Evaluating an Augmented Remote Assistance Platform to Support Industrial Applications

Mark Rice[1], Keng-Teck Ma[1], Hong Huei Tay[1], Joyce Kaliappan[1,2], Wei Ling Koh[1,2], Wah Pheow Tan[2], Jamie Ng[1]
[1] Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way, Singapore, 138632
[2] Temasek Polytechnic, 21 Tampines Avenue 1, Singapore, 529757
Email: {mdrice; makt}@i2r.a-star.edu.sg

***Abstract*—Remote assistance provides a communication bridge for users engaged in different locations. However, understanding how to design such systems in IoT is a challenging issue given digital representations are not the same as sharing a physical space. In this paper, we present a Remote Assistance Platform (RAP) that is designed to facilitate task guidance between an instructor and one or more remote operators. This includes the support of visual communication using annotation tools that augment information from a live video stream. Two user studies were performed to evaluate co-located and remote interaction. In the first study, dyads interacted with paper-based instructions while situated in the same location. In the second study, different dyads remotely performed the same tasks, assisted by using a smartphone or smart glass display. Overall, our findings found significant differences in communication behaviour based on the type of collaborative environment and information modality used. A short review of these results is discussed.**

Keywords—human-computer interaction; remote assistance; head-mounted display; communication behaviour; industrial IoT

I. INTRODUCTION

The term *remote assistance* offers the ability to collaborate at a distance to other end users via a digital network. Reflective of modern day changes in the working practices and organisation of businesses, remote communication systems have the potential to help reduce operational costs, train and develop skills, and analyse performance and resources. In particular, the ability to remotely solve problems is seen to be important in sectors such as mining and agriculture that may lack local amenities [1], while advanced manufacturing is driven towards ubiquitous solutions that can access operations across different geographical locations [2].

According to the World Economic Forum, over the coming years, Industrial IoT will play an important role in augmenting the workforce [1]. As part of this infrastructure, wearable and head-mounted displays are increasingly being utilised in industries such as aerospace, healthcare and logistics to connect and streamline workflow processes. Likewise, a growing number of commercial applications are targeted towards facilitating aspects of remote support, such as for maintenance or technical servicing [e.g. 3, 4, 5, 6].

In turn, digital remote assistance presents a number of important human factors challenges in the absence of face-to-face communication. Not only in terms of how systems can help enable the matching of mental representations to procedural tasks, but also facilitate attention and awareness outside a user's field-of-view. This includes being able to successfully mediate interaction from a limited camera viewpoint, or display resolution.

Building on our prior work [7], in this paper we report on the performance of a *Remote Assistance Platform* (RAP). This platform allows for the communication between two or more users over the Internet, or via a local Wi-Fi connection. A remote user wears a pair of smart glasses or a hand-held smartphone with a forward-facing camera to capture the input video, which is transmitted to a web-based application that an expert/instructor can see. A set of tools in the web-based application then enables the expert to convey instructions.

Two user studies were conducted to understand the dyadic interaction between what we describe as an *instructor* and *operator*. In the first study, we investigated the interaction between pairs in the same co-located workspace using paper-based instructions. In the second study, we then compared their remote interaction using a smartphone and a pair of smart glasses. Primarily focusing on the human aspects of engaging with a remote assistance system, the main contributions of this paper are two-fold: 1) to report on the dyadic behaviour in co-located and remote interaction, and 2) to analyse and compare these differences in relation to the collaborative environment and information modality used. We perceive the novelty of this work in reporting on the task performance and the communication differences identified between the two studies, as we hope this research will generate interest for practitioners working in remote guidance systems and broader application areas of IoT.

II. RELATED WORK

Situational awareness can be described as a clear understanding of an environment, with the ability to translate information to (near) future actions [8]. According to Kraut et al. *"visual information can help people communicate about the task, by aiding conversational grounding, or the development of mutual understanding between conversational participants"* [9, p. 15]. In environments where users are co-located, situational awareness is attained via a wealth of visual information (including body language), compared to remote collaboration, where associated cues are largely dictated by the technology used [9]. Thus, understanding how visual information can direct attention in the completion of

procedural tasks is an important aspect of helping to facilitate remote communication.

