

Mixed Reality Environment for Mission Critical Systems Servicing and Repair

Andrea F. Abate, Fabio Narducci, and Stefano Ricciardi

VRLab, University of Salerno, Italy
{abate,fnarducci,sricciardi}@unisa.it

Abstract. Mixed Reality (MR) technologies may play an important role in assisting on-site operators during maintenance and repair activities. Nevertheless, industrial equipment augmentation requires a high level of precision when co-registering virtual objects to the corresponding real counterparts. In this paper we describe a comprehensive proposal for a mixed reality environment aimed to improve the effectiveness of servicing and repair procedures in mission critical systems, while reducing the time required for the intervention. The tracking of the user's point of view exploits a multi-marker based solution for robust and precise augmentation of the operating field. The architecture also features a diminishing visualization strategy allowing the user to see only the fraction of real equipment that is relevant for the maintenance task. A finger color-based tracking provides powerful interaction capabilities by means of a not-instrumented interface exploiting colored fingertips caps. An evaluation study of the proposed MR environment, performed by technicians with no previous experience of MR systems, highlights the potential of the approach.

Keywords: Mixed reality, diminished reality, finger based interaction, AR based maintenance.

1 Introduction

Mission critical installations (e.g. radar, navigation and communication systems aboard military and civil vehicles or based in airports, ports, etc.) represent examples of high-tech environments featuring complex hardware and software components requiring prompt servicing as, in case of breakdown, the security of a large number of people may be at risk. In this context, Mixed Reality (MR) technologies may play an important role in assisting on-site operators during maintenance and repair activities but, ultimately, issues like tracking accuracy and coverage, augmentation strategies and interaction capabilities determine whether a MR application is useful or not. Henderson and Feiner explored the benefits of AR for maintenance and repair [1]. They also showed how MR can assist military mechanics conducting routine maintenance tasks inside an armored vehicle turret exploiting a NaturalPoint OptiTrack tracking system [2]. In 2002, the project ARVIKA [3] fostered the research and the development of AR technologies for production and service in the automotive and aerospace industries, for power/process plants, for machine tools and production

gears. Klinker et al. [4] presented an AR system for the inspection of power plants at Framatome ANP, while Shimoda et al. [5] presented an AR solution in order to improve efficiency in nuclear power plants (NPP) maintenance interventions and to reduce the risk of human error. The precision of tracking and visualization strategy adopted represent two key aspects that have to be severely taken into account in industrial components augmentation. Computer vision, in both the marker-based [5-6] and marker-less [7-8] variants, is generally recognized as the only tracking methodology that has the potential to yield non-invasive, accurate and low cost co-registration between virtual and real [9]. Concerning the visualization strategy, how and when to augment real objects and environment may have a great impact on the quality of user assistance provided. For instance, in some situations the scene observed should be possibly simplified rather than augmented. In our proposal, the system is able to remove distracting items from the field of view, leaving as visible only the elements required to perform the desired task.

Another aspect to be considered is represented by the interaction level available and the related interaction paradigm: the user should be able to select what kind of augmenting content to display according to his/her needs by interacting with the MR environment without complicated gear. Wei et al. [10] introduced a MR framework featuring voice commands, but a hand-based interface would rather be more suited to the scope. As a hardware solution (instrumented gloves plus wrists tracking) would provide accurate hands capturing but would also reduce system's acceptability, a more feasible option is to exploit image-based techniques to track hands in real time [11] exploiting the same camera video stream used for the augmentation. The simple and robust approach developed in this paper is based on the recent work by Mistry and Maes [12] and relies on colored caps worn on index and thumb fingers to track their position and gestures, providing effective and natural interaction within the MR environment.

