouds represent a generic approach to capture, model, analyse, and visualize digital twins used by operators in Industry 4.0 application scenarios. The management, analysis, and visualization of massive 3D Point Cloud data can provide general purpose services that support digital twins to be used by operators in Industry 4.0 application scenarios [13]. Due to the generality and cost-efficiency of 3D Point Cloud software and hardware technology, this technology provides a generic approach to 3D twin modelling. Based on up-to-date, highly. accurate digital twins, visual analytics tools can be implemented supporting operators in decision making and understanding.

There are several aspects that this kind of solutions would need to provide, in order to support 3D Point Cloud-based digital twins in the context of operator support. Current international research[3] is focusing in some of them.

1. Data management and continuous integration of 3D Point Clouds over time, leading to a "4D" point cloud model. Only with a temporal approach, operators get up-to-date models and precise information about the past of the artefacts being represented.

2. Real-time, web-based visualization of massive 3D Point Clouds and integrated views of 3D Point Clouds and 3D geometric reference models. Due to the massive data size, web-based viewing technique need to use efficient spatial data structures for view-dependent level-of-detail 3D Point Cloud representation and related data streaming. Complementary to LOD, semantic classification of Point Clouds allows for thinning 3D Point Clouds by skipping 3D points of irrelevant categories [14].

3. Semantics enrichment of 3D Point Clouds based on heuristics and deep-learning based approaches, deriving 3D objects and attribute data by interpreting the 3D point cloud. In particular, deep learning allows us to semantically interpret and segregate 3D Point Clouds to build meaningful digital twins that concentrate on relevant parts and features.

4. Interaction and collaboration among operators using 3D Point Cloud-based digital twins as base communication artefacts, for example, with support of highlighting in 3D Point Clouds, annotating 3D Point Cloud parts and derived objects. As 3D Point Cloud-based digital twins do not lend themselves to be physically transferred to web clients due to their massive size, a server-based, centralized dissemination can be used together with a link-based encapsulation of settings for views and annotations, which enables operators to collaboratively work on digital twins in standard communication infrastructures.

5. Automated analysis such as for change detection or comparing as-is versus should-be states of artefacts represented by digital twins. For decision making and support, analytics functionality is essential. As a both generic and fundamental operation, we identified 3D point cloud difference analysis, which can be accelerated by a GPU-affine algorithm.

[3] The PointTube project from HPI is a good example, a service platform for 3D Point Clouds

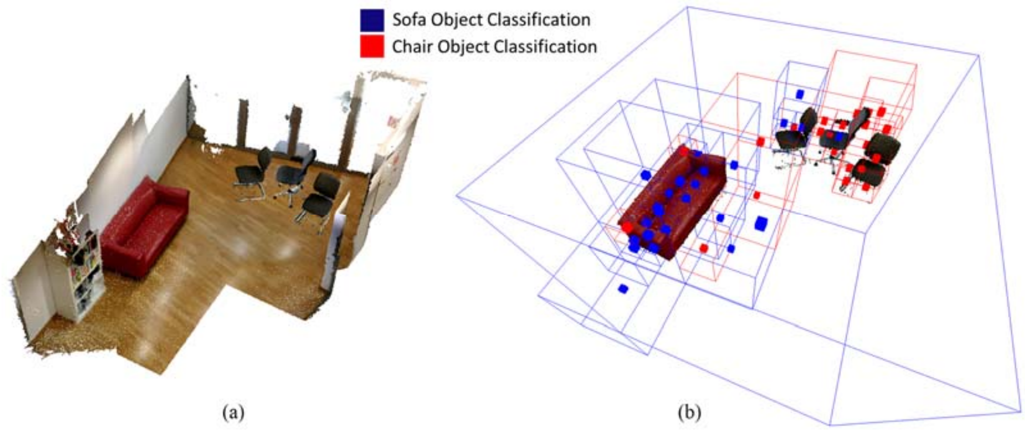

Figure 6. Classification results of an indoor office area containing a sofa and chairs [13]. (a) Shows the input Point Cloud and the segmented region that is to be classified, and (b) shows the classification output.