In remote assistance systems, visual annotations are one solution that have been reported to enhance the information shared, improve the identification of objects in the workplace, and reduce the amount of verbal communication needed to express instructions [10]. As such, the types of visual annotations studied have varied from the use of screen-based drawings [e.g. 10, 11, 12, 13], to laser guided projection [14]. Fussell et al. [10] found that annotations were most frequently used to highlight locations and objects in a scene. On the other hand, it has been reported that graphical annotations can have a negative effect in cluttering the display [11], with freehand drawings difficult to interpret due to poor legibility [12], and annotating directly onto paused video and snapshot images disorienting when returning back to a live scene [11, 13].

In terms of the development of related systems, nearly 30 years ago, Tang and Minneman [15] investigated a two-way drawing tool using whiteboard markers and video cameras to capture gestures that were visualised to a second party. Over a decade later, gesture-based interaction was further investigated on video displays [10]. While more recently, a few studies have compared hand-held and wearable devices. For example, Johnson et al. [16] found the use of a head-mounted display performed better than a tablet when engaged across multiple workstations, and Fakourfar et al. [13] reported an advantage in facilitating hands-free interaction. The roles of remote assistance have also been investigated to support assistive needs [17, 18], while other studies have explored performance criteria for remote assistance systems [19], but lack empirical comparisons in user interaction.

Yet, despite these works, research in remote assistance has not been exhaustively investigated. In some ways, this is unsurprising given the range of potential uses in industrial applications. For example, consider the potential of remote assistance systems to assist in the re-training of the older workforce, and the virtual monitoring of systems and processes at a distance to the shop floor [2]. At the same time, while there are recognised opportunities in industrial IoT for novel technologies to help bridge skill gaps in workers [20], relatively few known studies have empirically reported cross-device comparisons. Therefore, a strong motivational factor for undertaking this work is to understand the human components of developing a remote assistance platform, in leading to better user experience, and future commercial deployment.

III. The Communication Platform

As previously described [21], our system architecture consists of a signalling server, a Kurento media server that implements the WebRTC communications protocol [22], and a media exchange server (see Figure 1). Using either a camera-enabled smartphone or head-mounted display, a field operator captures an input scene. In real-time, the video data generated is streamed to a web-based application to be viewed by an expert. Using this application, annotations are added to snapshot images, which are extracted from a live video view, before being sent to the operator's device to assist in task completion.

Devices are typically connected to the platform using Wi-Fi, although a workable alternative is to use a SIM-compatible device on a fast 4G network. Video streams by default are encoded from each end point using a VP8 format, while the platform sends and receives data using pre-defined JSON commands. Communication can be established by any end-point, although it is generally invoked from devices where an operator requires assistance. These commands are sent using web socket connections to a signalling server to request resources, such as URLs to media, as well as to establish a video link to the platform.

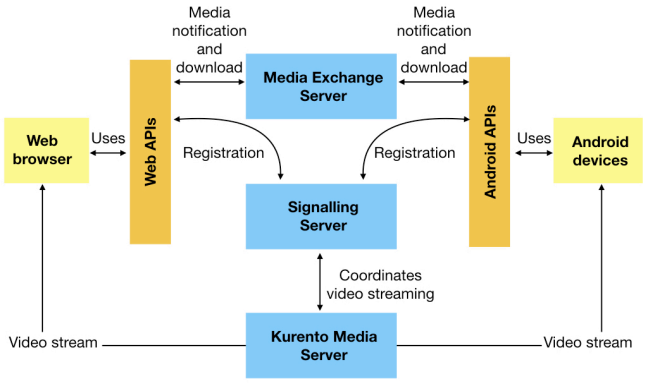

Fig. 1. Overview of the platform structure.

