2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)

Comparing Different Augmented Reality Support Applications for Cooperative Repair of an Industrial Robot

Doris Aschenbrenner* TU Delft IDE
Michael Rojkov† University of Würzburg HCI-Group
Florian Leutert‡ Zentrum für Telematik
Jouke Verlinden§ TU Delft IDE
Stephan Lukosch¶ TU Delft TBM
Marc Erich Latoschik‖ University of Würzburg HCI-Group
Klaus Schilling** Zentrum für Telematik

Abstract

Digitization and the growing capabilities of data networks enable companies to perform tasks via remote support, which previously required service personnel to travel. But which mixed reality method leads to better results regarding human factors, grounding and performance criteria? This paper reports on a collaborative user study, in which a local worker is guided by a remote expert with the help of different augmented reality methods, specifically see-through HMD, spatial projection, and video-mixing tablet. The task to perform is the exchange of a controller in a switch cabinet of an industrial robot, a task rather typical for failure detection within the field. Our study was conducted in collaboration with a technician school, where 50 technician apprentices participated in our study. Our results show clear advantages of using augmented reality (AR) versus traditional conditions (audio, video, screenshot) to enable remote support. It further gives significant indications for using a projection based AR method.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Treemaps; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

Computer-based interconnections of remote physical places enable us to communicate, interact, and collaborate across huge distances. Computer-supported collaborative work (CSCW) provides new collaboration opportunities for typical office work as well as for industrial production settings. Yet, there are still many open questions about typical usability aspects like efficiency, effectiveness, and user experience for the vast options of practical approaches and paradigms applicable in CSCW. In this article we investigate potential applications of different Mixed and Augmented Reality (MR and AR) techniques in tele-maintenance scenarios of industrial production lines. Specifically, we focus on remote support for maintenance and repair operations of industrial robots carried out in plants with active production lines.

Repairing an industrial robot is a difficult task: as in each complex computer-supported hardware system, problems and failures can be caused by faulty mechanical parts, electronics, or software. Active production lines typically run 24h a day and 7 days a week. Breakdowns are very expensive and hence the detection, analysis, and repair of the failure is constrained by time pressure and hence is costly [6]. As a result, it seems highly reasonable and desirable to support maintenance personnel during the aforementioned tasks.

A lot of research is targeting novel support methods based on various Mixed Reality paradigms to support repair or assembly tasks with only the worker in the loop [14]. Situated information, i.e., properly localized, spatially-anchored and contextualized information as provided by AR is particularly highly advantageous during complex tasks. Having all necessary information directly available during the required operation and spatially blended on the respective target of operation frees the personal at the site from potentially distracting media changes as caused by documentation look-ups or support calls, hence facilitating the overall maintenance process.

Still, there is a vast design space of potential interfaces as well as different contextualization and visualization techniques. For one, we cannot safely assume a *one-size-fits-all* approach for the remote experts as well as for the local staff. Each will exhibit different requirements, e.g., caused by specific ergonomic constraints w.r.t. the appropriate hardware interfaces or particular technical limitations at the respective work places. In prior work we mostly covered aspects of the remote expert side [1]. This paper reports on the development and evaluation of three different AR-methods targeting the local collaborator at the production site who is assisted by a remote expert. The three systems are based on (1) tablet-based AR, (2) optical see-through AR, and (3) projection-based AR. We compare these approaches with typical usability measures of human-computer interaction (e.g., QUESI, NASA-TLX, ISONORM, etc.) and highlight pros and cons of each approach. This work is carried out with experts from a school for technicians who provided us with the required participants.

1.1 Related work

The first studies in the area of collaborative work [3] investigated the conversations and information exchange of two workers working side by side on an assembly task. They primarily exchanged information related to the identification of objects to handle, instructions for performing tasks and the confirmation of finished tasks. Further observation studies [12] introduced a common working field, which can be used for sketches and writings. It has been shown that this common view helped to support the work process and to better convey information. For a remote constellation, research [8] showed that collaborative groups on the same workplace performed better than spatially separated coworkers, but also showed the potential of multimedia systems. Another study focused on the question which visual information provides a benefit for a co-working team. The study [7] measured the performance of a person working alone on a bicycle repair task with a group of a remote mechanic and a local worker on the same task. The experiment provided evidence for the importance of shared visual context for remote collaborative work.