6. Augmented Reality approaches allow for an immersive access to digital twins. Here, operators can get a near-realistic spatial impressions. If the view is combined with sensor data and, more general, sensor analytics results, a tight integration of digital twin and related data about its use, status, and history can be implemented within a single user-interface.

Dynamic and ubiquitous multi-screen dashboards

Traditional industrial HMIs are very "static", based on old interaction paradigms (WIMP). They include basically some alarms, basic production output data and production planning, and some sensors information. Most information offered to operators and engineers is based on static text, pictures in the best cases, and specially data (diagrams, statistics, etc), shown in external screens attached to the machines, in PC stations close to production, or on normal displays in the shop floor. Production dashboards are a very clear case for this. In order to have better knowledge of the real production environment, a very powerful tool, not sufficiently used in state of the art industrial practice, is to capture and use the relevant events in video. It is noteworthy that there is very few adoption of video recording, management, search, etc. in real-world factories.

There is a recent trend to include mobile devices for dashboards and other functionalities to some extent, but still the approach is based on expensive rugged hardware that runs proprietary software solutions with single screen paradigms, highly rigid and fixed distribution of information in the screen, and few interaction possibilities.

In this context, we see an interesting opportunity to migrate recent technology advances from other areas with intensive interaction based on dynamic, mobile, multiple screens towards Smart Factory scenarios. Very specifically, there is a field that apparently is far away from the context, but that is quite close in terms of user needs and specifications: we refer to the research advances in the *dynamic adaptation of multiple screen visualization and interaction for professional TV experiences*, and its synergies with other media (Internet, etc).

Operator and Engineers needs	TV spectator needs	Common approach
Ubiquitous production information, in multi-screen format	Ubiquitous media experience in TV screen, mobile devices, tablet, PCs [15]	Web based, HTML-5 and Web3D compliant. Fully mobile-compatible.
Harmonisation of heterogeneous data sources (production plan, sensors, CAD, ERP, MES…)	Harmonisation of TV broadcast signal with various Internet based information sources	Data fusion based on standards (e.g. OPC-UA for industrial production, HbbTV for Hybrid TV experiences)
Dynamic synchronisation of sensor data, product design and production planning, even video streams	Dynamic synchronisation of social network inputs, internet updates, multiple cameras input	Dynamic time-based alignment and synchronisation of data sources. Data Compression, streaming techniques and sampling.

Personalized experience, relevant to the particular Operator profile and Task requirements.	Personalized experience, relevant to user tastes and interests, interaction preferences, devices	Dynamic Layout engine, able to produce combined multi-screen fusion and interactive visualization based on user profile [16]. Optimal distribution of content between available screens.
Automatic localised information according to specific machine, location in plant	Automatic localised information according to environment (e.g. in a car, in a TV proximity, in a mobile)	Detection of location and adaptation of available screens. Information flow between screens according to user needs and location

A current research example involves the development of a core dashboard system for an industrial, export-oriented SME that builds tools for heavy lift operations and that owns a modular, portable and monitored Test Bench to perform Load Tests up to 1,000 ton[4]. This is a typical industrial case: An engineer controlling the load test through a specific hydraulic system, with its own HMI, while the outcome information of the test is displayed through different interfaces in machine-specific displays. A team of distributed operators is visually controlling the load test and following the different displays, interacting with the engineer to fine-tune the test. As the test are run in international locations, the engineer has to physically travel to the sites.

This is a case of *horizontal integration* of Industry 4.0, in which emerging internet-based systems allow a deeper integration between industrial clients and providers, empowering Operators and Engineers to work collaboratively even in remote international locations, and avoiding the need of costly physical travel and ad-hoc adaptations of user interfaces. In this context, a more versatile and smart solution was desired, providing:

1. A ubiquitous, adaptable and very flexible multi-screen dashboard to monitor multiple camera views and the outcome data in real time, in order to have all the information when controlling the load test through the hydraulic system.
2. A live adaptable and flexible multi-screen visualisation solution for engineers around the world to follow the load test remotely, as they were on the premises, combining graphics and media with all the relevant information in an interactive and customisable multi-device interface,
3. An on-demand inspection solution of the Load Test, personalising the view and verifying all the process and the adequate performance of the test, and the generation of an added value advanced and interactive report as a communication tool for engineers to learn more about the tools that are being measured.