In this paper we describe a comprehensive proposal for a mixed reality environment aimed to improve the effectiveness of servicing and repair procedures in mission critical systems, while reducing the time required for the intervention. As the MR system presented is targeted to rack sized equipment rich of small components and operating in the vicinity of strong electromagnetic fields, it features a computer vision based multi-marker tracking method to deliver high accuracy and robustness and allowing the user to observe the surrounding environment and the computer generated graphics by means of a video see-through HMD. The proposed system also provides powerful interaction capability with the co-registered virtual/real objects by means of a two-hand not instrumented finger based interface exploiting colored fingertips tags.

The rest of this paper is organized as follows. Section 2 describes the overall architecture of the system and each of its main components. Section 3 reports about the experiments conducted in a real industrial case to assess the advantages and the eventual limitations in the proposed approach. Finally, Section 4 draws some conclusions introducing future directions of study.

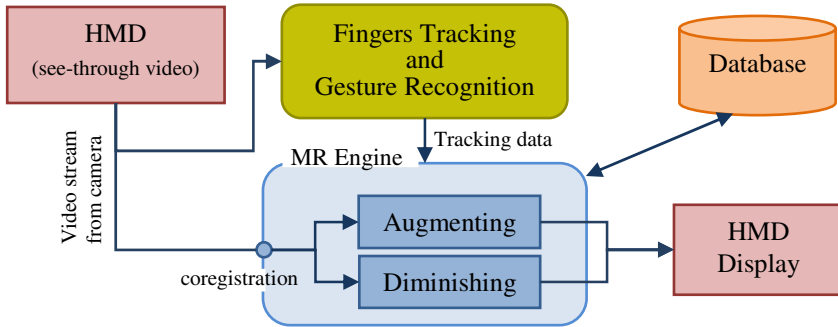


Fig. 1. The overall schematic view of the system

2 System's Description

The schematic view of the proposed system is shown above (see Fig.1). The architecture proposed integrates precise multiple marker based tracking, two modalities for scene augmentation and a not-instrumented finger based interface to provide effective and reliable visual aid during maintenance operations. It is designed around three main components. The Mixed Reality Engine (MRE) is in charge of user's head tracking, scene augmentation/rendering and servicing procedures management. The Finger Tracking module captures fingers position and the Gesture Recognition module recognizes gestures enabling the human-computer interaction, while the Database contains the working environment setup, the virtual contents and the maintenance procedures required for the system to work. To start the assisted servicing procedure, the user has to wear a video-based see-through HMD, a backpack enclosed notebook and a few fingertips caps required for contactless interaction. This architecture is further detailed in the following subsections.

2.1 The Mixed Reality Engine

The tracking system developed exploits the ARToolkit open source AR library for estimating user's perspective. Typical marker based tracking systems operating under controlled conditions (i.e. avoiding or at least reducing strong reflections, extreme shadows and excessive camera noise) are able to track the point of view of the user provided that a single marker is entirely captured in a frame. This simple solution, often adopted for desktop based AR applications, forces the user to continuously aim at the marker, holding it in the center of the visual field to reduce the risk of detection miss. Arranging a marker in the middle of operational environment could simply be unfeasible for many application contexts characterized by uneven surfaces or it could even interfere with the operations. In our proposal, many of these issues are addressed by exploiting multiple markers, thus delivering an inherently more robust and more accurate tracking even using small markers. A useful advantage of the multi marker approach is that it is easily scalable. For instance, by adding other six markers

