

Remote Real-Time Collaboration through Synchronous Exchange of Digitised Human-Workpiece Interactions

John Oyekan¹, Vinayak Prabhu^{1, 2}, Ashutosh Tiwari¹, Vinubha Baskaran³, Mark Burgess³, Rob McNally⁴

¹Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK

²Nanyang Polytechnic, 180 Ang Mo Kio Ave 8, Singapore, 569830

³Holovis International, Mere Lane, Lutterworth, Leicestershire LE17 4JH, UK

⁴Jaguar Land Rover, Abbey Road, Whitley, Coventry, CV3 4LF, UK

Corresponding author: John Oyekan; **Email:** j.o.oyekan@cranfield.ac.uk

Abstract In this highly globalised manufacturing ecosystem, product design and verification activities, production and inspection processes, and technical support services are spread across global supply chains and customer networks. Therefore, a platform for global teams to collaborate with each other in real-time to perform complex manufacturing-related tasks is highly desirable. This work investigates the design and development of a remote real-time collaboration platform by using human motion capture technology powered by infrared light based depth imaging sensors borrowed from the gaming industry and a synchronous data transfer protocol from computer networks. The unique functionality of the proposed platform is the sharing of physical contexts during a collaboration session by not only exchanging human actions but also the effects of those actions on the workpieces and the task environment. This enables teams to remotely work on a common engineering problem at the same time and also get immediate feedback from each other which is vital for collaborative design, inspection and verifications tasks in the factories of the future.

Keywords: Collaborative; Platform; Gaming; Virtual; Global teams; Industry

1 Introduction

A new wave of globalisation has resulted in product Research and Development hubs of various Multinational enterprises (MNE) being located to countries such as India and China as well as sourcing knowledge from globally dispersed knowledge hubs [1,2]. Furthermore manufacturing of

products have also been offshored to foreign countries where the low-wage economy enables MNEs to take advantage of lower production costs [3-5]. This has resulted in the need for MNE employees to travel to countries to collaborate on design tasks, address production problems as well as carry out maintenance tasks on production lines. This leads to longer downtimes in production plants as well as delays in design tasks.

In order to address these challenges, a technological solution for businesses to engage in collaborative activities as well as a platform to remotely provide expert services is needed. Such a solution should enable true collaboration between teams and facilitate a closed loop process through immediate feedback on the tasks being performed using bi-directional information flow. This requirement is highlighted by a large MNE automotive company as follows: "with the unprecedented expansion of our business across the globe, a cost-effective real-time collaboration solution is essential to enable an agile response to the growing demands of remotely located business units such as manufacturing sites, retailers and service centres".

This paper investigates the possibility of equipping geographically dispersed teams with the capability to work collaboratively on the same physical task at the same time by sharing the task context with each other. Even though there has been some work on collaborative virtual prototyping as in Xiao et al [6] and He et al [7], there are currently no solutions that can offer a true remote real-time platform for collaborative physical tasks other than the traditional means of communication using text and voice calls, video conferencing and file sharing.

In this work, a true collaboration platform is built on the theoretical foundation of digitisation of human-workpiece interactions as discussed in Prabhu et al [8]. According to this theory, any physical task can be broken down into a series of human-object interactions where every human action is followed by object feedback, which is analysed by the human on the fly, and the next appropriate action is chosen and implemented towards channelling the task to successful completion. In this work, human actions and corresponding changes to the workpieces are simultaneously tracked in real-time to produce a digital data stream that represents the task. This data stream is synchronously exchanged over a computer network between sites to enable remote collaboration. The proposed method is capable of hosting collaborations between virtual and virtual, real and real and virtual and real task environments. The unique aspect of this research is the use of off-the-shelf human motion-capture technology provided by depth imaging sensors to not only capture human actions but also their effects on the objects during the task together with a synchronous data transfer protocol.

This research contributes new knowledge as follows: (i) a method to share task contexts in a real-time collaborative environment, (ii) a method to digitise human activities during a task in simple data structures and the ability to convert them back to rich task information. The data is exchanged between multiple sites using a synchronous data transfer protocol. This enables low latency, low bandwidth internet transmission across two geographically separated sites, (iii) the use of low-cost gaming interface sensors to capture and digitise collaborative tasks and (iv) a method to extend the theory of human-workpiece interactions in simultaneous multi-site collaborations and map collaborative human actions to the corresponding changes to task workpieces in real-time.

The outcome of this research has the potential to significantly enhance the quality of manufacturing operations in the factories of the future by virtue of it facilitating the real-time sharing of best practices between global manufacturing sites at low cost. These operations range

from product and process design, verification, assembly, and inspection to through-life engineering and maintenance services.

The remainder of this article is organized as follows; Section 2 presents related research in the area of remote real-time collaboration. Section 3 describes the methods used to carry out the research and section 4 reports the research results. Section 5 presents a discussion on the research conducted and its outcomes followed by the conclusion.

2 Related Work

The internet has resulted in the possibility to achieve collaborative work between two geographical separated sites of a NME. Most collaborative work is often done using traditional communication media of e-mail, phone, video conferencing and fax. Even though useful this medium has been found to be ineffective and time consuming in tasks that involve physical artifacts due to the lack of tactile feedback from the artifact and inability to grasp the physical constraints and dimensions of the task environment as well as the artifact [9].

In order to address these challenges, the online gaming community could offer some insights. A form of collaboration occurs in massive player online gaming (MMORPG) communities where players co-operate or compete to fulfil various virtual game objectives generated by a computer [10–12]. Currently, these environments are mostly used for pleasure or entertainment purposes but could in future be used for the purposes of training a group of geographically separated individuals in collaborating to handle a physical task.

For such a technology to be useful for physical tasks on shop floors, there is still a need to relate what is happening in the real world with the virtual world and vice versa. In other words, a mechanism is needed to capture the effects of human physical actions on physical workpieces and convert this into a virtual form that is visible to a collaborator at a remote site towards generating feedback and next action steps. Furthermore, a way of transmitting the actions of the collaborator on virtual workpieces at the remote site to information that can be seen by the trainee in the physical world is required [13].

Some researchers have made some progress towards this. For example, Galambos et al. [14] discusses a virtual collaboration arena software frame that use a virtual reality engine for the purposes of training new employees in manufacturing and production engineering. Though, in their work only head movements and gaze direction were implemented.