1.2 Aim of this study

Before the experiment, we formulated the following hypotheses:

- Participants wearing augmented reality glasses (head mounted display, HMD) are likely to perform the task quicker than when

*e-mail: doris.aschenbrenner@tudelft.nl
†e-mail: michael.rojkov@gmail.com
‡e-mail: florian.leutert@telematik-zentrum.de
§e-mail: J.C.Verlinden@tudelft.nl
¶e-mail: S.G.Lukosch@tudelft.nl
‖e-mail: marc.latoschik@uni-wuerzburg.de
**e-mail: schi@telematik-zentrum.de

978-1-5386-7592-2/18/$31.00 ©2018 IEEE
DOI 10.1109/ISMAR-Adjunct.2018.00036

the tablet out of hand or bend back for the projection based method. That is also why we think that of all tested variants, the HMD will provide the best support for building up situation awareness and grounding.

using other applications, because the user does not need to put

- The participants might prefer to use the projection based method, because it is more intuitive and requires a lower task load. We think that this is due to the fact that it provides the most stable information presentation and is easiest to use, as the worker does not need to do much on his or her own. This will show in the values for the questionnaires.

- For the grounding we assume that the HMD is the best solution for supporting both workers. It is the only application which provides a continuous view for the expert and we also presume that this first person viewpoint enables the expert to help the participant in a better way.

2 User Study

The user study was carried out with $N = 50$ participants (48 male, 2 female) which were between 19 and 26 years old. The average age was 21.1 years (SD = 1.9 years). All participants were recruited from the technician school Franz-Oberthür-Schule in Würzburg. The highest educational qualification was higher education entrance qualification (German Abitur) by one participant. 17 participants already had completed vocational training, 29 participants finished junior high school (German Realschule) and two participants had a secondary modern school qualification (German Hauptschule). 19 participants wore glasses during the experiment for optical aid. 70 percent of the participants said that they use computers on a regular basis, 48 percent of which use it for office applications and 30 percent for gaming. 92 percent of the participants said that they need to solve mechanic tasks on a daily basis and spend approximately 23 hours per week with those tasks. 64 percent said that they solve those tasks from time to time in teamwork and 34 percent said that they work in a team regularly.

2.1 Repair scenario

In this specific case, we use the Robotstar V switch cabinet, which was provided to us by KUKA Industries. It is a typical switch cabinet for all the different robot configurations they are selling. The specific internals can differ slightly between different production setups, but the servo amplifiers are always present one for each robot axis. They can be seen in Fig. fig:brillelocal as blue boxes. They are interconnected with different cables and plugs and can only be exchanged with a complicated extraction procedure.

If a breakdown happens the plant is stopped, but the switch cabinet still has voltage. Correctly connecting the plugs in the right order is mandatory for installing or dismounting a controller. If a cable is connected the wrong way, short-circuits may occur inside the plant or the robot may drive erroneously (after it is restarted). This may cause a complete breakdown of the plant, damage to expensive robot parts or in the worst case injury to people. In order to protect property and personnel, it is indispensable to omit failures and increase the diligence of the maintenance.

In the specific case of a controller exchange in the switch cabinet, apprentices or new employees need extensive training to enable them to perform this maintenance action on their own. As already small mistakes may lead to severe consequences, their work needs to be supervised by a trained employee, or at least the result needs to be reviewed. Supporting an untrained worker with remote collaboration would help to minimize failures and obviate the need for review in the best case.

Figure 1: The external expert (performed by a researcher) looking at the HMD transmission

Figure 2: A participant working with the Head mounted Display (HMD) following instructions

Figure 3: The setup for projection based AR application (a Panasonic PT-VZ575N projector with an attached PointGrey Blackfly camera)

Figure 4: A participant working with the projection based AR application (red annotation, arrow)

Figure 5: A participant working with the tablet pc AR application with tracked annotations (green)

Figure 6: The screenshot tablet application (participant needed to lay down tablet pc)

2.2 Experimental setup

2.2.1 Head mounted display AR: condition ”HMD”

A head mounted display (HMD) is a display device worn on the head or as part of a helmet, that has a small display optic in front of one (monocular HMD) or each eye (binocular HMD). For this experiment we used the Epson Moverio BT200, which is shown in Fig. 2. Those are glasses which have a LCD polysilicium display integrated for each eye, illuminated by mini projectors on the rim. The displays provide a resolution of 960x540 pixels and cover 23 percent of the field of view. There is an external control device with a touchpad, using a 1.2GHz dual-core-processor and 1GB RAM. It uses Android 4.0.4.

