

Evaluation of a Mobile AR Tele-Maintenance System

Michael Kleiber and Thomas Alexander

Fraunhofer FKIE, Neuenahrer-Str. 20, 53343 Wachtberg, Germany
{michael.kleiber, thomas.alexander}@fkie.fraunhofer.de
<http://www.fkie.fraunhofer.de>

Abstract. Despite increased bandwidth in world-wide networks there are still situations in which only limited bandwidth is available for tele-collaboration between a technician maintaining a machine and an expert at a remote location. In this case, tele-cooperation is usually supported by synchronous audio but only asynchronous video exchange. We present an alternative approach for such a collaborative maintenance task. By utilizing techniques and technologies from Augmented Reality (AR) applications our approach can provide a synchronous shared visual context for the collaborators without a direct video link. The effectiveness of our experimental system for a tele-maintenance task was evaluated in a usability study.

Keywords: augmented reality, system design, evaluation.

1 Introduction

Technologies for audiovisual remote cooperation have been available for decades. Despite increasing network bandwidth there are still many situations where only limited data exchange is possible. Yet, powerful mobile devices and data storage allow a technician to access a comprehensive database of support information by means of a portable computer system. It can also provide step-by-step visual instructions. By utilizing AR-technologies additional visual information about an assembly or a maintenance procedure can be integrated into the real scene [1]. However, such systems are limited to standardized, well-known procedural tasks. Non-standard problems often require a remote expert supporting the technician during the maintenance task.

In these situations an interactive cooperation between a remote expert and a local technician is not possible because the available network does not provide sufficient bandwidth for a synchronous transfer of audiovisual information. The lack of visual information requires the technician to describe any visible malfunctions verbally. It also requires the expert to guide and direct the technician by means of verbal descriptions. However, sharing a visual space through the use of video improves communication and interactive cooperation [4]. Our approach to solve the dilemma between the need for visual information exchange and a narrowband network is based on a virtual reconstruction of the maintenance object through the use of AR techniques.

A typical tele-maintenance session consists of two interactive phases: Problem analysis and problem solving. In both phases a technician benefits from support by an expert at a remote location. In the problem analysis phase the expert can direct the technician towards identifying potential causes of the problem. During the following problem solving the expert can advise the technician. It is assumed that the support of the expert will increase the performance of the maintenance work. However, usual tele-cooperation methods and systems limit interactivity and, thus, efficiency and effectiveness of the joint maintenance work.

2 Earlier and Related Work

Earlier and related work can mostly be found in the application of AR for industrial purposes. There are also large collaborative research consortiums such as ARVIKA [3], STAR [14] and ARTESAS [5] which investigate the use of AR for almost all aspects of manufacturing. The according research activities describe rather elaborated system concepts which involve multiple cameras, multiple computers and are in general quite complex.

Utilizing AR for maintenance tasks has slightly different requirements. Henderson and Feiner categorize maintenance as consisting of activities involving the inspection, testing, servicing, alignment, installation, removal, assembly, repair, overhaul, or rebuilding of human made systems [7]. In these categories, assembly tasks have received the most attention. These tasks can range from assembling aircraft wire bundles [2] to assembling medical equipment for minimal invasive surgery [13].

However, in most of the related work tele-cooperation is of only minor importance. The AR applications rather resemble an extended electronic handbook which the technician uses without additional help by an expert. The user-friendly creation of AR scenes consisting of a set of maintenance instructions is therefore an important topic in the work of the research consortia cited above as well as in other projects [10].

The problem of network connections with limited bandwidth between the collaborators of a tele-maintenance session has hardly been considered. Although it seems obvious to limit data transfer to a numerical object ID and spatial information, no similar approach describing a tele-maintenance system consisting of an AR system for the technician and a VR system for the expert was found in our literature research. Therefore, the following contribution addresses the description of the concept of tightly interconnected AR/VR systems for tele-maintenance and in the practical evaluation of our prototype implementation.