(arranged in two strips of three, placed 60 cm. above and below the basic set) an optimal tracking volume of 180x60x60 cm (adequate for a full-size industrial rack) is seamlessly achieved. Besides an embedded calibration function aimed to measure and correct the lens distortion of the camera, a manual procedure allows the user to fine-tune co-registration between the real camera and its virtual counterpart in charge of rendering the required graphics. Each of the six degrees of freedom, including the focal length of the camera, and the marker thresholds, can be adjusted precisely. This task is performed only once unless physical or environmental changes occur in equipment's configuration. Concerning the augmentation of the scene, the system is designed to support two different approaches. The former is a classic augmentation strategy which consists of displaying different kind of virtual objects (e.g., arrows, labels, 3D models and so on) onto the captured scene. In this way the system literally "augments" the real environment (see Fig.2). However there are situations in which adding extra information to the scene may lead to an even more confusing effect. In particular, we refer to environments characterized by the presence of a large amount of interaction points (e.g., a control board, a rear panel of a complex device etc). In this case, showing further information in addition to those already present may be counterproductive. These are the reasons that led us to develop of the latter augmentation strategy, enabling the system to hide part of the operating environment to let the user focus on the physical elements on which to perform a particular task. This methodology is inspired to the concept of diminished reality [13-15], but it is based on the selective occlusion of unwanted elements rather than on image based object removal (see Fig.3). From a more technical point of view, to the purpose of hiding real objects or part of them, the engine renders an occluding (polygonal) surface on which can be applied either a diffuse map or an opacity map.

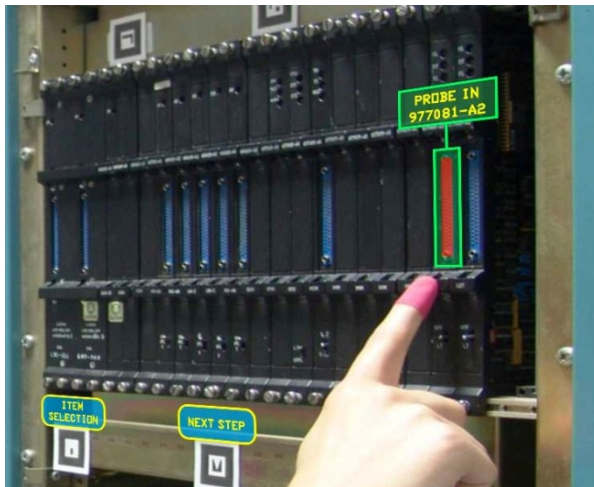


Fig. 2. User interacting with a panel augmented by virtual labels and GUI

By exploiting the device's formal representation stored in the system database (more on this later on), the engine either loads the associated textured polygons or it builds up a procedural texture consisting of a black background featuring white "holes" of various shapes (e.g, circles, squares, polygons and so on) corresponding to the hotspots that should stay visible. All the necessary information to perform this task is available in the working environment database. Once the textured polygons are built, the engine renders it over the real device occluding all the not relevant hotspots. Both the more common augmenting and the diminishing strategies are meant to improve user's operational capabilities. However, there are contexts in which one is more suited than the other. Which of the two visualization methods should be used depends on the total number of hotspots present in the surrounding of the virtual contents visualized at a given step of the intervention. If the density of the hotspots exceeds a threshold then the diminishing modality is preferred. Anyhow, the user may always switch to the other modality in any moment by means of a specific Augmented/Diminished View toggle present in the visual interface. Finally, by combining both augmented and diminished reality a third hybrid visualization approach could be realized, providing a simplified view of the operating field in which only the elements left visible are augmented with additional info. Whatever the strategy adopted, scene augmentation or diminution is made possible thanks to a formal scene representation based on XML, a markup language universally known for being easily extensible and non ambiguous. The XML database consists of a collection of files providing the necessary information to correctly locate each relevant element of the working environment within the 3D space. The MR engine also performs another crucial task: the maintenance procedure management. Each generic maintenance procedure can be represented as a deterministic finite automaton (DFA). According to this approach, a particular state represents a maintenance step and its links define the execution order. DFA result particularly suited to model both simple and complex maintenance procedure in an easy, verifiable and legible way. The DFA representation of a particular procedure is converted in a XML file where a `<step>` tag defines a state. Any possible path through the automaton defines a procedure's file. By this approach a single XML procedure file defines a specific execution order in a maintenance procedure.