In order to capture full body human motion, several research attempts tend to use gaming interface sensors such as the Microsoft Kinect. With the Microsoft Kinect sensor, detecting and tracking of physical objects as well as obtaining the physical constraints of workspaces is possible. This makes it possible to accurately project the physical world into a virtual space and vice versa. In Adcock et al. [15] for example, Kinect sensors are used to scan the physical workspace and detect object changes using fusion point cloud technique and a Spatial Augmented Reality (SAR) mechanism to project a remote expert's instructions as graphical annotations overlaid on the workspace. Similarly, Tecchia et al. [16] proposed a Kinect-based method to provide real-time aid to another person based on the capture and rendering of remote workspace and of the helper's hands using tele-presence. In their work, Kinect sensors were used to capture the point cloud of the objects manipulated in the workspace to create their digital versions, which are then transmitted to the remote expert who can view them in immersive space.

The expert then uses hand gestures to provide instructions on the task, which the user can view on a computer screen with the expert's virtual hands overlaid on top of his/her own live video of the task.

Piumsomboon et al. [17] proposed a framework that enabled face-to-face collaboration, allowing users present at the same site to use their hands to naturally interact with each other using virtual objects, enabled by Kinect sensors and AR viewing cameras. Users' hands are tracked and the manipulation of the virtual object rendered using AR is captured by mapping the real hand coordinates with virtual object positions in 3D space. Similarly, Sodhi et al. [18] reported a method that allows users to manipulate virtual objects in a real physical environment. Their work uses a mobile phone integrated with depth sensors to track the location of the user's fingers, as well as to capture the 3D shape of the associated objects. A second user can see this information on his/her mobile phone and using finger gestures provide instructions on the task to the first user, who can see these gestures on his/her own mobile phone, superimposed on the real video feed thereby enabling real-time collaboration.

In order to achieve full immersion in a virtual environment constructed from a physical one, a tool that facilitates this conversion is needed. The use of Unity 3D and a Kinect sensor was proposed by Mossel et al. [19] to achieve a real-time virtual collaboration method. In their work, one user's motion is tracked resulting in a virtual avatar that a second user interacts with in real time using the mouse interface on another computer. This leads to a virtual collaboration session housed within Unity3D gaming environment. Similarly, Kurillo and Bajcsy [20] proposed a 3D tele-immersion system to provide real-life-like interaction between remote teams. Kinect sensors are used to track the motion of the users involved in the interaction and rendered as 3D avatars using a virtual reality environment. These avatars are able to manipulate 3D objects or collectively work with virtual components in a virtual environment thereby enabling real time collaboration.

There are five marked differences between the above reported articles and this research in terms of the underlying concept, the methodology used as well as the functionalities proposed. The first difference is that this research proposes a true two-way real-time collaboration between remote teams in which human actions and their effects on the workpieces on both sides are exchanged. This enables both sides to not only demonstrate tasks but also obtain immediate feedback on their tasks from each other. The second difference is that the above articles confine their task capture zone to restricted workspaces, such as human hands only whereas in this work the workspace is the entire 3D space in front of the Kinect cameras and the whole upper body of the user is tracked and digitised. The third difference is that the above articles capture the objects and humans involved in the task as disconnected and independent entities but in this work the direct interaction between humans and the task objects is captured enabling a richer collaboration experience. The fourth difference is that apart from enabling collaboration between a real and a virtual or between two virtual task sites, this research also enables collaboration between two real task sites. Finally, the collaboration mechanisms proposed by the above articles involve the transfer of rich data such as videos, images and point clouds making them dependent on the network bandwidth for smooth data transfer whereas in this work, the ability to capture and transfer human-workpiece interactions in simple data structures yet render rich information from them eliminates any network bandwidth dependence.

3 Method

3.1 Platform Architecture

Prabhu et al.[21] proposed an architecture for a real-time collaboration platform for two collaborating sites. That platform was extended in this work for the purposes of supporting an automotive repair service use case as shown in Figure 1.

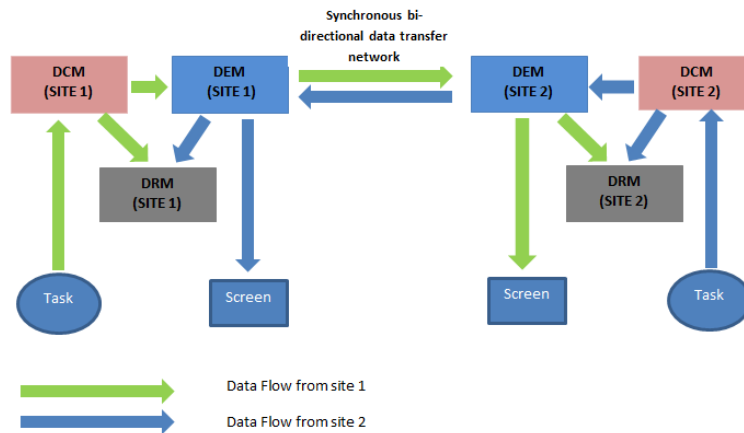


Figure 1: Software architecture of the real-time remote collaboration platform

The extended platform architecture consists of three main components, namely, a Data Capture Module (DCM), a Data Record Module (DRM) and a Data Exchange Module (DEM). The main functions of the DCM is to capture task related information such as (i) the actions of humans during the task through the use of motion capture techniques, (ii) the changes to the workspace through the use of object state change tracking techniques and (iii) displaying of topics relevant to the task the user is working on using virtual grid mapping. The output of the DCM is digitised human actions, workspace state data as well as digital grid data that is produced continually as the task progresses on both sides of the collaboration. The function of the DRM is to record the data captured by the DCM for knowledge capture, analysis and reuse. This digital data is passed on to the DEM for real-time transmission to the collaborating site.

The two DEMs are connected over a data network such as the Internet using a bi-directional data transfer protocol. The DEM has three main functions: (i) to receive data from the DCM, package it into appropriate data packets and send these packets over the network to a receiving DEM on the other side and (ii) to receive data packets sent by the DEM on the opposite side, (iii) extract the human action and workpiece tracking data from these packets. The extracted data is then passed to the data renderer for rendering using an appropriate graphical medium. The DEM performs the send and receive functions across the network synchronously and in real-time to enable lossless exchange of task information. In this work, we have modelled the internet as having three attributes namely: bandwidth, latency and data buffer. Text based data structures were utilized in

order to reduce these modelled attributes of the internet on our approach. Figure 2 shows the Entity Relationship (ER) diagram that shows the flow of data in one site. The same diagram can be mirrored to represent the data flow on the collaborating sites.

