Prabhu V, Oyekan J, Tiwari A, Advikolanu Y, Burgess M, Mcnally R, Remote Real-Time Collaboration Platform enabled by the Capture, Digitisation and Transfer of Human-Workpiece Interactions, Sustainable Design and Manufacturing 2015, proceedings of the 2nd International Conference on Sustainable Design and Manufacturing, 12 - 14 April 2015, Seville, Spain, pages 513-527
Paper number: sdm15-073

KES Transactions on Sustainable Design and Manufacturing II
Sustainable Design and Manufacturing 2015 : pp.513-527 : Paper sdm15-073

Remote Real-Time Collaboration Platform enabled by the Capture, Digitisation and Transfer of Human-Workpiece Interactions

Vinayak Prabhu[1, 2], John Oyekan[1], Ashutosh Tiwari[1], Yohann Advikolanu[3], Mark Burgess[3], Rob Mcnally[4]

[1]Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK
[2]Nanyang Polytechnic, 180 Ang Mo Kio Ave 8, Singapore, 569830
v.prabhu@cranfield.ac.uk
[3]Holovis International, Mere Lane, Lutterworth, Leicestershire LE17 4JH, UK
[4]Jaguar Land Rover, Abbey Road, Whitley, Coventry, CV3 4LF, 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 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. 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 task environment. This enables teams to remotely work on a common task 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.*

1 Introduction

Manufacturing today is a truly global activity in which production of a single product typically involves multiple locations across the globe (Kogut and Kulatilaka, 1994). In order to facilitate this, a solution for businesses to engage in collaborative activities as well as a platform to remotely provide expert services is needed. True collaboration between teams in which bi-directional information flow that facilitates a closed loop process through immediate feedback on the tasks being performed is in demand. This requirement is highlighted by a large multinational 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 how geographically dispersed teams can collaboratively work on the same physical task at the same time by sharing the task context with each other. Currently there are 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, the collaboration platform is built on the theoretical foundation of digitised human-workpiece interactions as discussed in Prabhu et al (2014). According to this theory any physical task is 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 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. This research also investigates the means of achieving synchronous bi-directional data transfer to ensure low-latency, robust and lossless exchange of data in real-time between the collaboration sites. The unique aspect of this research is the use of off-the-shelf motion-capture technology provided by depth imaging sensors that use infrared light to bounce off 3D scenes and capture human motion from within those scenes. This research extends the functionality of these sensors by using them to also detect and track moving objects during a task. Microsoft Kinect and ASUS Xtion are two examples of consumer-grade depth imaging sensors.

This research contributes: (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, (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

Several research attempts have been made to develop remote collaboration systems for people to remotely interact with each other during physical tasks. Many

of these attempts use gaming interface sensors such as the Microsoft Kinect to track human motion and/or physical objects involved in the task. Therefore, this concept is not entirely new.

Adcock et al (2013) used Kinect sensors 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 (2012) 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 (2012) 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 (2013) 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.

Mossel et al (2012) proposed a real-time virtual collaboration method using the Kinect sensor and the Unity3D game engine in which one user's motion is tracked producing a virtual avatar and the second user interacts with this virtual avatar in real time using the mouse interface on another computer. The virtual collaboration is housed within a gaming environment developed using the Unity3D game engine. Kurillo and Bajcsy (2013) 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 either a virtual reality environment or a display screen. 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 and 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

The software architecture of the proposed real-time remote collaboration platform for two collaborating sites is shown in Figure 1.