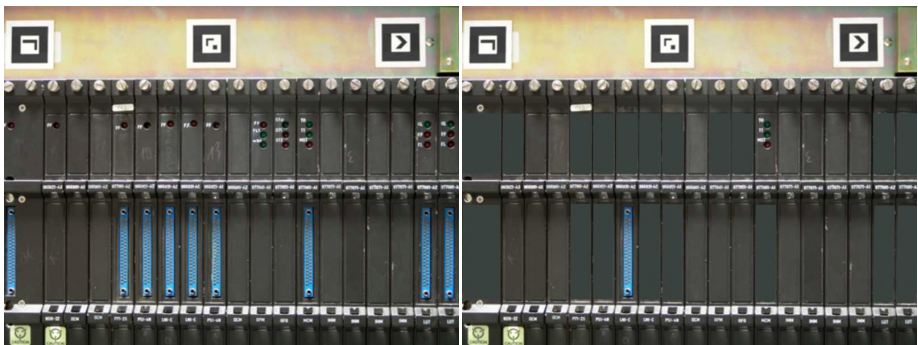


Fig. 3. Original frame (left) and diminished view (right). Only the required elements are left visible, while textured polygons hide potentially confusing items.

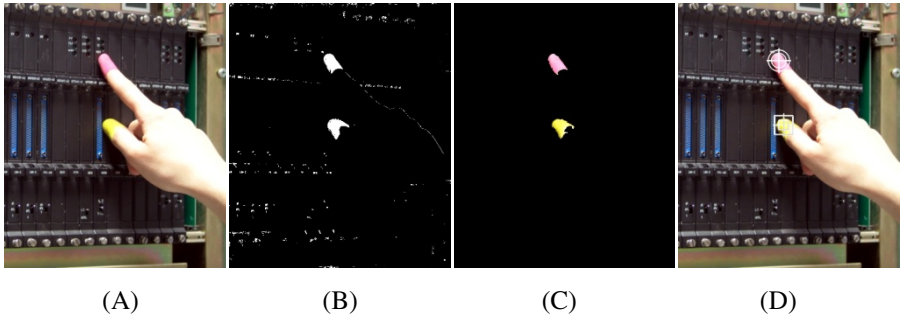


Fig. 4. Fingertips tracking in four steps: (A) original frame grabbed from the camera; (B) candidate pixels highlighted (C); matching regions found; (D) resulting tracking

2.2 Finger Based Contactless Interaction

The contactless interface developed frees the user from the usage of any tangible I/O device to communicate with the system. The user indeed, has only to wear small rubber caps of different colors over his thumb and index fingertips, eventually of both hands. The image-based tracking exploits the same video stream used for marker tracking to achieve fingertips detection and tracking, thus yielding a reduction of computational cost compared to a solution based on a dedicated camera and a simpler hardware configuration. Each finger is associated to a different color according to a simple enrollment procedure repeated for each finger and generally performed only the first time the system startup. The fingertips samples captured are analyzed in the HSL color space to extract the dominant hue and saturation ranges, while the lightness component is used to filter out eventual highlights. At runtime, the image grabbed from the camera is subsampled (by a factor of 4 to 8 times) to both reduce the effect of camera noise and to optimize system's performance. For each pixel in the subsampled image whose HLS levels fall within the ranges defined during enrollment, a recursive search for similar (color wise) neighbors is performed until a region of 20x20 pixels is explored. If at least one half of the pixels inside this region matches with the original pixel, then the engine recognizes that region as one of the colored caps to track (Fig. 4 summarizes the fingers detection steps). This approach resulted both reliable and responsive, granting a sustained frame rate always well above 30 frame per second for an image resolution of 640x480 pixels (typical for most HMD cameras). Finger tracking enables a rich interaction paradigm that can be exploited in many different ways. For instance the user may query the working environment to learn more about it by simply moving the index finger over any hotspot (i.e., a screw, a button, a handle, a led indicator and so on) according to his/her point of view to obtain visual info about a particular component. Moreover, finger tracking enables operating the system by using a graphical user interface (GUI). The main challenge with an intangible GUI is related to the interaction paradigm, which has to manage the lack of physical contact with the interface elements (buttons, slider, toggles etc.). Indeed, when using a conventional (tangible) interface, the kinesthetic feedback provides an important confirm of the operations performed. To address this issue, we

exploited a time based interaction paradigm, requiring the user to hold the finger in position for a defined (around one second) amount of time to trigger the associated function. A visual feedback, in the form of a small progress bar drawn over the GUI element selected, inform about the selection state (i.e. hold the finger until the progress is over).