3 System Concept

The system concept described in this section addresses the problem of providing a shared visual space for a technician and an expert collaborating via a narrowband data connection. It is an alternative to the commonly used video transmission requiring a lot of bandwidth. The expert is provided with an egocentric view

of the technician so that instructions and inquiries can be formulated relative to the technicians own view. Instead of transmitting a video stream our system first identifies the machine parts in the view and then transmits the ID of these parts as well as their location and orientation in a machine-specific coordinate system. The expert uses a VR system which reconstructs the 3D view from the technicians point-of-view (Fig. 1). A precondition for this concept to work is that 3D models of the machines and their subparts are available and their real world counterparts are identifiable.

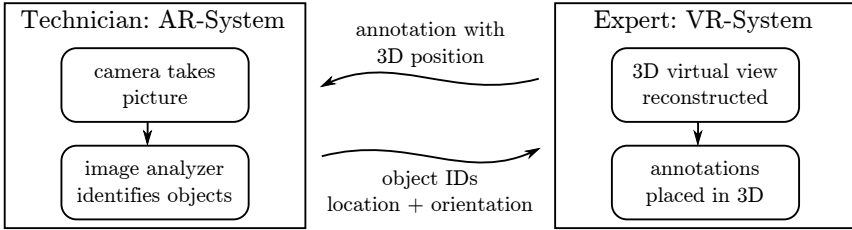


Fig. 1. The basic concept underlying our AR tele-maintenance system

In our case, the expert uses a desktop VR system to view and interact with the virtual 3D view. Besides the virtual view from the technician the expert can examine an interactive model of the maintenance object from an arbitrary viewpoint. This way he can explore the object interactively, e.g. to plan the repair of the machine.

For the AR system of the technician we first proposed using a head-worn display which frees the hands of the technician. However, the weak acceptance of such displays by actual maintenance personal prevented this approach. Instead, an AR system was constructed using a touch screen mounted on a multiple joint arm. The camera of the AR system was mounted on the backside of the screen so that the screen provides a "magic window" to the real world.

4 Implementation

We decided to use the XML-based XMP protocol for the interconnection of the systems of the expert and the technician. The protocol is the basis of many instant messaging services and was designed to be extensible. Similarly to e-mail a user is registered with an XMPP service provider. A user can exchange messages and presence information as well as arbitrary data with other users using any client software which implements the XMP protocol. The protocol is open, standardized and extensible. Since XMPP is text-based and all messages are relayed through a server the transmission of data with real time constraints must be carried out through a separate channel, e.g. audio and video data should be transmitted using the RTP. However, the XMP protocol can be used for signaling and control [12].

The desktop VR system of the expert as well as the mobile AR system of the technician are implemented using the open source OAW VR toolkit [19]. The toolkit is based on the X3D ISO standard [8] and includes many components of realtime 3D applications. However, it does not describe a way to exchange data through a network. There is an attempt by the *X3D Networking* working group to amend the X3D specification to allow such cases. They try to provide networking capabilities at the node level so it is easier for content developers to use networking with no or only limited scripting [18]. However, the proposed amendments, when implemented on top of raw network sockets, do not address issues such as NAT traversal or user identification and authentication. We implemented an X3D node called `XmppSensor` which follows the proposal of the *X3D Networking* working group and is based on an the XMPP extension Jabber-RPC [15]. This extension specifies how remote procedure calls (RPC) can be made through XMPP. We decided to not strictly adhere to the Jabber-RPC standard because it only includes support for basic data types. Instead, we directly encode the X3D types. Thus, the event `vpPos` with type `SFVec3f` and value (0, 1, 3) will be encoded as follows:

```
<iq type='set'
  from='remoteUser@xmppServer.local.net/x3d-rpc'
  to='userName@xmppServer.local.net/x3d-rpc'
  <query xmlns='jabber:iq:rpc'>
    <methodCall>
      <methodName>vpPos</methodName>
      <params>
        <param><value><SFVec3f>0 1 3</SFVec3f></value></param>
      </params>
    </methodCall>
  </query>
</iq>
```