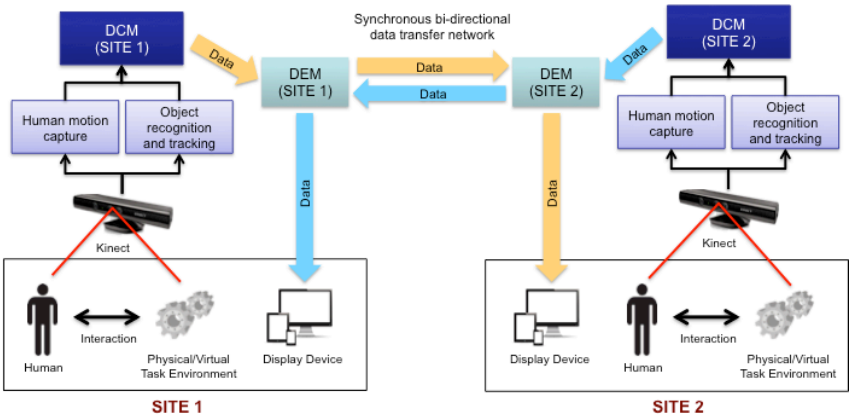


Figure 1: Software architecture of the real-time remote collaboration platform

The platform architecture consists of two main components, namely, a Data Capture Module (DCM) and a Data Exchange Module (DEM. The main function of the DCM is to capture the actions of the human by using motion capture techniques and to track the changes to the workpieces in the task environment through the use of object detection and recognition techniques. The output of the

DCM is digitised human action and workpiece tracking data produced continually as the task progresses on either side of the collaboration. This digital data is passed on to the DEM as and when it is produced in real-time.

The DEM has two 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, extract the human action and workpiece tracking data from these packets and render this data 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. The two DEMs are connected over a bi-directional data transfer network such as the Internet.

3.2 Use Case Design

The real-time remote collaboration platform is implemented and tested using the use case of Lego block assembly. In this use case, the users on both sides of the collaboration attempt to assemble a set of Lego blocks in a particular formation (Figure 2) with one user teaching the assembly process to the other in real-time.

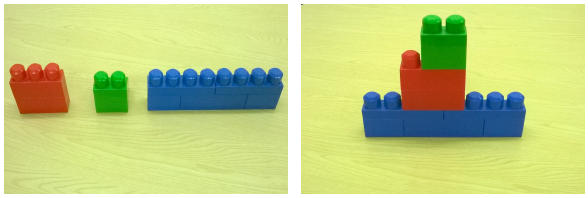


Figure 2: Lego blocks and the assembled formation

Two different collaboration scenarios are explored, namely, a real-real collaboration in which real Lego blocks are assembled on both sides and a virtual-real collaboration in which one side uses real Lego blocks and the other side uses virtual ones. In order to simplify the object detection and tracking, Lego blocks of the three primary colours (Red, Green and Blue) were used.

The process of manoeuvring a Lego block by the user on one side (site 1) is digitised and sent to the other side (site 2) where it is rendered as an animation which the user at site 2 observes and follows. The manoeuvre performed at site 2 is simultaneously digitised, sent to site 1 and rendered as an animation. Therefore, both users are able to see each other's actions in real-time, provide feedback to each other and cooperate to complete the assembly task successfully.

3.3 Setup

The setup at either side of the collaboration consists of an assembly table with individual Lego blocks (real or virtual), a gaming interface sensor such as the ASUS Xtion, a PC connected to the sensor, and the software architecture described above coded in a JAVA-based application. The sensor is placed at a

distance of 1.5m from the user performing the task and mounted on a tripod at a height of 1.5 metres from the floor. The sensor provides both RGB and depth image frames at the rate of up to 30 frames per second. The RBG image is provided at a resolution of 640 x 480 pixels whereas the depth image is provided at a lower resolution of 320 x 240 pixels. Figures 3a and 3b show the use case setup for real-real and real-virtual collaboration scenarios respectively.