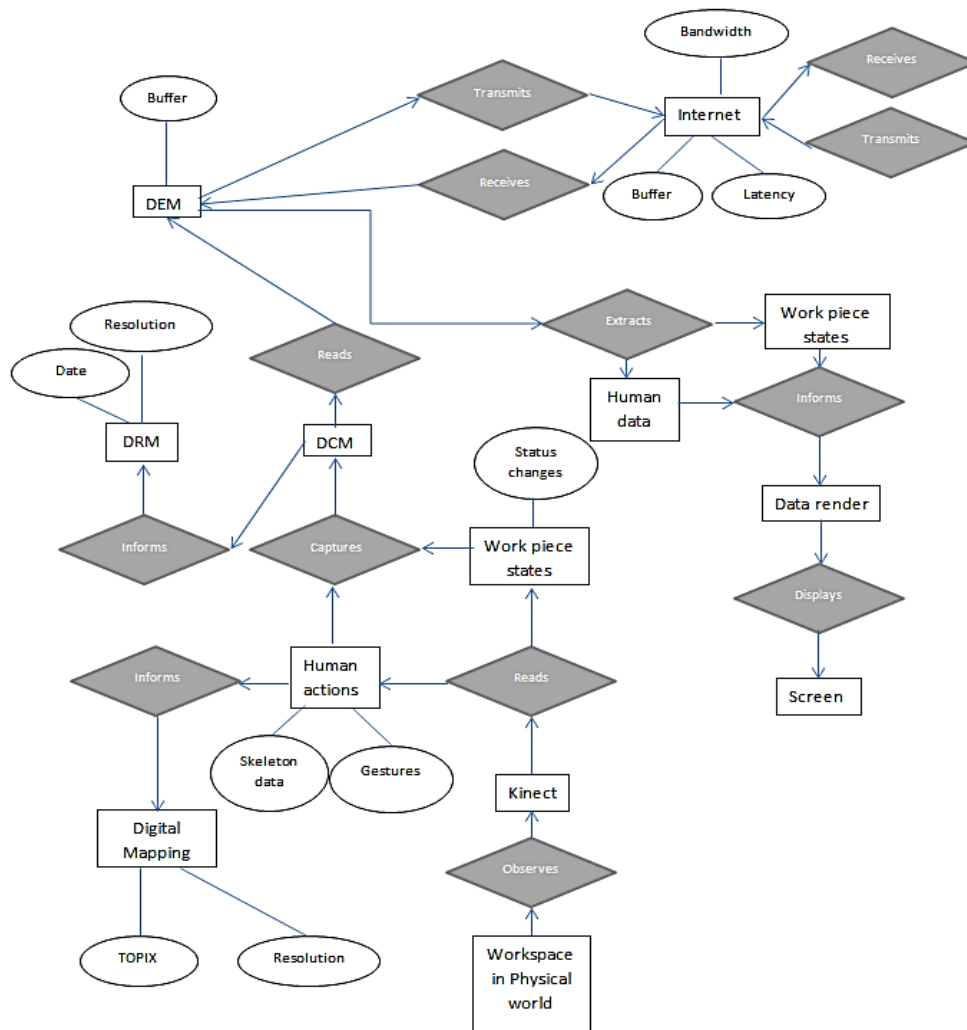


Figure 2: Entity Relationship diagram for the real-time remote collaboration platform showing data flows.

3.2 Use Case Design

In Prabhu et al.[21], a remote collaboration task that involved the construction of a lego brick structure across two geographical sites was used as a use case. In this paper, the real-time remote collaboration platform is implemented and tested using the use case of a roof rail assembly task on a vehicle.

The roof rail assembly task involves a mechanically complex series of steps that involve the use of tools to (i) remove strip and guides from the roof of the vehicle, (ii) install additional

components, (iii) replace the strip and guides onto the roof and (iv) install the roof rail. Figure 3 shows a chosen set of inserts extracted from the vehicle manual.

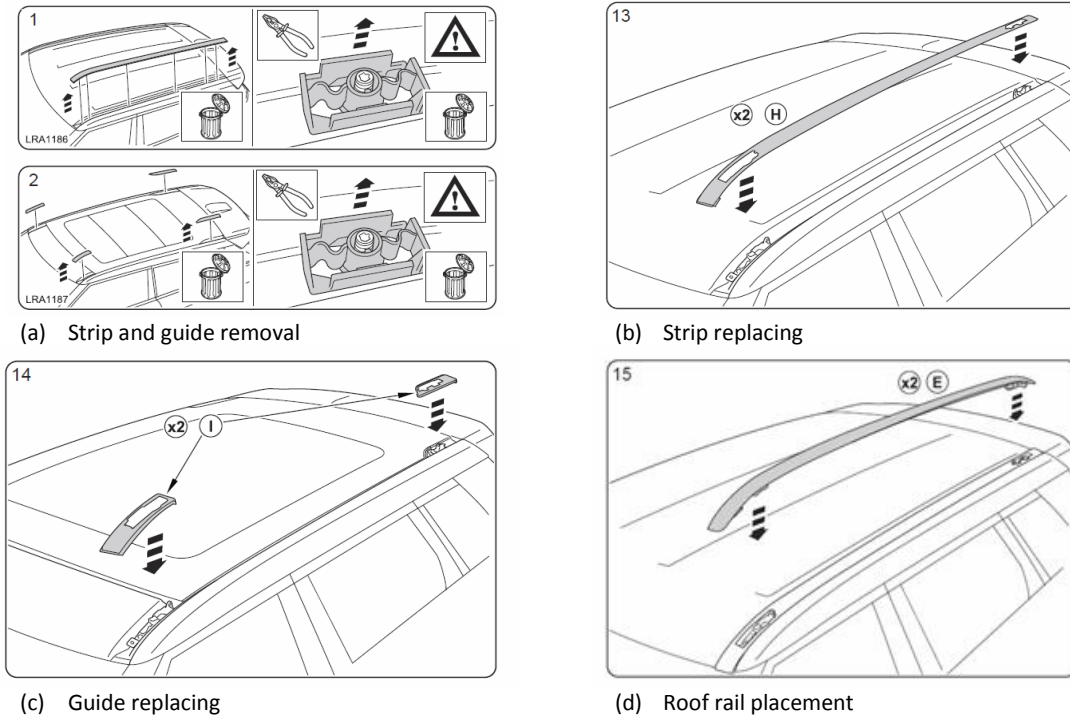


Figure 3: A selection of installation steps from a roof rail assembly task [22].

In addition to real-time remote collaboration, the developed platform attempts to provide important just-in-time information in a hands free manner to the technician. This enables the technician to work on the vehicle assembly task without the need of leaving the task in search for information in paper based or computer based manuals thus boosting productivity and efficiency.