A. Device integration

The RAP platform is designed to accommodate for information based on different display sizes of mobile and wearable devices. The UI and UX aspect of the platform takes into account available control schemes that may be hardware specific. For example, the Epson Moverio BT-200 has a tethered controller with a touchpad, while the Vuzix M300 uses a touchpad with control buttons on the headset, and the Microsoft HoloLens, voice and gesture input. This is achieved by abstracting core protocol implementation away from the UI elements based on a design pattern of the 'model-view-controller'. The resulting client components and their respective input can then be freely re-designed to suit different supporting devices, while maintaining a consistent model to execute functions on the platform.

B. User interfaces

The user interfaces are developed in Google Android. Specifically, the *instructor interface* is a web-based application, displayed on a desktop or laptop with a built-in or separate web camera. The UI consists of a live video stream sent via the remote operator's device camera. Right clicking on the video creates a snapshot image that is automatically displayed in a separate window (Figure 2). Separating these views is designed to reduce interference when watching the video stream. The instructor can then annotate this information by inserting symbols, text and freehand drawings, and dragging and dropping from a set of customised icons representative of the tasks performed. These objects can be replaced, moved, or deleted. Sizes and colours can also be changed. When ready,

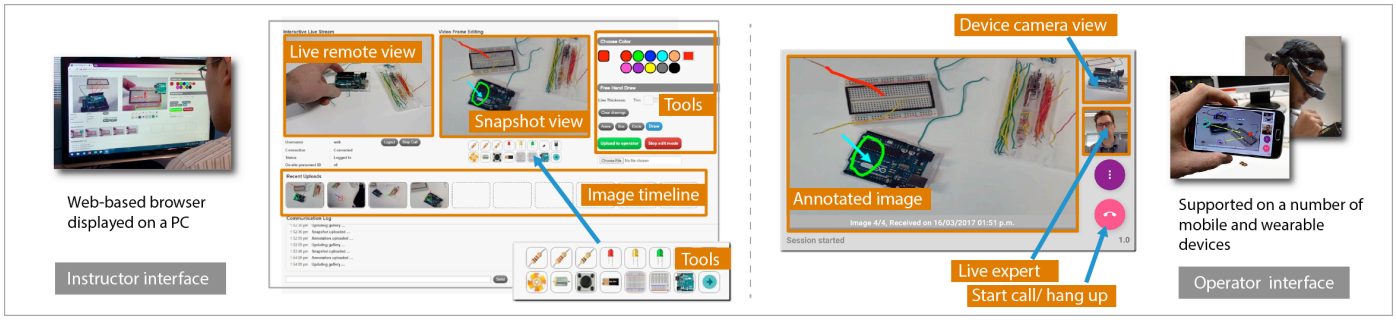

Fig. 2. Illustration of the RAP interfaces. (*Left*), the instructor interface, and (*right*), the operator interface for hand-held and head-mounted displays.

this information is uploaded to the operator as an annotated image.

Alternatively, for the *remote operator interface,* interactive features are divided in three areas: a main window containing the annotated image sent from the instructor, a live view of the instructor (or expert), and a camera view from the device. Swiping left or right on an annotated image allows for the toggling between different images received, which are sequentially ordered and time stamped.

IV. Study 1 - Co-Located Interaction Using Paper Instructions

A baseline study was first conducted to identify the ability of dyads to complete two assembly tasks, unaided by the use of any assistive technology. A total of 20 participants (10 males and 10 females) with an average age of 21 years were recruited. Participants were randomly assigned to either be an instructor or operator. As the names may suggest, the instructor's role was to guide the completion of the tasks, while the operator, to assemble them.

A. Tasks

The Arduino platform was used for the tasks [23]. Arduino offers the ability to combine electrical components in a number of interesting ways, and learn basic features quickly. The tasks required multiple parts: micro-controllers, wires, diodes and resistors, and a laptop to run the pre-loaded software. Instructions were designed to walk the instructor through the procedural steps, with visual examples given. Specifically, the two tasks were a *construction* task that involved building a pinwheel circuit, and a *troubleshooting* task that required re-connecting three misaligned wires, and three resistor errors in a traffic light circuit.