The same paradigm is used during a servicing procedure to move from a step to the next one or previous one as well as to play/pause/rewind an animated virtual tool showing how to perform a specific task. Additionally, as the system is designed to track up to four colored caps, multi-finger gestures can be used to provide more powerful interaction modalities, like object picking, zoom or rotation.

3 Experiments and Results

We conducted two kinds of experiments on the system described above to assess both the performance of the tracking approach and the overall usability of the MR environment applied to a radar system training facility. The hardware used for the experiments includes a notebook, featuring Intel I5 processor and Nvidia GeForce 9 series graphics board and an ARVision-3Dvideo see-through HMD from Trivisio, equipped with two 800x600 LCD display and two 640x480 cameras capturing the surrounding environment at 30 FPS (see Fig.5). Only the left camera has been used for scene capture. During operations the notebook was contained inside a small backpack. In any system evaluation, user testing is of great relevance in confirming the validity and the effectiveness of solutions adopted. To this aim, we prepared a user questionnaire to assess the perceived quality of the interaction after performing a number of tasks significant to the operating context considered. The evaluation sessions involved ten users, selected among specialized technicians with no previous experience of either MR systems or contactless interfaces. The following is a list of the tasks performed by the testers: **Load** a new servicing procedure; **Select** a particular hotspot; **Select** a function from the GUI; **Toggle** between two functions; **Perform** a servicing procedure.



Fig. 5. User wearing HMD and colored fingertip caps

In the final questionnaire, the questions were presented using a five-point Likert scale, where respondents specify their level of agreement to a statement. In order to avoid any bias, some statements were in positive form and others in negative one. This was taken into account in the final assessment of results. The following is the list of the proposed statements:

1. Available finger based functions are easy to perform;
2. Functions are too many to remember them;
3. Interacting by fingers is not intuitive;
4. It is easy to select objects;
5. Visual aids are clear and useful;
6. It is easy to operate the contact-less GUI;
7. The type and number of available functions to interact with objects is not sufficient;
8. Devices worn are not comfortable during operations.

The answers to the questionnaire are summarized in the table below (see Table 1). Most participants reported a good confidence feeling during the usage of the system, and some of them also reported an operational advantage in performing the proposed tasks with respect to their usual operating modality. All the participants to the evaluation sessions have also been interviewed to better understand the motivations behind the answers provided. Most comments showed a general agreement about the finger based interface, although most of them remarked, as we expected, the lack of a physical contact as something strange to which is not easy get used. Both the two visualization modalities were considered useful for improving the confidence and avoiding distraction errors during the operations, while, not surprisingly, the HMD caused a somewhat stressful experience to most users.

Table 1. Scores reported after subjective system evaluation according to five-point Likert scale

Question	I strongly agree	I agree	I do not know	I disagree	I strongly disagree
1	1	6	1	2	0
2	0	2	1	7	0
3	0	1	2	5	2
4	2	6	0	2	0
5	2	7	0	1	0
6	0	6	2	1	1
7	0	1	2	6	1
8	2	4	1	3	0