(a) The real site with the Kinect optimally placed to track the technician as well as to observe the physical vehicle



(b) The remote virtual site with a user using an Oculus rift for visualization and Kinect for tracking

Figure 4: The setup for use case scenario

In the event that the information in the manual is not comprehensive enough or unclear, then the technician uses the real-time remote collaboration platform to seek help from an expert trainer. Once established, the technician attempts to assemble the roof rail while the trainer observes the

progress of the task in a virtual environment created in Unity3D and visualised using an Oculus rift. When the technician requires visual aid on a task, an Oculus rift is worn and a CAD overlay of the vehicle is superimposed on the real vehicle. Any manipulation of the CAD in the virtual environment and actions of the trainer are transmitted over the network where they are seen as augmented reality in the real world. Therefore, both users are able to see each other's actions in real-time, provide feedback to each other towards completing the roof rail assembly task in a truly collaborative manner.

In this paper, we focus on capturing the actions of the technician as well as the state of the workspace (vehicle) after which they are transferred over the internet to a virtual environment for trainer observation and reaction.

3.3 Setup

The setup at the real site (Figure 4a) consists of a vehicle, a simplified model of a roof rail, a Microsoft Kinect gaming interface sensor, a PC connected to the sensor and the software architecture above coded in a C# application for integration with the Unity3D game development engine. The sensor was placed in such a way as to observe both the human's actions as well as the roof surface of the vehicle being worked upon. The depth and RGB streams from the Kinect were used in this work with frame rates of up to 30 frames per second. The RGB image provided by the Kinect is at a resolution of 1920 x 1080 pixels and the depth image is at 512 x 424 pixels.

In this work, bi-directional data transfer between collaborating sites was implemented using the socket connection API provided by Unity 3D for a peer-to-peer connection using a dedicated cross Ethernet cable. The platform was deployed on an intra-company LAN of one of the collaborating organisations in which multiple computers were connected to the intranet using Ethernet cables. In order to provide synchronicity of data transfer, a 2-way handshake protocol was used as discussed in section 3.5 below.

The setup at the virtual site (Figure 4b) consists of a PC to which a Microsoft Kinect gaming sensor and an Oculus rift is connected to. The Oculus rift is used to provide an immersion into a virtual environment that was constructed in Unity3D game development engine. The virtual environment also contained a virtual vehicle that was constructed using the CAD model of the real vehicle. The Microsoft Kinect was used to capture and track the gestures and motion of the trainer for transmission by the DEM to the real world site.

3.4 Data Capture Method

Three types of data are captured from the real site during the collaborative assembly task. These are the motion capture data and gesture of the human, the workspace state on the vehicle roof during the assembly task as well as the position of the human in relation to the digital vehicle grid (as explained below). The data captured from the virtual site are the gestures and actions of the trainer. The gestures captured include open and close states of the trainer's hand to simulate the grabbing and releasing of virtual objects as well as the skeletal motion of the trainer's body during posture and positional changes as well as during the action of moving grabbed objects in the virtual space.

3.4.1 Digital Vehicle Grid: The vehicle was digitally mapped through the use of a digital gridding system as shown in Figure 4. The digital map covered the entire right side profile of the vehicle. Each grid of the digital map had a resolution of 800 x 500 mm and corresponded to a region of the vehicle in which a part of interest was located. Standard service manuals related to each of the part were digitally created in PDF format and stored in a repository. As the technician moves from grid to grid, the corresponding digital service manuals were activated and displayed to the technician. In order to achieve accurate coverage of the vehicle, a starting datum point at the top right of the vehicle was used. As there could be multiple components, sub-components or sub-steps in each grid, a graphical user interface populated with required contextual information was developed (2 in Figure 5). User hand state was used to choose the appropriate PDF for display from a list of PDFs corresponding to the grid in which work was being carried out. An open hand state meant open PDF document while a close hand state meant close document. A combination of hand state and swiping motion was used to navigate the pages of the chosen PDF.

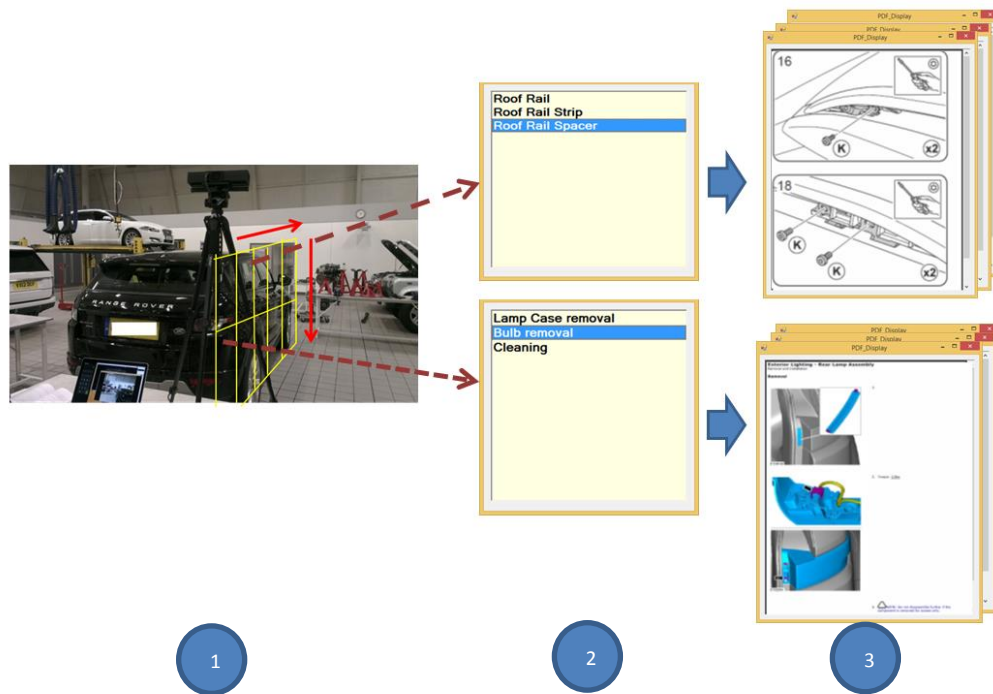


Figure 5: An example showing how user manuals are stored and called using contextual information in each grid. 1) is the digital vehicle grid; 2) is the graphical user interface showing the various components in each grid; 3) is the PDF chosen from a bundle of stored PDFs.