B. Procedure and data analysis

To begin, the instructor was given 15 minutes to familiarise themselves with the tasks. Once ready, pairs were seated next to each other, and asked to complete the two tasks, the order of which was counterbalanced in the study. During this time, the instructor was able to give verbal instructions, and physically reference by pointing to objects in the workspace (Figure 3). However, they were not allowed to show instructions to the operator, or physically manipulate the objects themselves. Pairs were given 20 minutes to complete each task. If they failed to complete one or more of the tasks within the time allocated, these were marked as incomplete.

All sessions were video recorded. Error counts and task completion times were gathered from two independent researchers. Given the sample size, non-parametric tests were used for statistical analysis.

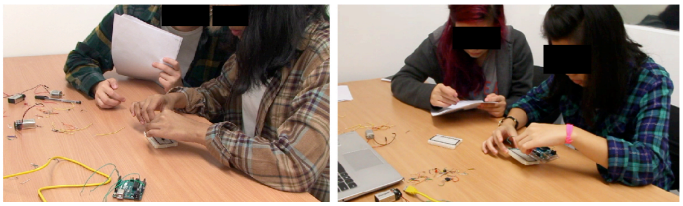

Fig. 3. Workspace set-up and illustrations of the paired interaction.

C. Results

Overall, we identified a significant difference in the task completion times, $z(n = 10) = -2.60$, $p < .01$, with the troubleshooting task ($Mdn = 247$ sec) taking less than half the time to complete compared to the construction task ($Mdn = 670$ sec). In contrast, for assembly errors, while the number of instances was higher for the construction task, this was statistically non-significant, $z(n = 10) = -1.28$, $p > .05$ (Figure 4).

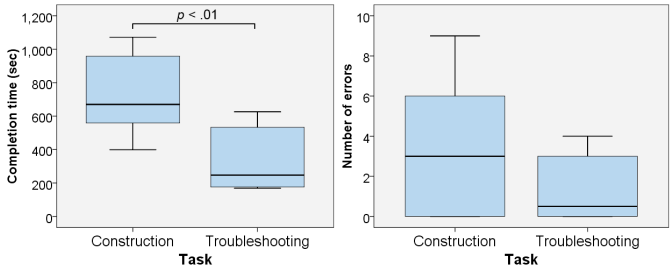

Fig. 4. Boxplots of the completion time and error rate for each task.

To account for these differences, a review of the video data indicated that the construction task required more precision and search time to locate and assemble individual components compared to the troubleshooting task, where the pre-positioning of electrical components on the breadboard acted as useful reference points to orientate towards. For example, in the construction task, to place an electrical switch, the operator had to first determine the orientation of the pins, which had to be precisely aligned to the circuit. Alternatively, in the troubleshooting task, they only needed to verify that the fitting was secure. Despite these differences, all participants completed the tasks in the time allocated.

V. Study 2: Remote Interaction Using Hand-held And Head-mounted Displays

Having identified participant abilities to complete the assembly tasks, a second study was undertaken to evaluate the RAP platform. Specifically, the study focused on comparing task performance across two different display modalities - a *smartphone* and a pair of *smart glasses*, to understand if they had any measurable impact on user interaction. Using a between-subjects design, a total of 40 participants (25 males and 15 females), with an approximate age of 29 years were recruited. Randomly assigned, for the operator, this consisted of using one of the display technologies. None of the participants were involved in the previous study, with no known experience using Arduino.

A. Tasks and devices

The two experimental tasks were the same as the first study, with an additional practice task that required the subject pairs to build a simple circuit with a single LED light. For the mobile device, a commercial smartphone was used, consisting of a 2560 x 1440 screen resolution, and 16 MP camera. In contrast, a pair of see-through binocular glasses was used for the head-mounted display, operated via a track pad, with a 960 x 540 display resolution, and VGA camera (Figure 5).

B. Procedure and data analysis

After giving their informed consent, both the instructor and operator were separately briefed on how to use the interfaces. Once they acknowledged their understanding of the interface usage, each pair completed a practice task together. To provide a realistic set-up, participants interacted in separate rooms, using headphones to ensure audio information was clearly understood. Tasks were counter-balanced, with a similar cut-off time of 20 minutes to the baseline study. During this time, the instructor could refer to, but not show the paper instructions to their partner. All sessions were video recorded, and non-parametric tests were largely used for statistical analysis. For those participants who did not complete the tasks in the time allocated, they were still counted in the analysis under the maximum time limit allowed.