The OAW toolkit can be extended by plugins so that additional functionality besides the core X3D specification can be used. Since the current X3D specification does not contain a method for integrating external sensors like cameras or tracking devices we added a node for the acquisition of live camera images which inherits the interface of the X3D `PixelTexture` node. The camera image is processed by the AR tracking software. It analyzes the image in order to identify and locate maintenance objects. The software also calculates the camera position in relation to the tracked objects. This information is transmitted to the expert and is used in the reconstruction of the technicians view. Since we are using `ArToolkitPlus` which employs marker tracking to estimate the camera pose, the tracked objects in our case are the markers [17]. We use multiple BCH-markers with ID numbers which can be related to specific maintenance objects or parts. The 8×8 BCH-markers we employ allow 4096 different IDs which is sufficient to distinguish all subparts of a typical maintenance object.

The camera image is also displayed on the technician's touch screen. The technician can select or point at locations in the image using his fingers or a pen. This information is also transmitted to the expert and visualized. Likewise, the expert can indicate a location on the 3D model and optionally annotate it which will be visualized as an augmentation of the live camera image on the touch screen of the technician. Additionally, the expert can select specific steps of available maintenance procedures and transmit them to the technician.

The collaboration between the expert and the technician can be solely based on the virtual view of the scene as it provides an interactive shared visual context. Nonetheless, when a detailed view of a specific part of the maintenance object is required, the expert can request a photo snapshot with high resolution.

5 Evaluation

The developed tele-maintenance system was evaluated for a practical maintenance task in a short usability experiment. This allows insights into the usability of the AR/VR system. In our case, the system was compared to a conventional system. For quantifying overall performance, completeness and correctness of the maintenance work and the time to complete the task were measured and analyzed. We also captured the subjective cognitive workload by a standardized questionnaire (NASA-TLX as in [6]).

Based on the above we formulated the following hypotheses for the comparative evaluation of the developed AR/VR tele-maintenance system ($S_{AR/VR}$) and a conventional system (S_{conv}):

- H_1 The overall time required to complete the maintenance task will be less when using $S_{AR/VR}$ compared to S_{conv} .
- H_2 The expert will need to give less instructions to the technician or ask him less questions when using $S_{AR/VR}$ compared to S_{conv} .
- H_3 When using $S_{AR/VR}$ the technician will need to ask less questions or give fewer descriptions compared to when using S_{conv} .
- H_4 The cognitive strain on the technician is lower when using $S_{AR/VR}$ compared to S_{conv} .

5.1 Method

Both tele-maintenance systems adhered to the same general conditions: Expert and technician could use a technical handbook, they could communicate using an audio connection with GSM quality and a maximum latency of 0.5 seconds and they could exchange data using a GSM channel (14.4 kBit/s). Because of these conditions, video could not be transmitted and even the transmission of a JPEG compressed medium resolution image took 20 seconds. However, the available data rate is sufficient for transmitting machine part IDs as well as their location and orientation relative to the camera. Both systems were implemented on the same hardware setup.

In the conventional setup the technician was able to take snapshots of the current view and send them to the expert. The expert was able to annotate the received photo and return it to the technician. In the AR/VR setup the current viewpoint position was continuously transmitted to the expert and used to reconstruct the view of the technician. The expert was also able to request a photo snapshot of the current view. Annotating the photo was not possible, but annotations could be placed on the interactive 3D model which were then displayed as augmentations on the touch screen.

5.2 Maintenance Task

The maintenance task was to replace the camshaft of an automobile engine. For this task a considerable part of the engine had to be disassembled following a specific order. The manual skills required for the task were not demanding, but it was rather difficult to complete the task by the technical handbook only. Broken, dirty or missing parts enhanced the difficulty. This was the case with the engine we used because marks required for aligning the cylinders were missing (Fig. 2). For the evaluation the single steps and the technical sketches scattered in the handbook were combined in consecutive order.