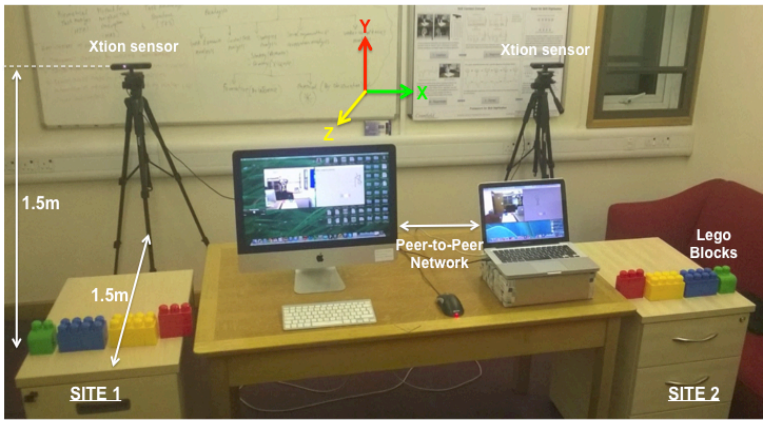


Figure 3a: The use case setup for real-real collaboration scenario

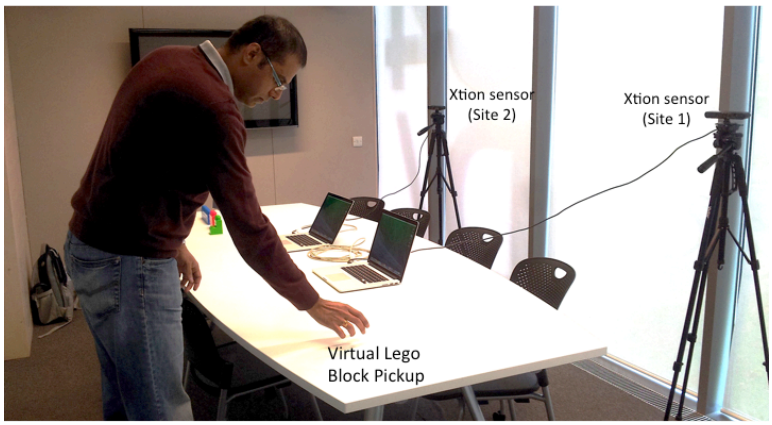


Figure 3b: The use case setup for virtual-real collaboration scenario

The network used for this use case was a peer-to-peer connection between the two PCs each belonging to a collaborating site. Physical connection between the PCs was achieved via a crossover Ethernet cable.

3.4 Data Capture Method

Two types of data are captured during the collaborative assembly task. These are the motion capture data of the human and the positon tracking data for each real or virtual Lego block during the assembly task at each site.

Human Motion Capture: Human motion capture is achieved by positioning the Xtion sensor to have a full view of the assembly workspace and the upper half of the user's body in order to simultaneously track skeletal motion and object manipulation. The standard java-based skeletal tracking library provided by OpenNI is used to obtain 3D positons of the human upper body joints such as head, neck, shoulders and arms in real-time. Therefore, this provides the DCM with a stream of digital (x, y, z) coordinates representing human motion involved in the assembly task at up to 30 times per second.

Lego Block Tracking: Along with human motion tracking, the Xtion sensors are used to simultaneously recognise and track the Lego blocks in the task scene. For this, pixels groups of red, green and blue colours are identified from within each RGB image frame and located using a simple edge detection mechanism on the screen space. Edge detection is performed by comparing the colour values of pixels belonging to each pixel group with the colour values of the background pixels. The centre point for each Lego block is representative of its spatial position and is obtained by calculating the midpoint of each pixel group in x and y-axis (Figure 4). Therefore, this provides the DCM with a digital stream of (x, y, z) coordinates of all the Lego blocks in the workspace at up to 30 times per second.

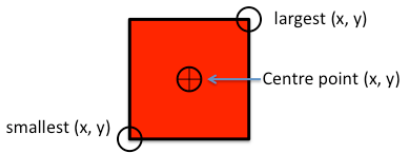


Figure 4: Red pixel group corresponding to a red Lego block and its centre point

Since the time stamps for human motion data and Lego block tracking data are the same, the human actions responsible for the manipulation of the Lego blocks during the assembly task can be identified. By mapping the user's left and right hand coordinates with the spatial positions of the Lego blocks, the nature of human actions during manipulation can also be captured. The software at either side of a real-real collaboration scenario performs this mapping and co-relation.

In real-virtual collaboration scenario, the spatial positions of the virtual Lego blocks are initially defined in the workspace and subsequently tagged with the user's hand coordinates whenever each hand is sufficiently close to a Lego block. In this work, the grasp and release gestures are not programmed therefore a virtual block is considered to be picked up when the user's hand coordinates are sufficiently near

that block and the block stays with the hand until the block coordinates change abruptly relative to the hand coordinates during the task.

This human motion and Lego block tracking data is sent to the DEM at the end of each image frame.