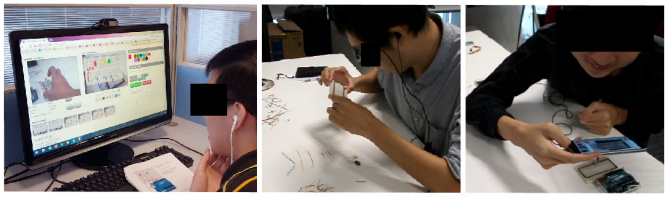

Fig. 5. Workplace set-up. (*Left*) instructor with display, (*middle*) operator wearing a pair of smart glasses, and (*right*) operator using a smartphone.

C. Results

In examining the effect of display-type on task *completion time*, no significant difference was found, $U(n_1 = 10, n_2 = 10) = 32$, $p > .05$. Only one participant in each condition completed the construction task in the time allocated, compared to all the participants completing the troubleshooting task in time. For task errors made, a few types were found. These included identifying wrong components and their placement on the breadboard for the construction task, and missing out procedural steps in the troubleshooting task. However, the type of display did not have a statistical effect on the number of errors made, $U(n_1 = 10, n_2 = 10) = 38$, $p > .05$ (Figure 6).

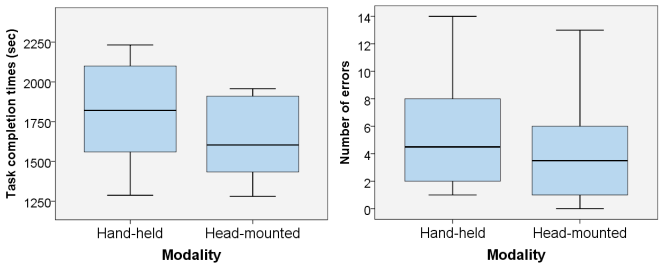

Fig. 6. Boxplots of the combined task completion times and error rates comparing hand-held (smartphone) and head-mounted (smart glass) displays.

Based on a review of the video data, a lack of statistical differences in display performance, and problems completing the arguably more complex construction task, appear to be attributed to hardware, software or practical limitations in the devices used. For the smart glasses, low video quality was a noticeable constraint in completing the tasks. Similarly, the camera's head-mounted position was often not aligned with a user's viewing perspective. This resulted in participants having to adjust their head position to capture an aspect of the task for the instructor to see, before re-adjusting back to complete an assembly. In contrast, for the smartphone, limitations were identified in two areas. First, the need to hold the smartphone with one hand, which slowed down the task. Second, poor camera auto-focus at a close distance, causing problems in seeing board components, and obtaining a clear image from which the instructor could modify. As a result, several design suggestions were provided to improve the user interaction. These included the following:

- Ensure the use of a high-resolution camera, or include post-image processing to enable the viewing of fine detail.
- Consider the deployment of a second camera to provide a more holistic view of the operator's actions, with options to switch views.
- Provide a highlighting tool that allows the instructor to pinpoint areas of interest on the live video stream.
- Enable the operator to annotate information back to the instructor.

In facilitating the two tasks, on average, a similar number of visual annotations were generated for the smartphone (M = 26.9, $S.D.$ = 13.9) and smart glass (M = 28.4, $S.D.$ = 17.4) conditions. Specifically, annotations were either used to illustrate the *placement* or the *identification* of objects on the breadboard. Although statistically non-significant between conditions ($p > .05$), *placement* annotations (*smartphone*, M = 20.9, $S.D.$ = 12.8; *smart glass*, M = 21.0, $S.D.$ = 17.0) were used more often than *identification* annotations (*smartphone*, M = 6.0, $S.D.$ = 1.9; *smart glass*, M = 7.4, $S.D.$ = 3.2).

Further, the use of annotations was incremental in the tasks. Multiple variations were often sent using the same snapshot image. This avoided the need for unnecessary repetition.