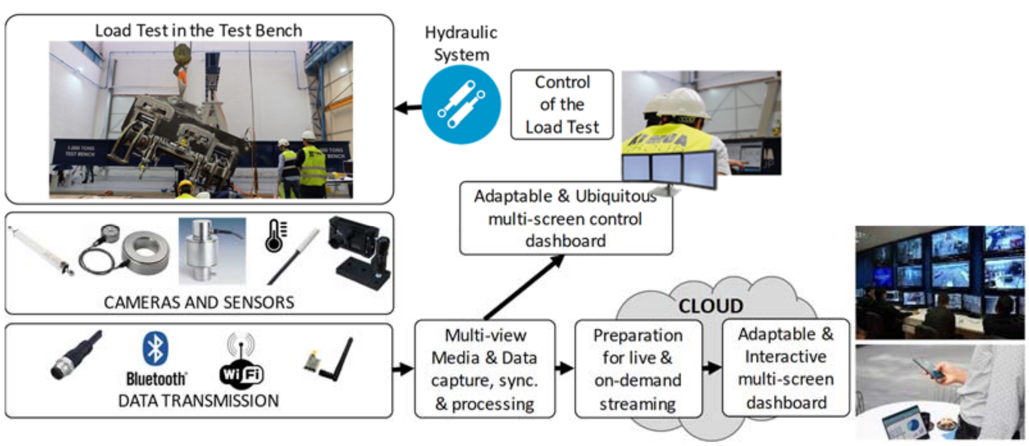

Figure 7. Dynamic Multi-screen dashboards for Operators in Industry 4.0

It provides an adaptable & ubiquitous multi-screen control dashboard, where different video sources and heterogeneous outcome data are captured and prepared to be exposed through a Web-based HTML-5 application. All information is shown in any kind of Web-based device, such as a laptop, a tablet or a smartphone. The pieces of information and User Interface layout

[4] Kimua Group's Load Test services, where PLATMON 360 research project has provided multi-screen dashboards pilot: https://www.kimuagroup.com/en/services/#load-tests..

is dynamically configured according to the target device and enables the interaction of the operator. Different devices can be easily associated, such as multiple laptops and mobiles being used by the same operator, and the User Interface dynamically adapts to that multi-screen configuration, showing parts of the application on each screen, and enabling the operator to move components from one device to another. Moreover, the operator can move from one place to another in the premises, dissociating big screens and having all the relevant information in the mobile, while associating new big screens in the new location.

Information flow between screens with drag-and-drop is possible, but especially automatic machine location detection and user profile adaptation brings a more intelligent, physically distributed and flexible user interface. Current deployment is well accepted by operators and engineers as it opens more efficient ways to get information and communicate the tests.

It is also important to take into account the integration of these dashboards, including video streams and multimedia content, in ERP (Enterprise Resource Planning) and/or MES (Manufacturing Execution Systems), to avoid isolated solutions not connected with the information systems of the factories; another research example is working in integrating this media capabilities in a commercial MES Software[5].

Challenge 3: Learn from the activity of operators

AI-enabled Learning from the actual activities of Operators, including their individual tasks' performance and reactions to the environment, is a cornerstone to enable Industry 4.0 new scenarios. There has been always a concern about the capture and transmission of Operator knowledge in the factory. One of the already widely used Computer Graphics technologies to support operator training is the use of Virtual Reality training systems, and slowly even traditional Augmented Reality is gaining acceptance (based in pre-defined sequences with graphic instructions overlaid on camera input). However, these systems mainly focus on training new procedures, designed by a specialised author in a computer system, and in general do not capture the reality of the real-world performance to improve the training. Besides, learning from Operators activity is not only useful for training purposes; it can help to dynamically rethink and re-shape the operator tasks, and even provide valuable input for intelligent and flexible production systems in the Smart Factory, for instance in scenarios where robots adapt to human behaviour and performance.