The Android application connects to the expert's Desktop PC application, which is displayed in Fig. 1 via Wi-Fi and tries to visually track camera picture features at the location of the expert's annotations. We have published a separate paper about the tracking algorithms [11].

2.2.2 Projection based AR: condition ”Projection”

The projection based augmented reality application (or spatial augmented reality application [2]) uses a Panasonic PT-VZ575N projector with a resolution of 1920x1200 pixels. The setup is displayed in Fig. 3: Mounted rigidly to the projector is a PointGrey Blackfly camera that offers a static third person perspective of the work surface. After the camera and the projector have been calibrated, a Structured Light approach is used to gather a surface model of the working site. With the help of this model, the expert can directly send visual instructions to the surface of the working environment, which is displayed in Fig. 4. The worker also can use a pointing device tracked by a Polaris Spectra system in order to visualize his view. More publications to this specific setup can be found at [9].

2.2.3 AR on tablet computer: conditions ”Video”, ”Screenshot” and ”Tracking”

Tablet PCs have been commercially available for several years and are regularly used in everyday life as well as in industrial settings. This is why this device serves as a representative of the current common practice in industry, and its performance in the described experiment is also meant to encourage more computer supported collaborative work with a shared visual view in the field. In our prior work, we used tablet computers for cooperative work in an active production environment [1].

This is also one reason why we compared three different applications running on the tablet PC: a version which provides just the camera picture, a version which enables the expert to make a screenshot and annotate it (see Fig. 5) and a permanent tracking approach (see Fig. 6). The latter figure also shows that the tablet PC needs to be placed on the ground during the actual repair work.

We use a ASUS MEMO ME302C Tablet with an 1.6GHz Intel Atom Z2560 Processor with 2GB RAM. The device runs Android 4.3 and the display has a resolution of 1920x1200 pixels.

2.2.4 Audio-only: condition ”Phone”

In order to provide a baseline to the AR application experiments, we chose contact between worker and expert without shared visual context by introducing a ”telephone” condition. In our experimental setup, both users cannot see each other (see Fig. 7) but can have a normal conversation in the same room. This is the optimal condition which eliminates all latency of the audio, and this audio condition is used for all of the experiments. There is of course an influence of the audio quality and latency on the quality of CSCW, but it is intentionally left out in this experiment in order to compare the visual conditions only.

Figure 7: Experiment setting

2.3 Experimental procedure

Every participant undertook the task of mechanically removing a malfunctioning controller and inserting a new one, using one specific application (see section 2.2). We used a between-subject study design in order to omit learning effects. It was not possible to randomize the application, because setting up the process for each application took too long. We therefore chose specific days for specific applications and tried to randomize the participants. Conducting an experiment took 40 to 60 minutes. The technician school Franz-Oberthür-Schule provided a separate room for the study and enabled the attending students to perform the study during his or her normal school time.

The experimental setup can be seen at Fig. 7: the researcher on the left works with a laptop with an application presenting the visual feedback and an instruction sheet (in order to make every run as comparable as possible). The participant on the right in front of the switch cabinet uses the specific augmented reality application (in this case the HMD). Between both there is a visual barrier (a partition screen), so that they cannot see each other but can hear each other via normal speech.

Each experiment followed the following structure: After filling out preliminary questionnaires, the participant was introduced to the task and the specific AR application. After an exploration phase, the controller exchange was executed and the duration measured. Final questionnaires have been followed by free qualitative feedback.

2.4 Measured data, grounding analysis

This study wants to measure the effectiveness of the different applications in supporting a collaborative industrial repair task. The focus of this study was the intuitive usage of the applications by the local worker; that is why the remote expert has been doubled by the same researcher who followed a specific script. The gathered data of the local worker comprises completion time, detected errors after task completion, QUESI [5], NASA TLX [4], SART [13] and ISONORM [10] questionnaires and demographic data. Additionally, the voice interaction of the participant and the researcher (in the role of the remote expert) was recorded in order to get insights on the grounding.

In order to measure grounding, the recorded conversation between both partners had been transcribed afterwards. Following a coding scheme of [7] we categorized the single phrases into ”question”, ”description” and ”acknowledgement”. Additionally, the descriptions have been categorized into ”reactive” and ”proactive”, according to whether the specific speaker initiates the conversation or not.