Commonly, lines were colour coded to replicate the wires used, freehand dots and arrows for placement positions, and graphical objects to illustrate what to connect too. Variations in these drawing styles are illustrated in Figure 7.

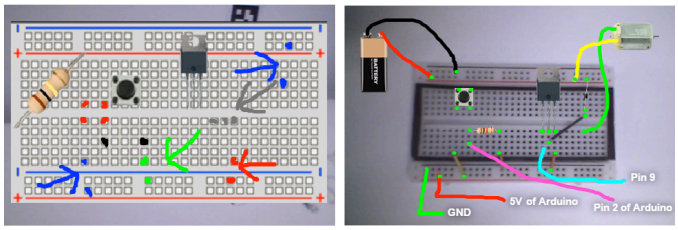

Fig. 7. Examples of the annotations created. Notice the difference in style, particularly the use of a graphical object to depict the breadboard, compared to using a captured image through the camera, as well as arrows compared to wire lines to illustrate their placement.

VI. Study Comparisons

For additional analysis, we compared the quantitative results from the two studies to get an overview of the potential differences between the co-located and remote interaction.

In examining the effects of collaboration-type for task completion times, *co-located* interaction (*Mdn* = 1135.5 sec) was significantly quicker than *remote* interaction (*Mdn* = 1763 sec), $U(n_1 = 20, n_2 = 10) = 4$, $p < .001$. Similarly, in comparing the task completion times across the type of modality used, a significant difference was identified, $H(2) = 18.67$, $p < .001$. Namely, completion times took significantly longer using the *smartphone* ($U(n_1 = 10, n_2 = 10) = 2$, $p < .001$) and *smart glass* ($U(n_1 = 10, n_2 = 10) = 2$, $p < .001$) compared to the *paper only* instructions (Figure 8). Alternatively, there were no statistical differences in the number of errors made across the conditions ($p > .05$).

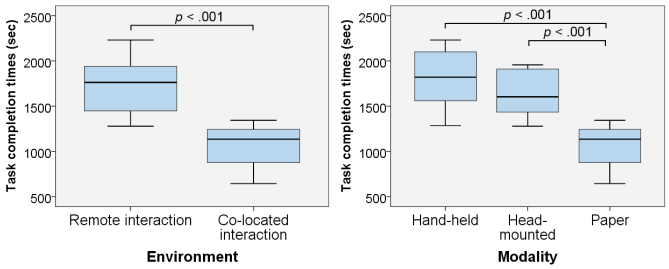

Fig. 8. Boxplots comparing the combined task completion times for co-located and remote interaction, and for paper only instructions, hand-held (smartphone) and head-mounted (smart glass) displays.

To further understand these differences, we reviewed actions related to the collaboration between the instructors and operators. Based on video analysis, these actions were enacted for task explanation and clarification, and to understand the attentional direction towards specific areas of the breadboard. In summary:

- Co-located instructors and operators tended to employ verbal communication more exclusively for task explanation and clarification, while instructors in the remote interaction also employ verbal communication for directing attention. However, verbal communication was rarely used for directing attention in co-located interaction as the instructor could readily direct the attention of the operator through pointing gestures.
- In the co-located condition, directing attention was often performed in synchrony with a task explanation (i.e. an instructor pointed to the relevant area while they explained the task). Comparatively, these two functions were often performed asynchronously in the remote condition, with task explanation occurring sequentially after directing attention.
- While the RAP platform allowed for a live-view of the instructor, this was often not sufficient to allow high quality non-verbal communication to be transmitted. For example, in the co-located condition, instructors often demonstrated how the task should be completed through gesturing (i.e. without touching the breadboard). This non-verbal form of information was naturally absent from the remote condition.

VII. Discussion

In relation to digital remote assistance, our findings support recent arguments for user-centred approaches in IoT [24], as we identified how environments, interactions and technologies play important roles in adoption and use. Drawing a number of parallels to the earlier work of Kraut et al. [9] and Fussell et al. [10], we found the ability to monitor actions improved when located side-by-side. Like Kraut et al. [9], we identified that limitations in the viewing of visual information in remote assistance imposed time and conversational resources.