We present in this section an example of how Artificial Intelligence and Computer Vision technologies can be used to automatically learn from operators' activities, and how it is possible to use this acquired knowledge to configure Augmented Reality supported assistance systems in challenging assembly scenarios.

Video-based Deep Learning for Operator Assembly tasks

Maintenance of complex machines is a task usually performed by skilled employees in a company. This knowledge however is often neither documented nor accessible in digital form. It is therefore desirable to transfer human skill to an artificial intelligence that could help the human operators -or even replace them- when performing specific tasks. Skill transfer from human to machine has been the focus of research for long time, but real progress can be attained with the advent of modern learning technique such as Deep Learning.

Operator Assembly tasks are particular to each product; however, human operators are able to adapt easily to completely different assembly setups. This human capability is directly related with the ability to learn (by imitation and training) different assembly steps in the production line, that even when they differ from one setup to another. For machines, this acquisition of knowledge and the application of it in the guidance of an Assembly task is a challenging problem. Providing a holistic approach to it allows to alleviate the work effort of content creation and handcrafted tracking solutions in current AR assistance systems.

[5] Vicomtech also works for IDS for the MES integration of video capabilities in the MEMORI project.

An assembly, maintenance or operation task can be learned by the system from an example video of a human performing the task. In [17] the system automatically segments the video into a set of actions using an unsupervised method. The specific single steps for the procedure are extracted through the analysis of these segments. Using a hand-tracking approach, the actions are classified, and a training set is generated by augmenting the video data with additional synthetic hand gestures to generalize the learned model. A similar approach could use Video-based Deep Learning to automatically recognize elementary tasks. This can be done for a global dataset of human actions first, in order to have a first action classifier, followed by a second, task-specific training, which would be able to learn the succession of tasks and automatically suggest the logical next step according to the use case at hand. Once elementary actions are recognized, and predicted, from the extracted video snippets, Augmented Reality visualisations to guide the repetition of the particular or of similar tasks can be created and overlaid in real-time following the identification of the current status by the trained model.

To ensure the reliability and error tolerance of such an AR assembly application, precise tracking of the position and orientation of the involved objects is required, along with recognition of the current object state at a level of different objects components. Deep Learning techniques utilizing the potential of Convolutional Neural Networks (CNN) are applied to identify objects at different states and to directly estimate their pose [18]. Successful training of such systems requires large amounts of data with large variation in pose and lighting. Since capturing real-data is often not scalable, synthetic training datasets are conveniently created using CAD models of industrial objects. The problem of transferring models learned on synthetic data to the real world is then confronted but can be mitigated by applying proper domain adaptation techniques on the training.

The computational load of tracking and rendering AR augmentations in real-time can sometimes be overwhelming for mobile devices such as AR glasses or some smartphones especially over an extended time. Therefore, an edge computing architecture that transmits the video stream to a server for processing and transmitting back with annotated can prove very beneficial to the overall system latency and user experience given a fast, reliable wireless network as presented in [19].

Augmented Reality is a powerful Media technology already used in various contexts. The use for Industry 4.0 represents a higher challenge, since industrial applications place higher demands on precision, accuracy and robustness. The ultimate goal for an AR system in an industrial setting is the full knowledge of the environment at any time at a structural as well as semantic level. An interplay of traditional geometrical Computer Vision SLAM systems with Deep Learning methods highly fit for segmentation and classification tasks may lead to a complete environmental mapping of known objects and unknown areas enabling the automation of interactions and seamless, immersive AR integration.

Figure 8. The assembly task learning pipeline and live operator support application screenshots.