3.4.2 Datum capture: The starting datum was needed in order to place the virtual vehicle accurately in the virtual environment while respecting the real world physical spatial distances and orientation. This process is very crucial in ensuring an accurate translation of the real world technician's motions, gestures, and actions on the real vehicle to the virtual vehicle in the virtual environment. This datum point was set by touching the roof tip of the vehicle with the right hand and then using closing the left hand. In order to obtain accurate orientation of the vehicle for the CAD model in the virtual site, two more datum points (one in the z direction of the Kinect and the other in the x direction) were captured as explained in section 3.7.

3.4.3 Human Motion and Gesture Capture: Human motion and gesture capture was achieved by positioning the Kinect so that a complete view of the body of the user in the real site was obtained as well as a full view of the vehicle's roof. The standard C# Kinect SDK 2.0 library was used to obtain 3D positions of the human upper body joints such as head, neck, shoulders and arms. Therefore, this provided the DCM with a stream of digital (x, y, z) coordinates representing human motion involved in the roof rail installation task at up to 30 times per second.

Left hand gesture such as grasp and release were used to communicate with the PC when setting the reference datum of the vehicle for the vehicle digital mapping algorithm and "lock" the relevant service manual (example: to install the roof rail) in place when the user is in a particular virtual vehicle grid and continue working with the selected component of the vehicle.

3.4.4 Assembly task tracking: Along with human motion and gesture tracking, the Kinect is used to simultaneously track the progress of the roof rail assembly task. In order to achieve this, the red, green and blue pixels groups are identified from within each RGB image frame in real-time. These pixels groups are located on the roof rail space where work is to be carried out. A simple frame comparison and differencing mechanism of the roof rail workspace is performed by comparing the colour values of pixels belonging to each pixel group at time t with the colour values of the previous pixels at time $t - 1$. Any difference in frame above a threshold is recognised as a state change in the roof rail assembly task and was used to trigger a state change in a Markov chain model of the assembly task. The Markov chain modelled the sequence of three states namely: a strip remove state, a strip replace state and a roof rail placement state. Each change in state was digitised and communicated via the DCM to the virtual site.

3.5 Data Transfer Method

In this work, the real time collaboration is enabled between two sites by considering one site as the server and the other as a client. The trainer on the server side operates in a virtual environment while the technician on the client side operates in the real environment. The technician on the client side initiates the communication with the server when technical support is needed. Communication is achieved via a peer-peer Ethernet connection between two PCs. Before the client-server connection can be established, the client side ensures that the datum of the vehicle is acquired via the 3 points as explained above. The datum points are used to place the virtual vehicle (3D CAD model) in the virtual environment in the same relative position as the real vehicle at the real site and to overlay the CAD onto the real vehicle on the client side in the future. A connection is then established with the server after which the client sends the name of the required vehicle model to be loaded in the virtual environment along with the Datum point. The server then loads the model appropriately and starts the collaboration. In the chosen scenario, the client user cannot interact with the CAD in the virtual environment. Instead, actions of the technician on the client side are sent to the server in the form of skeleton data, data representing workspace state changes (e.g. when the technician removes the strip), and data representing the grid position of the technician relative to the real vehicle. On the server side, the skeleton data from the client-side is rendered in order to display the technician's position in space, workspace data is used to switch CAD components on or off depending on the state at the real site and the grid data is used to bring up the relevant service manual that the technician is also seeing at the real site. As a result, collaboration between the virtual and real technician can be maintained (See Figure 11).

The DEM receives the data from the DCM in the form of human skeletal coordinates and the spatial coordinates of workpieces during the vehicle upgrade task as two separate streams with a common timestamp. These streams from each image frame are combined to form a comma separated data string (Figure 6).

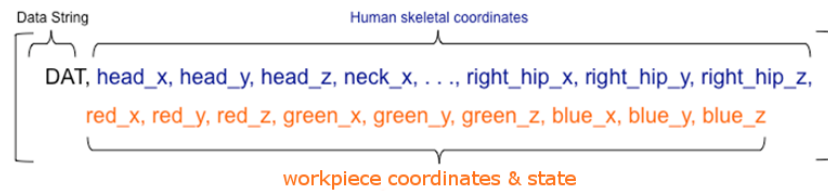


Figure 6: Data string representing the task information at any given instance

This data string is generated up to 30 times per second on both sides of the collaboration. As the DEM on both sides must work synchronously to maintain data sequence and avoid data loss, a bi-directional handshake protocol is implemented for data transfer. The use of request / start character (S) and acknowledgement / end character (E) ensures that the data transfer does not get jumbled up and the data strings in their entirety are transferred across the network in the sequence they were created. This protocol is illustrated in Figure 7.