3 Results

3.1 Exercise Duration

In order to analyze differences regarding the averages M of the measured duration of the controller exchange, the one-factorial ANOVA with independent probes has been used. The independent variable is the used application with their six different forms described in section 2.2: HMD, Projection, Video, Screenshot, Tracking and Phone. The dependent variable is the measured duration of the experiment task. A significant main effect has been found: $F(5,53) = 3.906$, $p = .004, \nu^2 = 0.269$. The pairwise comparison showed a significant difference between the Projector condition ($M = 15:08, SD = 03:04$) and the Phone condition ($M = 20:46, SD = 04:37$) with $p = .004$.

3.2 QUESI

The questionnaire for intuitive use (QUESI) has been calculated for each subscale out of the average of the corresponding items. The overall value is calculated as the average of the subscales. The values for the the different applications are visualized in Fig. 8. Regarding the overall value, no significant effect could be found. The subscales have been tested of significant results with ANOVA and a significant difference in the subscale ”Cognitive load” ($p = .037$) has been found. In the pairwise comparison, the Screenshot application ($M = 4.70$, $SD = 0.36$) performed significantly better than the Phone application ($M = 4.00$, $SD = 0.78$) with $p = .05$. Furthermore, in the same subscale the Tracking application ($M = 4.70$, $SD = 0.29$) has been found to be significantly better than the Phone application ($M = 4.00$, $SD = 0.78$) with $p = .05$.

3.3 Grounding

At first, we examined whether visual feedback for the expert influenced the communication compared to the Phone condition. For each category of the coding scheme (”question expert”, ”description expert reactive”, ”description expert proactive”, ”acknowledgement expert”, ”question technician”, ”description technician reactive”, ”description technician proactive” and ”acknowledgement technician”) results are displayed in Fig. 9. For the difference of the average values a T-test was conducted leading to the results in Tab. 10. The visual media performed significantly better in nearly all categories than the Phone. For the differences between all the categories with respect to grounding, a multi-variant ANOVA for independent probes has been calculated resulting in a significant main effect $F(40,255) = 2.933$, $p = .000$, $\nu^2 = 0.315$. The pairwise comparisons lead to several significant results, which are also included in Tab. 10.