CONCLUSIONS

Computer Graphics, and other related vision and media technologies, have a lot to contribute to new scenarios in the Smart Factories under the Industry 4.0 concept. From all the different dimensions, this article focuses in the role of the new Operator and how the fusion of Graphics, Vision and Media Technologies can enhance their role in this new context.

In particular, this paper addresses three main key challenges having the operators in the middle of the equation: a) how to train operators and support them to perform tasks, where the improvement of Augmented Reality experiences is crucial, b) how to support operators understanding and decision-making, thanks to advanced Digital Twins from 3D Point Clouds, Visual Analytics and multi-screen interactive dashboards, and c) how to learn from the activity of operators, with video-based monitoring, management and knowledge of operators and production ecosystem. We have also shown some relevant current research activities, as well as technical and scientific open questions related to them.

An Industry 4.0 that is really focused in enhancing the role of the operator requires a seamless fusion of the cyber and physical dimensions. This counts not only for the automation inherent to cyber-physical production systems (also known as CPPS), but also, and especially, for the integration of the Human Operators, with all their diverse physical and mental capabilities, in the configuration of the new smart factories, including (but not only) new ways of human-robot collaboration.

There are important open research issues to make this happen. Augmented Reality has been around since early 90's, and still is not a widespread industrial practice (although we are living now a strong renaissance due to major improvements in costs, hardware, software, and usability). However, ergonomic aspects are still a major obstacle, and the challenge of a transparent and seamless augmentation of information in physical space is still there. Projection based AR is just a step forward but not the final solution.

HMI Usability and adaptation to various Operator profiles is also a challenge: mainstream research on graphics are not focusing on the actual diversity from operators in real world. Real solutions for elderly or operators with disabilities are still far away from AR devices, gesture interaction and so on. We have shown some advances in this line but there is still a lot to do.

Automatic acquisition of 3D models of production reality (machines, environment, even humans and robots), with the purpose of generating dynamic digital twins is still an unexplored area with challenges in the processing in massive Point Clouds. On the other hand, ubiquitous multiscreen dashboards are quite feasible, but requires that research from other domains can find a stronger application pull from industrial operator needs in the Smart Factory.

Finally, the ability to learn is intrinsic to intelligence, and Artificial Intelligence approaches to capture the knowledge and learn from operators' activities, based on video recordings, is a major enabler for more intelligent production. In contrast with other applications (such as autonomous driving, face recognition, object classifications, and so on) relatively few research has been done in these technologies for the industrial operator domain. The challenge to perfect these new approaches and integrate them as important input in intelligent production configuration and human assistance systems for the Operator is very relevant.

In summary, a human-centered Industry 4.0 vision, that takes the human Operator as a key element, poses many challenges and interesting lines for Computer Graphics research.

REFERENCES

1. Stork, A. (2015). Visual computing challenges of advanced manufacturing and Industrie 4.0. IEEE computer graphics and applications, (2), 21-25.
2. Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., De Amicis, R., ... & Vallarino, I. (2015). Visual computing as a key enabling technology for industrie 4.0 and industrial internet. IEEE computer graphics and applications, 35(2), 26-40.