3.5 Data Transfer Method

The DEM receives the data from the DCM in the form of human skeletal coordinates and the spatial coordinates of the Lego blocks during the assembly 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 5).

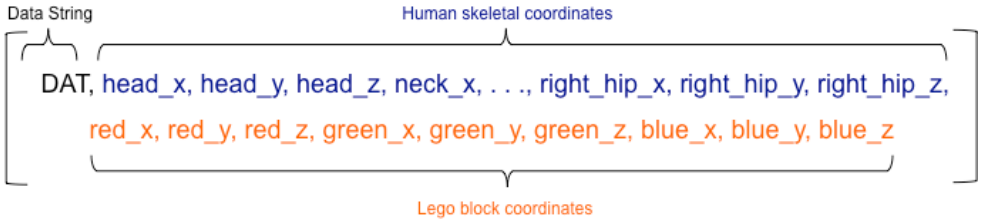


Figure 5: Data string representing the task information at any given instance

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 6.

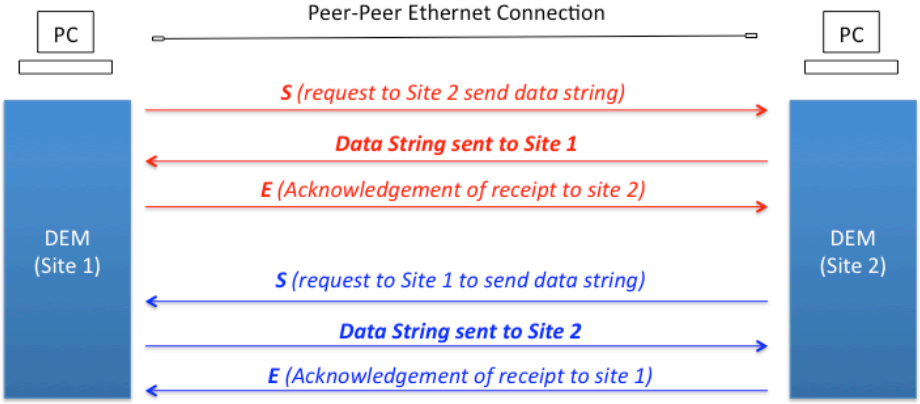


Figure 6: 2-way protocol for bidirectional data transfer

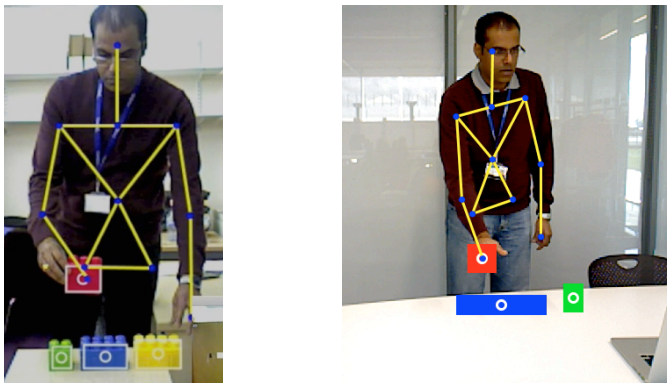
3.6 Data Rendering

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, 2D animation using OPENGL was used. The incoming data packet is parsed to extract human motion and Lego block position data. The user's skeleton is then rendered on screen using a stick figure corresponding to the spatial positions of the skeletal joints whereas the Lego blocks are rendered by positioning 2D rectangular blocks of the same size and colour at the corresponding spatial positions on the screen. A constant stream of such incoming data and the rendering of this data produce the effect of a live animation.

4. Results

4.1 Simultaneous tracking of human motion and Lego blocks

The results of the key components of the collaboration platform that is simultaneous human skeletal motion tracking and Lego block tracking are shown in Figure 7.



(a) (b)

Figure 7: Human upper body motion simultaneously tracked with (a) Real Lego blocks tracking and (b) Virtual Lego blocks tracking

The stream of data strings produced at one end of the collaboration gets parsed and rendered as an animation at the other end. The result of this process is shown in Figure 8.