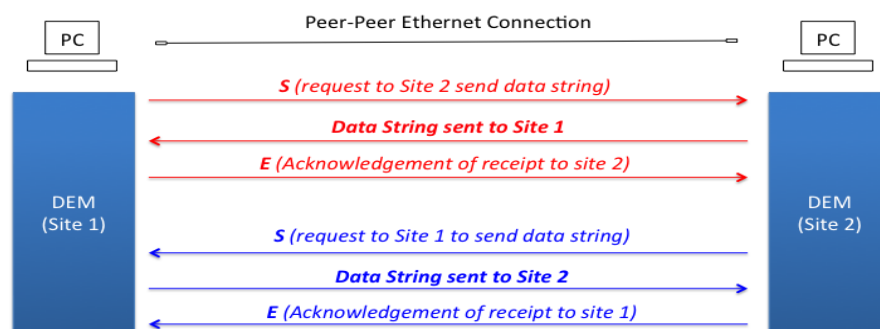


Figure 7: 2-way protocol for bidirectional data transfer

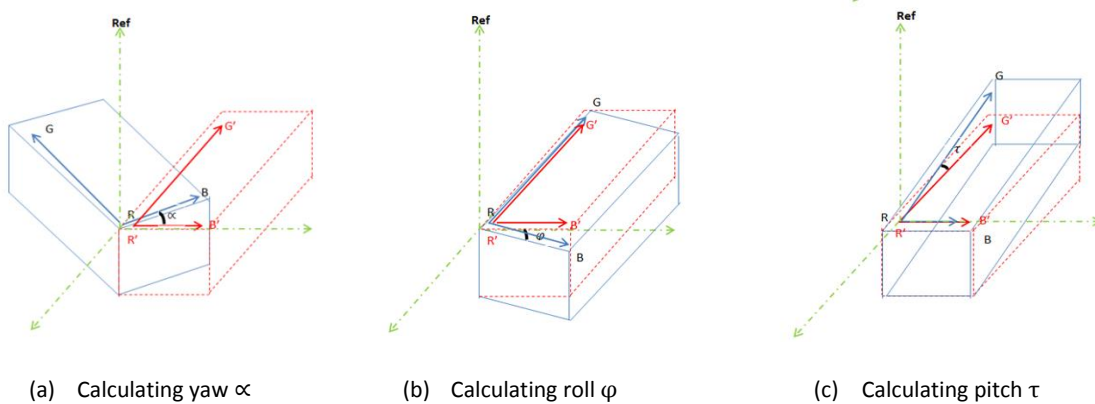
3.6 Data Rendering

The other major function of the DEM is to graphically render the data it receives from the DEM on the other side using a suitable visual medium. An example of such a medium is 2D animation on a computer screen or an immersive environment created using virtual reality. In this work, Unity3d was used to create the immersive environment. The incoming data packet is parsed to extract human motion and work piece position data. The technician’s skeleton is then rendered on screen using a stick figure and subsequently an avatar corresponding to the spatial positions of the skeletal joints whereas the work pieces are rendered by positioning the corresponding CAD models at the corresponding spatial positions in the virtual environment.

3.7 Placement of objects

In order to ensure that real time collaboration occurs correctly between geographically separated sites, it is necessary to have a common frame of reference for both sites. This should involve correct assignment of the orientation of the vehicle as well as its correct locational placement in the virtual environment. In order to achieve this, three datum points were used. These datum

points were obtained from the technician by using the approach discussed in section 3.4.1 and 3.4.2. For each of the datum points, the real world co-ordinates (x, y) and depth information were obtained. These datum points were used to calculate the orientation of the vehicle according to Figure 8 and Equations 1 to 3. The boxes in the figure below indicate the position and the orientation of the real vehicle.



(a) Calculating yaw α

(b) Calculating roll ϕ

(c) Calculating pitch τ

Figure 8: Using datum approach to calculate orientation of an object

$$\alpha = \sin'(R_{depth} - B_{depth} / \overline{RB}) \quad (1)$$

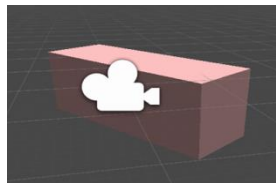
$$\phi = \sin'(R_Y - B_Y / \overline{RB}) \quad (2)$$

$$\tau = \sin'(R_Y - G_Y / \overline{RG}) \quad (3)$$

Where B_{depth}, R_{depth} , are the depths of the points B and R; R_Y, B_Y and G_Y are the real world y co-ordinates of points R, B and G.



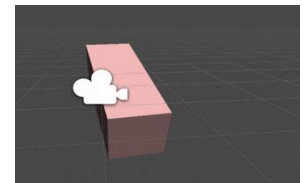
(a) Physical Yaw



(b) Virtual CAD Yaw



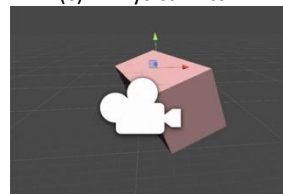
(c) Physical Pitch



(d) Virtual CAD Pitch



(e) Physical Roll



(f) Virtual CAD Roll

Figure 9: Correlation of physical box orientation and its corresponding virtual CAD orientation

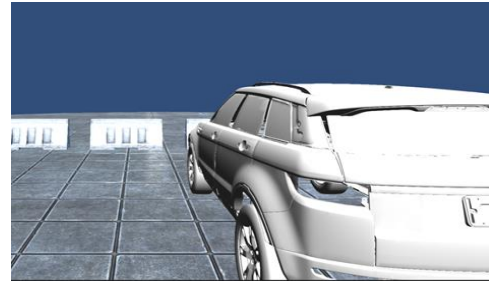
4 RESULTS

4.1 Placement of the vehicle

Using the approach discussed in section 3.7, a box was used to simulate a vehicle. Three datum points were obtained from the box on the desktop. These points were used to place a box CAD model as shown in Figures 9a to 9f. Similarly, the approach was used to place the vehicle CAD model in the virtual environment as shown in Figure 10.

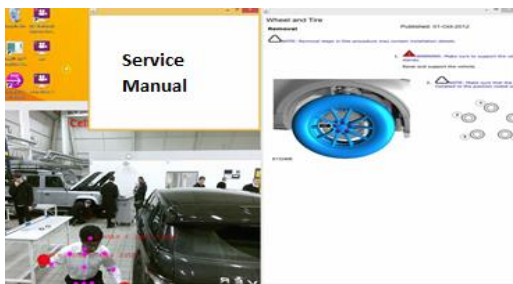


(a) Real vehicle

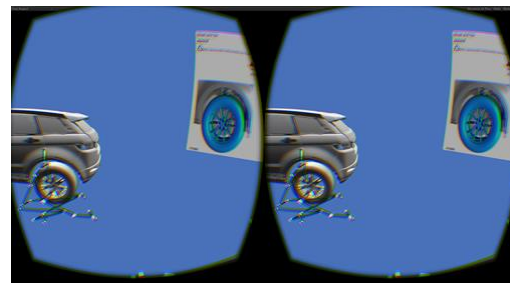


(b) Virtual model of vehicle and environment

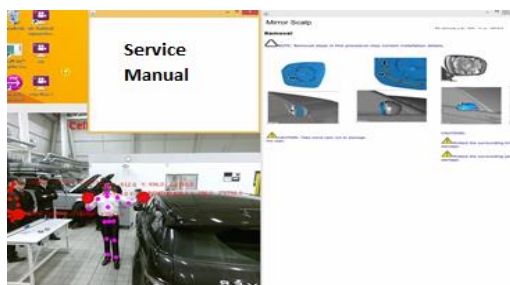
Figure 10: Example of vehicle placement in CAD using real world co-ordinates



(a) Shows a user of the system close to the wheel a vehicle with the related service manual displayed in the top right window.



(b) Showing a user in the virtual world seeing the position of the real world user and the wheel service manual seen by the real world user.

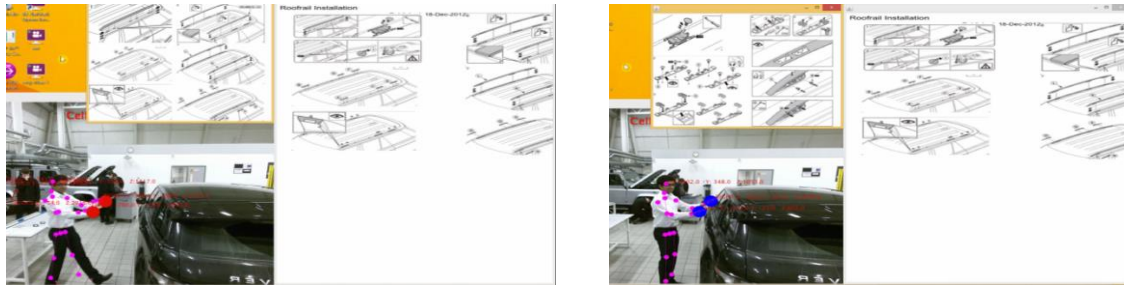


(c) Shows a user of the system close to the wing mirror of a vehicle with the related service manual displayed in the top right window.