To help explain differences in the collaborative interactions observed, one possible explanation is the multiple resource model posited by Wickens [25]. That is, in both co-located and remote conditions, task explanation requires cognitive resources from the instructor. Directing attention via pointing (in the co-located condition) draws from spatial cognitive resources, while doing so via annotations (in the remote condition) draws from verbal cognitive resources. As each type of resource is limited [25], the instructor may be unable to annotate to direct attention, while explaining the task concurrently, as both actions are competing from the same pool of processing resources. Subsequently, the asynchronous nature of both attentional direction and task explanation in the remote condition may have affected the quality of interaction between the dyads, which is reflected in the slower task completion time in the remote condition.

Thus, our results reinforce the importance of ensuring a clear representation of the task environment between the operator and the instructor. Segmented and low fidelity representations in a wearable or mobile display will undoubtedly mismatch with real-world scene representations. Examples of this included the low image quality captured by the smart glass camera, and problems auto-focusing with the smartphone at a close proximity to objects. These problems appear to be exacerbated by working with small electrical components that require good image resolution, and highlight the importance of testing across a range of wearable and hand-held devices to find the optimal hardware solution.

In addition, as other researchers have commented on user strategies to accommodate for viewpoint limitations in remote assistance [e.g. 16], we believe it would be useful to establish a set of human factors design guidelines that can help match device capabilities with task and environmental requirements of the workplace. In particular, our results indicate that task stimuli and environment representation should leverage on technological products that can create an immersive and high-fidelity representation.

Regarding the visual annotations created, they benefited in helping to discriminate from uniformed features that can take time to verbally describe. The drawing gestures were often used as specific placement markers, and flexibility in how they are utilised is an important aspect of the platform, given the different ways the users articulated procedural steps and instructions. Consequently, we believe the inclusion of design tools need to accommodate for these different styles of drawing, including giving the remote operator more control in the capture and manipulation of content that can be directed back to an expert. Given the different ways these annotation tools could be used in industrial IoT, this raises further questions in how annotations are to be intuitively drawn by an operator when using a wearable display.

To extend the platform further, there is a possibility of combining with other types of sensory data for diagnostics, decision-making and event tracking. However, as our system is primarily visual, we are particularly interested in developing more *visual intelligence* through the deployment of machine learning and computer vision algorithms that can interpret aspects of the video stream. Recent studies have successfully combined eye tracking with smart glass displays to measure aspects of cognition [26], in addition to demonstrating how unsupervised learning approaches can detect scene objects from egocentric videos [27]. Therefore, being able to automate the detection of task errors, and infer an operator's attentional demands from a first-person view, raises the question over the extend machine intervention may substitute, or help reinforce the role of the expert in inspection and field services - particularly for procedural steps that require repetitive actions from which a computational model could systematically learn.

Finally, as none of the participants in this study were experts in a domain area, we acknowledge that in the real-world, tasks may be far more complex, involve more users, and potentially operate over a much larger workspace. Future studies would therefore benefit from focusing on specific domain areas (e.g. in MRO, transportation, logistics, etc.), using use cases that can evaluate the platform in practical settings, and gather representative data in situations they may be deployed for.

REFERENCES

[1] World Economic Forum, "Industrial internet of things: unleashing the potential of connected products and services," 2015 [Online]. Available: http://reports.weforum.org/industrial-internet-of-things/

[2] European Commission, "Factories of the future. Multi-annual roadmap for the contractual PPP under Horizon 2020," European Union, 2013.

[3] Upskill, 2017 [Online]. Available: https://upskill.io/

[4] Scope AR, 2017 [Online]. Available: http://www.scopear.com

[5] Proceedix, 2017 [Online]. Available: https://proceedix.com

[6] Fieldbit Ltd, 2017 [Online]. Available: https://fieldbit.net

[7] M. Rice, S. C. Chia, H. H. Tay, M. Wan, L. Li, J. Ng, and J. H. Lim, "Exploring the use of visual annotations in a remote assistance platform," in *Proc. of CHI EA '16*, ACM, pp. 1295-1300, 2016.