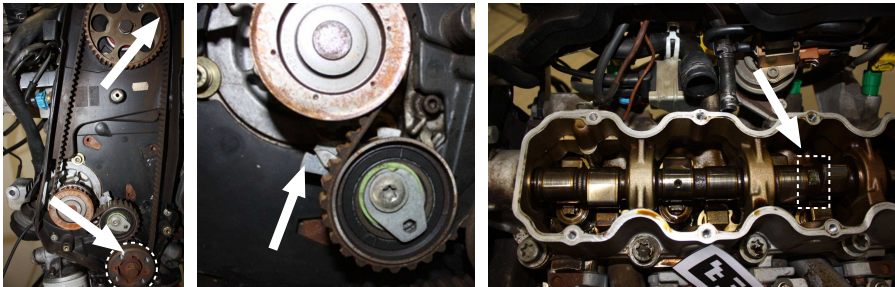


Fig. 2. Before removing the vee-belt the first cylinder has to be brought to the top dead center by aligning the marking on the gear wheel of the camshaft with another marking on the engine block

5.3 Experimental Design and Procedure

We anticipated learning effects in the maintenance task and therefore used a between-subjects experimental design. The two systems served as independent variable. The task completion time was measured as the first dependent variable. We also recorded the audio communication of the expert and the technician for later analysis. The supervisor and instructor of the experiment served as the expert during the tele-maintenance session. The expert was seated in a separate room during the actual experiment so that no means of communication besides through the tele-maintenance system was available. The participants acted as technicians.

We recruited 10 male subjects of similar age (25 – 32 yrs) from the institute's staff who were naive of the experiment's hypotheses. None of the participants

had prior experience in automobile maintenance or repair. To prevent further artifacts because of group differences two pretests were performed. To test the mechanical and technical comprehension we used the MTV test by Lienert [11]. The hose-figure test by Stumpf and Fay was used to quantify the spatial abilities of the subjects [16]. The participants were assigned to two groups according to their combined rank in these two pretests.

Prior to the actual maintenance session the participants received written instructions about their general task. This also included an introduction of the functions of the tele-maintenance system. The participants were allowed to inspect the engine thoroughly and the camshaft housing was pointed out. The available tools were ratchets and wrenches of different size, pliers and screwdrivers. The participants wore a protective coverall and gloves during the experiment and used a pen to interact with the touch screen. Prior to the start of the experiment they received the combined repair instructions. The expert opened the tele-maintenance session a few seconds later. The experiment ended when the participants finished unscrewing the camshaft housing. Directly afterwards they rated their subjective workload by the NASA-TLX questionnaire.

5.4 Results

For evaluating the first experimental hypothesis the overall time of the maintenance procedure was analyzed. For H_2 and H_3 the recorded audio communication was analyzed. A talk cycle for the expert was detected when the audio amplitude was louder than -26 dB (with 0 dB counting as the maximum) for at least 0.07 seconds and ended when the amplitude was lower than -26 dB for 0.75 seconds. Since the level of noise at the technicians work place was louder the required amplitude was increased to -22 dB for the recorded audio of the technician (Table 1).

Although the overall maintenance duration was slightly smaller (36.00 minutes versus 37.44 minutes) for the conventional tele-maintenance system the effect is not significant (standard deviation for S_{CONV} is ca. 5.8 minutes and for $S_{\text{AR/VR}}$ ca. 6.4 minutes).

The amount of speak cycles of the expert for the two groups are also similar. However, the standard deviation of the number of talk cycles seems to be different. For S_{CONV} the standard deviation is 47.58 cycles and for $S_{\text{AR/VR}}$ it is 20.78 cycles. An analysis of the similarity of the variance using the Levene test revealed weak significance ($F(1,8)=6.13$, $p=0.038$). The overall talk duration of the expert for the two groups differs by ca. 160 seconds but the standard deviation is almost as high (for S_{CONV} $sd = 182.71$ s, for $S_{\text{AR/VR}}$ $sd = 151.77$ s). This means there is no statistically significant difference in the talk duration of the expert. The Levene test does not show a significant difference between the two error variances.