(d) Showing a user in the virtual world seeing the position of the real world user and the wing mirror service manual seen by the real world user.

Figure 11: Showing the displayed service manual in the real environment and virtual environment.



(a) Showing a zoomed out view of the roof rail service manual (b) Showing a zoomed in view of the roof rail service manual

Figure 12: Showing how the service manual changes in relation to distance from a component

4.2 Digital vehicle mapping

Each grid of the vehicle was associated with a digital service manual that shows the sequence of steps to perform in order to complete a task. The technician's hand was used to detect in which grid he/she is working in (Figure 11(a & c)) and the grid information is transmitted over the network to the virtual site where it is rendered as shown in Figure 11(b & d).

4.3 Using Gestures for Human-Computer Interaction

In order to prevent the digital vehicle mapping from changing topics while working on a particular vehicle's component, the technician can lock the current service manual being displayed by bringing both hands together. Also, dynamic zoom changes of the service manual was implemented so that if the technician is further away from the component, an overview of the service manual is displayed while if closer, a closed up view is displayed as shown in Figure 12.

4.4 Tracking of Workspace State Changes

As discussed in section 3.2, a simplified roof rail assembly task was used as a use case. This involved removing a strip from the vehicle, placing the strip back and then putting on the roof rail onto the strip. These stages represented states that were captured using an object state change tracking module.



(a) Roof rail strip on vehicle



(b) Roof rail strip removed from vehicle



(c) Roof rail strip on virtual vehicle



(d) Roof rail strip removed on virtual vehicle

Figure 13: Tracking of roof rail strip status on the real vehicle and transferring to the virtual environment

These changes in states were transmitted to the server where they will be used to activate or deactivate the CAD elements associated with each of the states. Figure 13 shows how the module is able to track changes in roof rail strip state status and the corresponding changes in CAD model in the virtual environment.

5 Discussion

The global manufacturing industry needs to explore modern means of communication and collaboration for exchanging ideas, knowledge and best practices. The companies need to look for new ways to increase operational efficiency and decrease duplication of resources across their global operations to remain competitive and sustainable.² Also, with skilled and experienced personnel in short supply, skill transfer as well as centralised delivery of expert services becomes desirable. This work proposes a collaboration platform that will potentially enable geographically dispersed teams to collaborate with each other in real-time for jointly performing tasks and solving problems.

This work investigates the design and development of such a collaboration platform through the use of human motion capture technology borrowed from the gaming industry and grounded by the theory of digitisation of human-workpiece interaction. The unique functionality of this platform is the sharing of physical task contexts during a collaboration session, especially the workpieces involved, and the exchange of human actions and the effects of those actions on the workpieces. This enables teams to remotely work on a common problem at the same time and also get immediate feedback from each other.

Existing commercial collaboration platforms facilitate communication using exchange of text as well as exchange of rich data such as voice, images, videos and files. Rich data requires a high-bandwidth network infrastructure for smooth transfer thereby excluding resource-deprived settings and increasing the cost of collaboration. This work investigates a new method of transferring simple data structures such as character strings to represent task information and convert this simple data back to rich information on either side of the collaboration. Therefore, this method of collaboration does not need high bandwidth networks. Furthermore, the plug and play architecture of the collaboration platform is designed to be independent of the type of devices

used and thus users would not be tied down to specific products and brands. Use of robust, proven, consumer-grade devices to capture and digitise tasks enables users to keep the ownership and running costs of collaboration down.

In this work, we are currently focusing on using object recognition techniques to detect state changes of external large scale assemblies such as the wheel, roof rails and door panels. State changes trigger the next state of a markov chain of the assembly sequence that is then sent to the virtual side where it is used to switch on or off corresponding CAD models of the real world assemblies. As a result, the expert at the virtual side can view whether the technician is actually carrying out the instructions correctly. However for objects beyond the detection capability of this work, the technician communicates using verbal commands which are transcribed into text and sent to the virtual side for the benefit of the expert. Using the above methods, the virtual world and real world is able to exchange information about the procedure being carried out.

Nevertheless, there are some limitations with our developed platform. The limitations are:

(i) This work uses the first generation of low cost gaming interface sensors, which also bring with them some disadvantages. The resolution of the depth and RGB imaging cameras used in this work are relatively lower than their superior industrial counterparts that are available at much higher costs. As a result, it was noticed that it was sometimes difficult to track the orientation of hands in the real world causing the avatar hands to twist at odd angles. This could be a challenge in collaboration tasks that require the communication of minute and precise hand movements. Nevertheless, this challenge could be solved using a commercially available RealSense depth sensor.

(ii) The simultaneous tracking of human actions and vehicle status is not possible all the time. This is because the depth imaging capability of the Kinect requires that the whole human's body be in constant line-of-sight of the sensor for tracking. This limits the kind of tasks that can be digitised to enable collaboration and as a result, is currently restricted to work outside the vehicle. Additionally, the Kinect can track humans up to a distance of 4 meters making it impossible to work on longer vehicles. This could be remedied through the use of multiple Kinects.

(iii) The frame differencing technique used in this work to track workspace changes is also susceptible to noise and reflections of moving objects onto the vehicle's roof. This resulted in false positive triggers. In the future, this will be solved through the use of more advanced object recognition and tracking approaches as well as through the use of RFID tags or other wireless object tagging technologies. This will ensure that rotation of objects can also be tracked.

(iv) Due to the limitations of current technology it was discovered that Oculus rift would result in health and safety issues in an active workshop due to restricted peripheral vision. Furthermore, its need to be constantly attached to a computer means that a realistic superimposition of virtual vehicle onto the real vehicle for augmented task training can only be achieved on very restricted bases. In order to overcome these limitations, we are currently investigating the use of a hand held display to achieve augmented reality based on the technician's location in relation to the real vehicle. On this display, the technician will be able to view the actions of the trainer.