Figure 8: DEM rendering the live animation from the data string

4.2 Data flow during Collaboration

The flow of data during the collaboration session (Figure 9) is according to the software architecture (section 3.1) of the real-time remote collaboration platform.

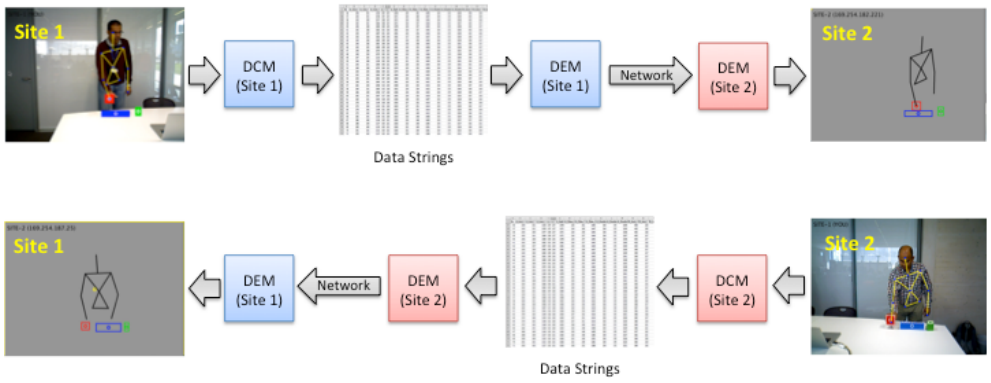


Figure 9: Flow of data during collaboration

4.3 Real-Real Collaboration Scenario

In this scenario, the Red, Green and Blue Lego blocks placed on the workstation are collaboratively assembled in the formation shown. This activity data is constantly streamed in real-time to the other site where it is rendered as an animation as shown in Figure 10.

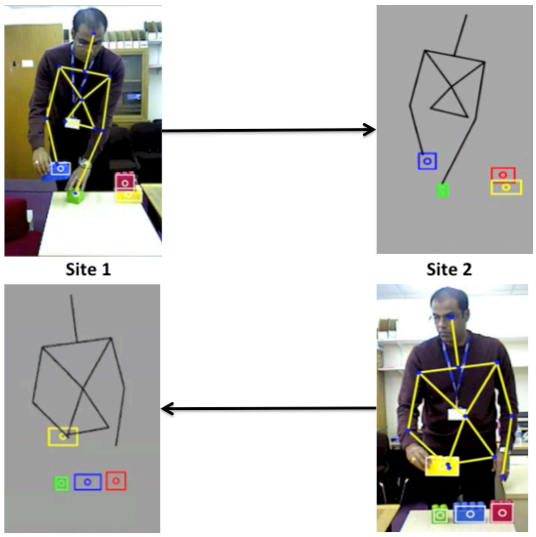


Figure 10: Real-real collaboration scenario

4.4 Virtual-Real Collaboration

A virtual-real collaboration session does not need both sites to have access to the physical parts. This helps reduce training costs and increase collaboration accessibility. Figure 11 shows an example where the trainer (site 1) assembles virtual Lego blocks to remotely demonstrate to the learner (site 2) who assembles real Lego blocks according to the process he observes in the live animation.

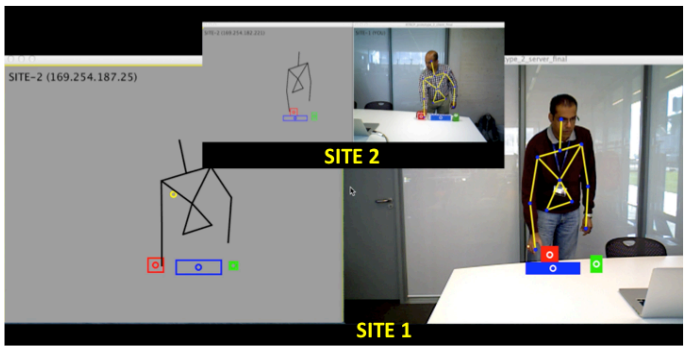


Figure 11: Virtual-real collaboration scenario

Note that in both the scenarios, the user demonstrating the Lego block assembly is also able to obtain live feedback on how the other user is performing on the task through the animation. Therefore, immediate corrective steps can be taken by the users to resolve problems as and when they occur during the task.