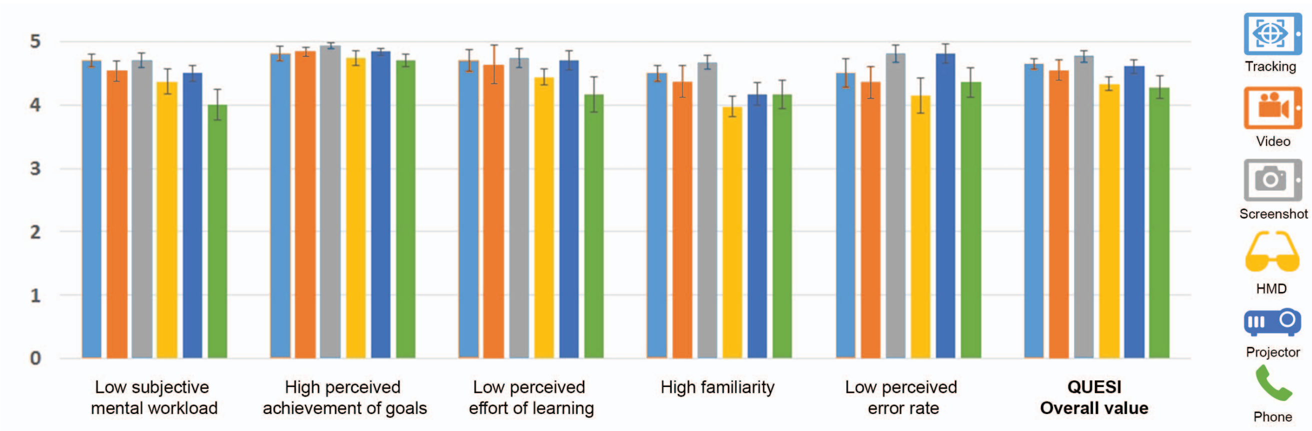

Figure 8: QUESI values of different applications

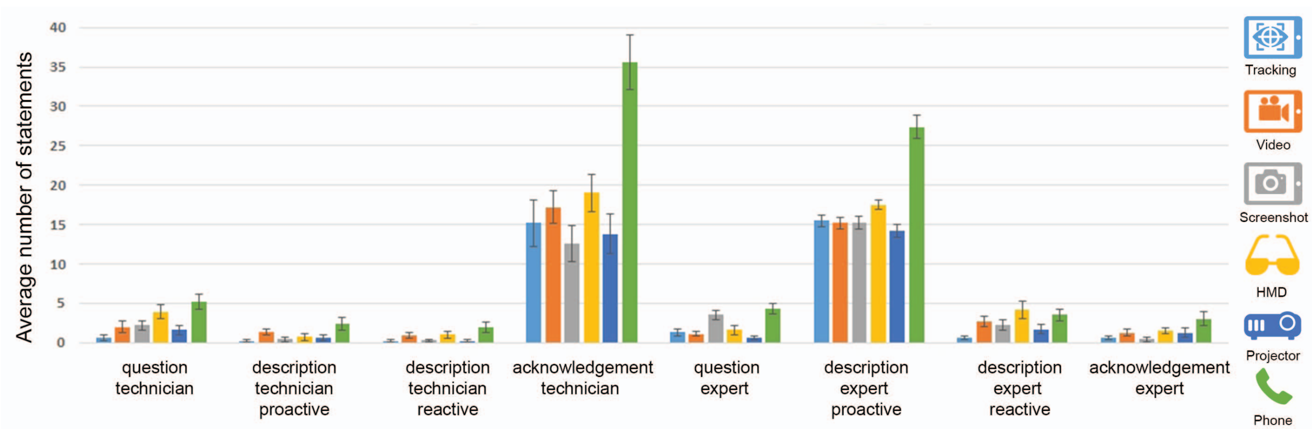

Figure 9: Grounding measured in statement amount of different applications

Factor	Scala	Application	Mean	Effect
Duration	Time	Projection vs Phone	15:08 vs 20:46	.004
QUESI	Cognitive load	Screenshot vs Phone	4.70 vs 4.00	0.05
QUESI	Cognitive load	Screenshot vs Phone	4.70 vs 4.00	0.05
Grounding	question technician	Tracking vs HMD	0.60 vs 3.90	.036
Grounding	question expert	Tracking vs Screenshot	1.30 vs 3.50	.036
Grounding	question expert	Video vs Screenshot	1.10 vs 3.50	0.15
Grounding	question expert	Projection vs Screenshot	0.60 vs 3.50	.001
Grounding	description expert reactive	Tracking vs HMD	0.60 vs 4.20	.012
Grounding	question technician	Visual media vs Phone	2.06 vs 5.20	.000
Grounding	description technician proactive	Visual media vs Phone	0.64 vs 2.40	.001
Grounding	description technician reactive	Visual media vs Phone	0.52 vs 1.90	.002
Grounding	acknowledgement technician	visual media vs Phone	15.56 vs 35.60	.000
Grounding	question expert	visual media vs Phone	1.64 vs 4.30	.000
Grounding	description expert proactive	visual media vs Phone	15.16 vs 27.40	.000
Grounding	acknowledgement expert	visual media vs Phone	0.98 vs 3.00	.001

Figure 10: Summary of significant results

4 Discussion

All significant findings are summarized in Tab. 10. At first we found a significant result regarding duration the participants needed in order to conduct the controller exchange. The participants using the projector based augmented reality application achieved the best results ($M = 15:08, SD = 03:04$). The tablet PC based screenshot application ($M = 18:18, SD = 02:58$) reached similar results as the tracking application ($M = 18:19, SD = 02:42$). The HMD application induced an average measured duration of 18 : 21 ($SD = 03:27$) and lies ahead of the Video ($M = 20:04, SD = 03:12$) and the Phone application ($M = 20:46, SD = 04:37$). These results oppose our first hypothesis, in which we thought that the HMD would have the fastest results.

As noted above, the projector application leads to a significantly shorter experiment duration compared to the other applications. Similar results can be derived from the questionnaires: the Projector application achieves the lowest task load of 25.58 (average NASA TLX Overall value). Also the situation awareness was best with 21.4 (average overall SART value). This supports our second hypothesis: The direct projection of the instructions at the work surface apparently helped to reduce work load and increase situation awareness. In contrast to the duration, in those questionnaires the result have not been significant.