6 Conclusion

This work proposes a unique method of providing a platform to enable remote real-time collaboration between geographically dispersed teams. The unique aspect of this work is the use of infrared light based low-cost depth imaging sensors to capture human actions and the effects of those actions on the workpieces and the task environment on both sides of the collaboration thereby enabling sharing of task contexts. This work contributes new knowledge to areas of digitisation and synchronous transfer of human activity during physical tasks to enable remote real-time collaboration. We have presented: (i) a low cost remote collaborative platform that is within the budget of most organisations, (ii) a platform that enables the exchange of knowledge, skills and transfer of expert services over a network so that task contexts can be shared, (iii) a method to enable the capturing of task information as well as the digitisation of human-workpiece interactions using simple data structures. This ensures low latency and low bandwidth transmission between two geographically separated sites, and (iv) an extension to the theory of human-workpiece interaction for simultaneous multi-site collaborations.

Collaborative working benefits cannot be over emphasised especially in the highly globalised manufacturing industry of the present and the future. Operations such as product and process design can vastly benefit from teams working together, sharing expertise, experiences, engineering and manufacturing constraints, and market needs in the same contextual platform. Such a platform will also enable experts to remotely demonstrate their skills using a one-to-many collaboration medium to improve manual tasks that are prevalent in machining, assembly and disassembly of complex parts during manufacturing and through-life maintenance operations. Finally, the ability to deconstruct an engineering environment into the simplest of data forms and construct it back at the remote sites in real-time potentially means that the same complex part could be worked upon by multiple teams at the same time for verifications and inspections. Such a collaborative capability between teams in the factories of the future will be vital to raise operational efficiencies and competitiveness.

Acknowledgement

The authors of this papers would like to thank Innovate UK for funding this research as part of the technology-inspired innovation - collaborative R&D in ICT - grant 101774 (August 2013).

References

- [1] G.J. Da Silveira, An empirical analysis of manufacturing competitive factors and offshoring, *International Journal of Production Economics*. 150 (2014) 163-173.
- [2] G. Tzeng and C. Huang, Combined DEMATEL technique with hybrid MCDM methods for creating the aspired intelligent global manufacturing & logistics systems, *Annals of Operations Research*. 197(1) (2012) 159-190.
- [3] C. Davis and A. Naghavi, Offshoring production: A simple model of wages, productivity, and growth, *Economic Inquiry*. 49(2) (2011) 334-48.
- [4] S. Kinkel and S. Maloca, Drivers and antecedents of manufacturing offshoring and backshoring—A German perspective, *Journal of Purchasing and Supply Management*. 15(3) (2009) 154-65.

- [5] B. Michel and F. Rycx, Does offshoring of materials and business services affect employment? Evidence from a small open economy, *Applied Economics*. 44(2) (2012) 229-51.
- [6] H. Xiao, Y. Li, J. Yu, J. Zhang and W. Tang, Dynamic assembly simplification for virtual assembly process of complex product in cloud computing environment, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*. 228(10) (2014) 1198–1213.
- [7] B. He, Y. Wang, W. Song and W. Tang, Design resource management for virtual prototyping in product collaborative design, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*. 229(12) (2015) 2284-300.
- [8] V. Prabhu, A. Tiwari, W. Hutabarat and C. Turner, Monitoring and Digitising Human-Workpiece Interactions during a Manual Manufacturing Assembly Operation Using Kinect™, *Key Engineering Materials*. 572 (2014) 609-612.
- [9] A. Shamsuzzoha, C. Toscano, L.M. Carneiro, V. Kumar and P. Helo, ICT-based solution approach for collaborative delivery of customised products, *Production Planning & Control*. (2015) 1-9.
- [10] G. Christou, E. Lai-Chong Law, P. Zaphiris and C.S. Ang, Challenges of designing for sociability to enhance player experience in Massively Multi-player Online Role-playing Games, *Behaviour & Information Technology*. 32(7) (2013) 724-34.
- [11] M. Peterson, Learner interaction in a massively multiplayer online role playing game (MMORPG): A sociocultural discourse analysis. *ReCALL, the Journal of EUROCALL*. 24(3) (2012) 361.
- [12] Y.L. Lin, H.W. Lin and Y.C. Jhan, Are We After the Same Thing? Differences among Different MMORPG Players, *Journal of Management and Strategy*. 6(1) (2015) 21.
- [13] A. Kofod-Petersen and J. Cassens, Using activity theory to model context awareness, In *Modeling and Retrieval of Context*, Springer Berlin Heidelberg. (2006) (pp. 1-17).
- [14] P. Galambos, Á. Csapó, P. Zentay, I.M. Fülöp, T. Haidegger, P. Baranyi, and I.J. Rudas, 2015. Design, programming and orchestration of heterogeneous manufacturing systems through VR-powered remote collaboration. *Robotics and Computer-Integrated Manufacturing*, 33, pp.68-77.
- [15] M. Adcock, S. Anderson and B. Thomas, RemoteFusion: real time depth camera fusion for remote collaboration on physical tasks, *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, Hong Kong, Hong Kong, November 17 - 19, pp. 235-242. New York: ACM.

- [16] F. Tecchia, L. Alem and W. Huang, 3D helping hands: a gesture based MR system for remote collaboration. Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry, Singapore, Singapore, December 2 - 4, 2012, pp. 323-328. New York: ACM.
- [17] T. Piumsomboon, A. Clark, A. Umakatsu and M. Billinghamurst, Poster: Physically-based natural hand and tangible AR interaction for face-to-face collaboration on a tabletop. In: IEEE Symposium on 3D User Interfaces (3DUI), Orange County, CA, USA, March 4-5, 2012, pp. 155-156.
- [18] R.S. Sodhi, B.R. Jones, D. Forsyth, B.P. Bailey and G. Maciocci, BeThere: 3D mobile collaboration with spatial input. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, April 27 - May 2, 2013, pp. 179-188. New York: ACM.
- [19] A. Mossel, C. Schönauer, G. Gerstweiler and H. Kaufmann, Artifice-augmented reality framework for distributed collaboration, International Journal of Virtual Reality. 11(3) (2013) 1-7.
- [20] G. Kurillo and R. Bajcsy, 3D teleimmersion for collaboration and interaction of geographically distributed users, Virtual Reality. 17(1) (2013) 29-43.
- [21] Prabhu V, Oyekan J, Tiwari A, Advikolanu Y, Burgess M and McNally R. Remote real-time collaboration platform enabled by the capture, digitisation and transfer of human-workpiece interactions. Proceedings of the international conference on sustainable design and manufacturing, Seville, 12- 14 April, 2015.
- [22] Jaguar Land Rover, Fitting instruction: Roof rails. Part no. VPLGR0110/111/112/113/. No. 1.