Similar to the analysis of the talk durations and cycles of the expert, the analysis for the technician's recorded audio can be done. The talk duration differs by ca. 65 seconds. However, because of the large standard deviation of the data (for S_{CONV} $sd = \text{ca. } 100$ s, for $S_{\text{AR/VR}}$ $sd = \text{ca. } 96$ s) this difference is not

Table 1. The results of the experiment

participant	system	duration [min]	talk cycles [num]		talk duration [s]	
			expert	technician	expert	technician
5	S_{conv}	44.58	253	258	594.48	200.77
10	S_{conv}	33.25	222	270	469.04	258.98
3	S_{conv}	33.59	162	132	187.81	101.05
7	S_{conv}	38.88	219	281	520.91	344.94
4	S_{conv}	29.69	137	182	223.05	122.17
average	S_{conv}	36.00	198.6	224.6	399.06	205.58
1	$S_{AR/VR}$	42.57	201	181	722.85	345.93
9	$S_{AR/VR}$	30.04	189	151	669.75	309.06
8	$S_{AR/VR}$	34.01	229	230	611.05	360.28
2	$S_{AR/VR}$	45.42	172	179	362.96	137.39
6	$S_{AR/VR}$	35.13	198	196	449.59	204.33 (1.04)
average	$S_{AR/VR}$	37.44	187.4	197.8	563.24	271.4

significant. The number of talk cycles is similar for both systems but just as is the case for the expert the variance of it varies significantly. The standard deviation for S_{conv} is 64.7 cycles and for $S_{AR/VR}$ it is only 28.8 cycles. According to the Levene test this is a significant difference ($F(1,8) = 6.736, p=0.032$).

After they finished the maintenance task the participants answered the NASA-TLX questionnaire. There was no significant difference for the two systems.

5.5 Discussion of Results

The usability experiment was performed in order to ensure that the developed system works in a practical setting and to gather first insights about the applicability of the AR/VR tele-maintenance system in contrast to a conventional tele-cooperation system.

The observed overall maintenance time did not differ between the two systems so that hypothesis H_1 must be discarded. The same is true for hypotheses H_2 and H_3 since there is no significant difference in the number of talk cycles and in the length of the talk duration of the expert and the technician. However, the significantly smaller standard deviation of the number of talk cycles for the expert as well as for the technician could mean that when using the AR/VR system individual differences of the subjects have less impact on the results. Yet, when evaluating the performance data the small sample size of the experiment must be kept in mind.

In the subjective rating (NASA-TLX) all but one participant answered the question regarding the achieved performance with "very good" or "perfect". This could be an indication that the task was too easy and that therefore the

expert was not really needed to fulfill the task. Furthermore, the participants' motivation was high (according to the NASA-TLX scale but also when asked directly). These factors, combined with the availability of the repair handbook and the technical background of the participants has led to the situation that the expert support was helpful but not necessary.

6 Conclusion and Future Work

The tele-maintenance system is proposed to be useful and applicable for non-standard situations when a predetermined work plan is not available. However, recreating this setting completely for an evaluation in a comparative experiment is not possible. Future evaluations should therefore be specifically designed so that the technician is dependent on the expert.

We want to extend the system to support the acquisition and display of stereoscopic images. When adjusting the stereoscopic display parameters accordingly the expert can view the scene and perceive correct depth [9]. Furthermore, we want to adjust the authoring process of the work steps shown in the AR/VR systems. Currently, manually generated animations are used. This is very time and labor consuming and not possible for all of the different maintenance procedures. We therefore plan to integrate methods to enable the expert to create these animations during the tele-maintenance session.