Our third hypothesis considered the HMD application to lead to the best grounding. The data does not support this: On the contrary, both worker and expert needed on average more statements using the HMD (49.4) than in the Tracking (34.2) or Screenshot (36.8) condition. The best result was achieved by the Projector application (33.9 statements per experiment on average). There has been a significant difference in the category "worker question" between Tracking (on average 0.6 questions) and HMD (on average 3.9 questions). This is mainly due to the fact that the worker has put the tablet PC aside for his work. If he or she wanted to ask a question, they needed to pick up the tablet PC anew. Also the significant effect measured in the category "expert reaction reactive" is a direct result of that behavior. The expert had on average 0.6 reactive descriptions per experiment with the Tracking application and 4.2 with the HMD. Additionally, the "expert question" category has been significant between Tracking ($M = 1.3, SD = 1.4$) and Screenshot ($M = 3.5, SD = 1.9$) and Projector ($M = 0.6, SD = 0.7$) and Screenshot ($M = 3.5, SD = 1.9$). This was due to the fact that the expert tends to ask the worker to send him another screenshot of the work environment. It has been observed that the participants used visual feedback over the applications in order replace auditive feedback. For example the worker indicated the start of a task via putting the tablet PC aside without directly saying it. This is also supported by the comparison of the statement category "worker acknowledgement", which is shows a significant difference between tablet applications and non-tablet applications ($t(58) = -2.858, p = .006$).

5 Summary and conclusion

This study and its findings contribute to the improvement of remote maintenance tasks by using concepts of industry 4.0, namely computer supported collaborative work. In an industrial remote maintenance task, a worker is supported in a repair task by a remote supervisor. The service technician is there to execute maintenance orders while the supervisor trusts on the technician to obtain situational awareness. Using a practical repair task, research on the impact of information highlighting using different means of multimedia representation was carried out. Different optical tracking algorithms were compared and implemented according to their suitability for industrial remote maintenance tasks. A tablet PC, optical see-through AR-Glasses, and a spatial AR projector were used as visual output media.

The field test has been carried out at the technician school Franz-Oberthür-School Würzburg, covering 50 participants in sessions of 40-60 minutes. To measure the task performance, the task time was recorded, while QUESI and NASA TLX-questionnaires were used to evaluate the quality of interaction. Furthermore, a linguistic coding was conducted on the communication during the experiment, and situation awareness and grounding was measured. The results suggest that the use of optical tracking algorithms, i.e. spatial awareness, improved the support of industrial remote maintenance tasks. Even though only minor differences in the performance of the different systems were found, the linguistic coding shows significant differences in how the worker and the supervisor communicate and interact with each other. From the point of view of the worker, the spatial projection of the instructions at the work surface does reduce work load and increase situation awareness, while task execution was significantly shorter than with the other conditions.

Acknowledgments

The authors wish to thank KUKA Industries for providing the switch cabinet and the repair instructions as well as Franz-Oberthür Schule in Würzburg for enabling us to perform the user study with their students.

References

[1] D. Aschenbrenner, F. Sittner, M. Fritscher, M. Krauß, and K. Schilling. Cooperative remote repair task in an active production line for industrial internet telemaintenance. *IFAC-PapersOnLine*, 49(30):18–23, 2016.

[2] O. Bimber and R. Raskar. *Spatial Augmented Reality - Merging Real and Virtual Worlds*. A K Peters, Ltd., 2005.

[3] N. V. Flor. Side-by-side collaboration: A case study. *International Journal of Human-Computer Studies*, 49(3):201–222, 1998.

[4] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[5] J. Hurtienne and A. Naumann. Quesi – a questionnaire for measuring the subjective consequence of intuitive use. *Interdisciplinary College*, 536, 2010.

[6] R. Isermann. *Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems*. Springer Science & Business Media, 2011.

[7] R. E. Kraut, S. R. Fussell, and J. Siegel. Visual information as a conversational resource in collaborative physical tasks. *Human-computer interaction*, 18(1):13–49, 2003.

[8] H. Kuzuoka. Spatial workspace collaboration: a sharedview video support system for remote collaboration capability. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 533–540. ACM, 1992.

[9] F. Leutert, C. Herrmann, and K. Schilling. A spatial augmented reality system for intuitive display of robotic data. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pp. 179–180. IEEE Press, 2013.

[10] J. Prümper. Der benutzungsfragebogen isonorm 9241/10: Ergebnisse zur reliabilität und validität. In *Software-Ergonomie'97*, pp. 253–262. Springer, 1997.

[11] M. Rojkov, D. Aschenbrenner, K. Schilling, and M. Latoschik. Tracking algorithms for cooperative telemaintenance repair operations. *IFAC-PapersOnLine*, 50(1):331–336, 2017.

[12] J. C. Tang. Findings from observational studies of collaborative work. *International Journal of Man-machine studies*, 34(2):143–160, 1991.

[13] R. Taylor. Situational awareness rating technique(sart): The development of a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations 17 p(SEE N 90-28972 23-53)*, 1990.

[14] X. Wang, S. Ong, and A. Nee. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing*, 4(1):1–22, 2016.