[8] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, issue 1, pp. 32-64, 1995.

[9] R. E. Kraut, S. R. Fussell, and J. Siegel, "Visual information as a conversational resource in collaborative physical tasks," *Human-Computer Interaction*, vol. 18, issue 1-2, pp. 13-49, 2003.

[10] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. I. Kramer, "Gestures over video streams to support remote collaboration on physical tasks," *Human-Computer Interaction*, vol. 19, issue 3, pp. 273-309, 2004.

[11] S. Kim, G. Lee, N. Sakata, and M. Billinghurst, "Improving co-presence with augmented visual communication cues for sharing experience through video conference," in *Proc. of ISMAR '14*, IEEE, pp. 83-92, 2014.

[12] V. Domova, E. Vartiainen, and M. Englund, "Designing a remote video collaboration system for industrial settings," in *Proc. of ITS '14*, ACM, pp. 229-238, 2014.

[13] O. Fakourfar, K. Ta, R. Tang, S. Bateman, and A. Tang, "Stabilized annotations for mobile remote assistance," in *Proc. of CHI '16*, ACM, pp. 1548-1560, 2016.

[14] D. Palmer, M. Adcock, J. Smith, M. Hutchins, C. Gunn, D. Stevenson, and K. Taylor, "Annotating with light for remote guidance," in *Proc. of OzCHI '07*, ACM, pp. 103-110, 2007.

[15] J. C. Tang, and S. L. Minneman, "VideoDraw: a video interface for collaborative drawing," in *Proc. of CHI '90*, ACM, pp. 313-320, 1990.

[16] S. Johnson, M. Gibson, and B. Mutlu, "Handheld or handsfree? Remote collaboration via lightweight head-mounted displays and handheld devices," in *Proc. of CSCW '15*, ACM, pp. 1825-1836, 2015.

[17] M. Avila, K. Wolf, A. Brock, and N. Henze, "Remote assistance for blind users in daily life: a survey about be my eyes," in *Proc. of PETRA '16*, ACM, article no. 85, 2016.

[18] S. Ikeda, Z. Asghar, J. Hyry, P. Pulli, A. Pitkanen, and H, Kato, "Remote assistance using visual prompts for demented elderly in cooking," in *Proc. of ISABEL '11*, ACM, article no. 46, 2011.

[19] M. Schneider, J. Rambach and D. Stricker, "Augmented reality based on edge computing using the example of remote live support," in *Proc. of ICIT '17*, IEEE, pp. 1277-1282, 2017.

[20] E. Ras, F. Wild, C. Stahl, and A. Baudet, "Bridging the skills gap of workers in industry 4.0 by human performance augmentation tools: challenges and roadmap," in *Proc. of PETRA '17*, ACM, pp. 428-432, 2017.

[21] W. L. Koh, J. Kaliappan, M. Rice, K-T. Ma, H. H. Tay, and W. P. Tan, "Preliminary investigation of augmented intelligence for remote assistance using a wearable display," in *Proc. of TENCON '17*, IEEE, in press.

[22] WebRTC, 2017 [Online]. Available: https://webrtc.org

[23] Arduino, 2017 [Online]. Available: https://www.arduino.cc

[24] A. Soro *et al*., "Designing the social internet of things," In *Proc of CHI EA '17*. ACM, pp. 617-623, 2017.

[25] C. D. Wickens, "The structure of attentional resources," in R. S. Nickerson (ed.), *Attention and Performance VIII*, Hillsdale, NJ, Lawrence Erlbaum, pp. 239-257, 1980.

[26] K. Essig, B. Strenge, and T. Schack, "ADAMAAS: towards smart glasses for mobile and personalized action assistance," in *Proc. of PETRA '16*, ACM, article no. 46, 2016.

[27] D. Damen, T. Leelasawassuk, and W. Mayol-Cuevas, "You-Do, I-Learn: egocentric unsupervised discovery of objects and their modes of interaction towards video-based guidance," *Computer Vision and Image Understanding*, vol. 149, pp. 98-112, August 2016.