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 (Tzeng and Huang, 2012). 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.

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. This work uses the first generation of gaming interface sensors, which also bring with them some disadvantages. The skeletal tracking function degrades considerably as the level of occlusion increases between itself and the user being observed. In this work, only the human upper body could be tracked because of the presence of the assembly table and occasionally skeletal tracking data was missing. This could have a significant impact especially if the tracking is lost during critical moments of a collaborative task. Also, depth imaging requires the human to be in the constant line-of-sight of the sensor to track motion. This limits the kind of tasks that could be digitised to enable collaboration. Finally, the resolution of the depth and RGB imaging cameras used in this work are relatively lower than their superior industrial counterparts available at much higher

costs. Therefore the complexity of objects that can be tracked during collaboration is limited to simple straightedge geometry and primary colours.

In the future, this work plans to investigate the use of the second generation of Kinect sensors that claim to offer better skeletal tracking due to a change in the underlying motion capture technique. Object detection capability will also be enhanced due to the higher resolutions of the RGB and depth imaging cameras (1920 x 1080 pixels and 512 x 424 pixels respectively). The use of new-age visualisation devices such as Oculus Rift is also planned in order to introduce immersive collaboration experience created using game design engines such as Unity3D. Also, the use of multiple Kinect sensors to capture the same task will be investigated to reduce the line-of-sight constraints that currently exist within a single sensor setup. Validation of the proposed collaboration platform will be conducted using industrial real-life cases.

The positive impact of collaborative working cannot be emphasised more 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. The platform also enables 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 also means that the same complex part could be worked upon by multiple teams at the same time for verifications and inspections.

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. Exchange of knowledge and skills and transfer of expert services over a network is enabled using digitised human-workpiece interactions and context sharing. The use of simple data forms to reliably digitise and exchange rich information makes the collaboration platform independent of network bandwidths. The initial results of this work with collaborative Lego block assembly in real-real and real-virtual scenarios are encouraging and further research is planned to expand the scope of the platform. The potential use of real-time remote collaboration spans across industries where communication and

collaboration 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 (formerly the Technology Strategy Board) for funding this research as part of the technology-inspired innovation - collaborative R&D in ICT - grant 101774 (August 2013).

References

Adcock, M., Anderson, S. and Thomas, B. (2013), "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,* ACM, pp. 235.

Kogut, B. and Kulatilaka, N. (1994), "Operating flexibility, global manufacturing, and the option value of a multinational network", *Management science,* vol. 40, no. 1, pp. 123-139.

Kurillo, G. and Bajcsy, R. (2013), "3D teleimmersion for collaboration and interaction of geographically distributed users", *Virtual Reality,* vol. 17, no. 1, pp. 29-43.

Mossel, A., Schönauer, C., Gerstweiler, G. and Kaufmann, H. (2013), "Artifice-augmented reality framework for distributed collaboration", *International Journal of Virtual Reality,* vol. 11, no. 3, pp. 1-7.

Piumsomboon, T., Clark, A., Umakatsu, A. and Billinghurst, M. (2012), "Poster: Physically-based natural hand and tangible AR interaction for face-to-face collaboration on a tabletop", *3D User Interfaces (3DUI), 2012 IEEE Symposium on,* IEEE, pp. 155.

Prabhu, V. A., Tiwari, A., Hutabarat, W. and Turner, C. (2014), "Monitoring and Digitising Human-Workpiece Interactions during a Manual Manufacturing Assembly Operation Using Kinect™", *Key Engineering Materials,* vol. 572, pp. 609-612.

Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P. and Maciocci, G. (2013), "BeThere: 3D mobile collaboration with spatial input", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* ACM, pp. 179-188.

Tecchia, F., Alem, L. and Huang, W. (2012), "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, ACM, pp. 323-328.

Tzeng, G. and Huang, C. (2012), "Combined DEMATEL technique with hybrid MCDM methods for creating the aspired intelligent global manufacturing & logistics systems", *Annals of Operations Research,* vol. 197, no. 1, pp. 159-190.