Acknowledgments. We gratefully acknowledge funding of the project by the German federal office of defense technology and procurement (BWB), Team T-5.2.

References

1. Barfield, W., Baird, K., Shewchuck, J., Ioannou, G.: Applications of wearable computers and augmented reality to manufacturing. In: *Fundamentals of Wearable Computers and Augmented Reality*, pp. 695–713. Routledge, London (January 2001)
2. Curtis, D., Mizell, D., Gruenbaum, P., Janin, A.: Several devils in the details: making an AR application work in the airplane factory. In: *Proceedings of the International Workshop on Augmented Reality: Placing Artificial Objects in Real Scenes: Placing Artificial Objects in Real Scenes, IWAR 1998*, pp. 47–60. A. K. Peters, Ltd, Natick (1999)
3. Friedrich, W.: ARVIKA-augmented reality for development, production and service. In: *Proceedings of International Symposium on Mixed and Augmented Reality, ISMAR 2002*, pp. 3–4 (2002)
4. Fussell, S.R., Kraut, R.E., Siegel, J.: Coordination of communication: Effects of shared visual context on collaborative work. In: *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, pp. 21–30 (2000)
5. Haberland, U., Brecher, C., Possel-Dlken, F.: Advanced augmented reality-based service technologies for production systems. In: *Proceedings of the International Conference on Smart Machining Systems (March 2007)*

6. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task load index): Results of empirical and theoretical research. *Human Mental Workload* 1, 139–183 (1988)
7. Henderson, S., Feiner, S.: Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In: 8th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2009, pp. 135–144 (2009)
8. ISO/IEC 19775-1:2008: Information technology – Computer graphics and image processing – Extensible 3D (X3D) – Part 1: Architecture and base components. ISO, Geneva, Switzerland (2008)
9. Kleiber, M., Winkelholz, C.: Distortion of depth perception in virtual environments using stereoscopic displays: quantitative assessment and corrective measures, vol. 6803. pp. 6803C–11. SPIE, San Jose (February 2008)
10. Knopfle, C., Weidenhausen, J., Chauvigne, L., Stock, I.: Template based authoring for AR based service scenarios. In: Proceedings of Virtual Reality, VR 2005, pp. 237–240. IEEE, Los Alamitos (2005)
11. Lienert, G.A.: *Mechanisch-Technischer Verständnistest*. Verlag für Psychologie - Hogrefe, Göttingen (1964)
12. Ludwig, S., Beda, J., Saint-Andre, P., McQueen, R., Egan, S., Hildebrand, J.: XEP-0166: Jingle. Draft (December 2009), <http://xmpp.org/extensions/xep-0166.html>
13. Nilsson, S., Johansson, B.: Fun and usable: augmented reality instructions in a hospital setting. In: Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces, OZCHI 2007, pp. 123–130. ACM, New York (2007)
14. Raczynski, A., Gussmann, P.: Services and training through augmented reality. In: Proceedings of the 1st European Conference on Visual Media Production (CVMP), pp. 263–271 (March 2004)
15. Saint-Andre, P.: XEP-0009: Jabber-RPC. Final Standard (February 2006), <http://xmpp.org/extensions/xep-0009.html>
16. Stumpf, H., Fay, E.: *Schlauchfiguren. Ein Test zur Beurteilung des räumlichen Vorstellungsvermögens*. Hogrefe, Göttingen (1983)
17. Wagner, D., Schmalstieg, D.: Artoolkitplus for pose tracking on mobile devices. In: Proceedings of 12th Computer Vision Winter Workshop, pp. 139–146 (2007)
18. Web3D Consortium: X3D Network Working Group Website (2007), <http://www.web3d.org/x3d/workgroups/x3d-networking/>
19. Winkelholz, C., Alexander, T., Weiß, M.: Open ActiveWrl: a middleware based software development toolkit for immersive VE systems. In: Proceedings of the Workshop on Virtual Environments, pp. 321–322. ACM, Zurich (2003)