4 Conclusions and Future Directions

We presented a mixed reality environment for on-site servicing assistance targeted to mission critical systems. According to the testing conducted the main design requirements (within two millimeter co-registration error, context dependent augmentation strategy and natural interaction) have been overall satisfied. The subjective system evaluation, performed by testers in a radar system's training facility, highlights the potential of the proposed approach, though issues related to the hardware used (the reduced HMD's resolution/field-of-view, brief tracking failure under rapid head movement) might sometimes detract from the MR experience. The selective visual removal of possibly confusing components from the actual scene has proved to be an effective option under certain circumstances. According to questionnaire answers, the combination of augmentation and finger-based interface worked well, providing an intuitive interaction paradigm that proved to be suited to the application context. Overall, the MR aided servicing environment produced a valuable improvement in user's confidence during simulated interventions, which could eventually lead to a measurable reduction of time required to tasks completion. For those reasons we are planning and experiencing more formal statistical tests to better evaluate the strength and weak aspects of the proposed system. Tracking accuracy aside, the tests take into account other useful parameters such as the users's effort in terms of concentration, memorization of the maintenance procedures and load on the user's head. We are currently implementing a stereoscopic version of the whole architecture, to improve depth perception of real environment through binocular scene capture and stereo rendering of virtual contents. We are also experimenting infrared-based scene lighting and capture to improve markers extraction from the background and therefore to reduce both tracking errors and markers size. Finally, we are experimenting a predictive approach to head tracking to further improve augmentation accuracy under rapid movements.

References

1. Henderson, S., Feiner, S.: Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In: International Symposium on Mixed and Augmented Reality (ISMAR 2009), pp. 135–144 (2009)
2. Henderson, S., Feiner, S.: Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics* 17(10), 1355–1368 (2011)
3. Friedrich, W., Jahn, D., Schmidt, L.: ARVIKA-augmented reality for development, production and service. In: Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002), pp. 3–4 (2002)
4. Kiyokawa, K., Billinghamurst, M., Campbell, B., Woods, E.: An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In: Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, p. 133. IEEE (2003)

5. Zhiqiang, B.I.A.N., Ishii, H., Shimoda, H., Yoshikawa, H., Morishita, Y., Kanehira, Y., Izumi, M.: Development of a tracking method for augmented reality applied to NPP maintenance work and its experimental evaluation. *IEICE Transactions on Information and Systems* 90(6), 963–974 (2007)
6. Wagner, D., Langlotz, T., Schmalstieg, D.: Robust and unobtrusive marker tracking on mobile phones. In: *7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR 2008)*, pp. 121–124. IEEE (2008)
7. Maldi, M., Preda, M.: Markerless tracking for mobile augmented reality. In: *IEEE International Conference on Signal and Image Processing Applications (ICSIPA 2011)*, pp. 301–306. IEEE (2011)
8. Comport, A.I., Marchand, E., Pressigout, M., Chaumette, F.: Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *IEEE Transactions on Visualization and Computer Graphics* 12(4), 615–628 (2006)
9. Fue, P., Lepetit, V.: Vision based 3D tracking and pose estimation for mixed reality, pp. 1–22. *Interfaces and Design, Emerging Technologies of Augmented Reality* (2005)
10. Wang, W., Qi, Y., Wang, Q.: An Augmented Reality Application Framework for Complex Equipment Collaborative Maintenance. In: Luo, Y. (ed.) *CDVE 2011. LNCS*, vol. 6874, pp. 154–161. Springer, Heidelberg (2011)
11. Wang, R.Y., Popović, J.: Real-time hand-tracking with a color glove. *ACM Transactions on Graphics (TOG)* 28(3), 63
12. Mistry, P., Maes, P.: SixthSense: a wearable gestural interface. In: *ACM SIGGRAPH ASIA 2009 Sketches*, p. 11. ACM (2009)
13. Jarusirisawad, S., Hosokawa, T., Saito, H.: Diminished reality using plane-sweep algorithm with weakly-calibrated cameras. *Progress in Informatics* 7, 11–20 (2010)
14. Enomoto, A., Saito, H.: Diminished Reality using Multiple Handheld Cameras. In: *ACCV 2007 Workshop on Multi-Dimensional and Multi-View Image Processing*, Tokyo (November 2007)
15. Herling, J., Broll, W.: Advanced self-contained object removal for realizing real-time Diminished Reality in unconstrained environments. In: *9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2010)*, pp. 207–212. IEEE (2010)