3. Tzafestas, S. (2006). Concerning human-automation symbiosis in the society and the nature. Int'l. J. of Factory Automation, Robotics and Soft Computing, 1(3), 6-24.
4. Romero, D., Bernus, P., Noran, O., Stahre, J., & Fast-Berglund, Å. (2016, September). The operator 4.0: human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In IFIP International Conference on Advances in Production Management Systems (pp. 677-686). Springer, Cham.
5. Funk, M., Kosch, T., Kettner, R., Korn, O., & Schmidt, A. (2016). motioneap: An overview of 4 years of combining industrial assembly with augmented reality for industry 4.0. In Proceedings of the 16th international conference on knowledge technologies and datadriven business.
6. Toro, C. , Barandiaran, I., Posada, J. A perspective on knowledge based and intelligent systems implementation in industrie 4.0 (2015). Knowledge-Based and Intelligent Information & Engineering Systems KES-2015, Singapore. Procedia Computer Science, Volume 60, Pages 1-1886 (2015)
7. Marner M. R., Smith R. T., Walsh J. A., Thomas B. H. (2014). Spatial user interfaces for large-scale projector-based augmented reality. IEEE Computer Graphics and Applications, vol. 34, pp. 74–82.
8. Gard N., Eisert P. (2018). Markerless content independent projection plane tracking for mobile projector camera systems. Proc. Int. Conference on Image Processing (ICIP). Athens, Greece.
9. Seibold C., Hilsmann A., Eisert P. (2017). Model-Based Motion Blur Estimation for the Improvement of Motion Tracking, Computer Vision and Image Understanding.
10. Barré, R., Chojecki, P., Leiner, U., Mühlbach, L., Ruschin, D. (2009). Touchless Interaction - Novel Chances and Challenges. In Proc. Int. Conf. on Human-Computer Interaction. San Diego, USA
11. Endsley, M. , Kaber, D. Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics (1999). Vol. 42, NO. 3, 462 -492, Taylor & Francis.
12. Toro, C. , Sanín, C., Vaquero, J., Posada, J., Szczeribci, E. Knowledge based industrial maintenance using portable devices and augmented reality (2007). Lecture Notes in Computer Science (Springer) 4692 LNAI(PART 1), pp. 295-302.
13. Stojanovic, V., Trapp, M., Richter, R., Hagedorn, B., Döllner, J. (2018). Towards The Generation of Digital Twins for Facility Management Based on 3D Point Clouds. 34th Annual ARCOM Conference, 3-5 September 2018, Belfast, UK.
14. Discher, S., Richter, R., Döllner, J. (2018) A scalable WebGL-based approach for visualizing massive 3D point clouds using semantics-dependent rendering techniques. Web3D '18: Proc. 23rd Int. ACM Conf. on 3D Web Technology.
15. Claudy, L. (2012). The broadcast empire strikes back. I SpEctrum, 49(12).
16. Domínguez, A., Agirre, M., Flórez, J., Lafuente, A., Tamayo, I., & Zorrilla, M. (2018). Deployment of a hybrid broadcast-Internet multi-device service for a live TV programme. IEEE Transactions on Broadcasting, 64(1), 153-163.
17. Petersen, N., & Stricker, D. (2015). Cognitive augmented reality. Computers & Graphics, 53, 82-91.
18. Rambach, J., Deng, C., Pagani, A., & Stricker, D.(2018). Learning 6DoF Object Poses from Synthetic Single Channel Images. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR).
19. Rambach, J., Pagani, A., Schneider, M., Artemenko, O., & Stricker, D. (2018). 6DoF Object Tracking based on 3D Scans for Augmented Reality Remote Live Support. Computers, 7(1), 6.

ACKNOWLEDGEMENTS

We acknowledge the contribution of other members of the institutions of GraphicsMedia.net who have also supported the common research vision presented in this paper, especially Ivan Vallarino Jr. (Mivtech), Yan Cui (CGAII) and Larent Lucat (CEA-List). We would like to thank the following institutions for the partial financing of the research projects addressed in this paper: The EU Commision for the H2020 project MANUWORK (grant no° 723711), the German Government for the projects Easy-Cohmo, Prowilan and

Innoprom; the Portugal 2020 program for the project Intelligent4DMould and the Basque Government for the LangileOK, LANA II, MEMORI and PLATMON 360 projects.; as well as the companies IDS and Kimua Group.

ABOUT THE AUTHORS

Dr. Jorge Posada is President of the international network GraphicsMedia.net. Email: jposada@vicomtech.org

Dr. Mikel Zorrilla, Ana Dominguez and Bruno Simoes (Vicomtech), **Prof. Peter Eisert** (Fraunhofer HHI), **Prof. Didier Stricker and Jason Rambach** (DFKI), **Prof. Jürgen Döllner** (Hasso-Plattner Institute), **Prof. Miguel Guevara** (CCG) are international collaborators working together in the GraphicsMedia.net network.

Contact department editor André Stork at andre.stork@igd.fraunhofer.de