Proc. of the 11th Model-Based Enterprise Summit (MBE 2020), Gaithersburg, Maryland, USA, March 31 - April 2, 2020

Leveraging standard geospatial representations for industrial augmented reality

Teodor Vernica[1,2], Aaron Hanke[1,3], and William Z. Bernstein[4]

[1] Associate, Systems Integration Division, Engineering Laboratory,
National Institute of Standards and Technology, Gaithersburg, MD, USA
[2] Department of Computer Science, Aarhus University, Aarhus, Denmark
[3] Chair of Engineering Design and CAD, Technische Universitat Dresden, Dresden, Germany
[4] Systems Integration Division, Engineering Laboratory,
National Institute of Standards and Technology, Gaithersburg, MD, USA
{teodor.vernica, aaron.hanke, william.bernstein}@nist.gov

Abstract

Due to its tremendous potential, Augmented Reality (AR) has experienced a recent surge in adoption and integration within the manufacturing enterprise. While industrial AR has been successfully implemented and shown to have significant benefits in a variety of applications, proper use case development, application-specific evaluation, and data interoperability remain open research challenges. In this work, we demonstrate an AR-enabled use case that allows for remote monitoring and inspection of manufacturing systems by overlaying contextual information, such as machine execution data, over the video feed of the manufacturing floor. Additionally, we discuss challenges related to our prototype's implementation and potential opportunities to mitigate such issues through standard indoor geospatial representations.

1 Introduction

In recent years, Augmented Reality (AR) has proven to be a versatile technology that has been leveraged in a multitude of domains including many industrial applications, such as manufacturing planning [5], assembly guidance [14] and maintenance and repair [6] among many others [11]. In this paper, we discuss a new use case for industrial AR that demonstrates remote inspection and monitoring of manufacturing systems by streaming and contextually representing real-time machine process information over the video stream of an Internet Protocol (IP) camera that can be controlled over the network. In doing so, our prototype system provides users, e.g., foremen, operators, and shop managers, with additional capabilities that leverage existing data structures already deployed within smart manufacturing systems. Moreover, by displaying manufacturing information in AR rather than in a strictly digital environment, dynamic elements common to a workshop floor that are much more difficult or even infeasible to track or model are included, such as humans or tools. Our presented prototype also accepts Computer-Aided Design (CAD) models and other virtual 3D objects as inputs that can be layered onto the video feed for a better representation and correlation between the digital (or *as-designed*) models and their physical (or *as-realized*) instances.

During prototyping, we encountered a number of design challenges related to object tracking and camera pose estimation due to the scale of the scene and the rather large distance between the camera and target. These issues are common across other crowded, complex environments akin to busy production floors. This means that traditional tracking methods, such as marker-based tracking, implemented in existing frameworks are not immediately applicable to this scale or are simply infeasible. In other words, popular marker-based recognition methods used in *large target, small field* situations seem to fail when applied to *small target, large field* scenarios.

DOI:https://doi.org/10.6028/NIST.AMS.100-29#page=194

2 Background

Over the years, numerous different AR tracking and registration techniques have been proposed with various benefits and drawbacks depending on the application context [1, 13].

Fiducial markers have been used in AR applications for decades with diverse designs and implementations [15]. The maturity of marker detection and tracking techniques allows for efficient and reliable pose estimation of the camera that can be computed leveraging the four corners of a marker. In spite of their simplicity and robustness, fiducials have the disadvantage of requiring setup and could cause aesthetic issues, making them inappropriate for certain application environments. In this sense, feature-based markers offer an alternative that trades simplicity for aesthetics, enabled by popular frameworks such as Vuforia[1] and Wikitude[2]. Koch et al. [9] demonstrate how *natural markers* within a building, such as exit signs, can be used for tracking while being seamlessly immersed in the environment.

While both *artificial* and *natural* markers are well-suited for most AR applications, where the markers are relatively close to the camera, success of marker detection and the precision of pose estimation generally decreases with distance. This challenge is evident when applying techniques to wider, more complex environments such as manufacturing floors (often to the point where markers become undetectable). *Artificial* markers seem to perform better than the *natural* markers in such conditions, being more robust to far-field detection and bad camera focus conditions due to their purposeful design.

Even so, most markers were not designed with far-field detection in mind, as most of them report detection ranges of around 3 m for a 20 x 20 cm marker [10]. Cho and Neumann [3] acknowledge this range limitation and present a multiring fiducial design that is able to smoothly zoom-out from near-field to room-sized detection, promising a detection range of up to 4.5 m using a 4 cm diameter circular fiducial. However, at least four of them are required to be in view to calculate the camera's pose. Claus and Fitzgibbon [4] propose a machine learning approach to marker detection, using a marker comprised of four circles on a white background, that shows a significant decrease in error rates compared to square marker detection systems for challenging environment conditions, e.g., bad lighting and far-field detection.

Recently, marker-less methods, such as Simultaneous Localization and Mapping (SLAM) [12], have gained popularity as an alternative to marker-based tracking. Such techniques aim to dynamically build a 3D map of the environment using the already existing natural features without the need of a former setup. This approach is particularly useful for unknown environments or when careful marker placement in the environment is impossible or impractical. Researchers have extended SLAM-based methods to deal with large, complex and dynamic environments. To this extent, Castle et al. [2] present a technique for wide-area tracking by creating multiple distinct maps of different scales that can be used in unison by transitioning from one map to another appropriately. This modular approach has the advantage of only needing to rebuild a subsection of the maps when a change occurs in the target room's configuration. While marker-less tracking techniques can be valuable for tracking the position and orientation of the camera within the environment without the limitation of needing a marker in view at all times, additional work can be required to correctly and automatically register the digital objects in the tracked scene.

Despite its limitations and advances in marker-less tracking methods, marker-based tracking is still effective and commonly used for prototyping purposes due to its robustness and the effortless implementation afforded by frameworks. These limitations can be alleviated in

[1]For more information, visit http://www.ptc.com/en/products/augmented-reality.
[2]For more information, visit https://www.wikitude.com.

combination with marker-less techniques but ultimately there are still challenges when applying these to large, busy environments. These issues were made apparent during the development of our room scale prototypes, which means that they would only be amplified on a larger scale and would require alternative solutions.

3 Camera-Supported Monitoring of Production Systems

To explore opportunities for far-field tracking in the context of industrial AR, we developed a prototype to interface with the National Institute of Standards and Technology (NIST) Smart Manufacturing Systems (SMS) Test Bed[3]. Streaming near-real-time data via a web portal, the SMS Test Bed is representative of a contract manufacturer with a good mix of machine tools. Such an environment offers appropriate testing conditions, e.g., occlusion due to crowded spaces. Our initial testing described here was conducted in the Data Information Visualization and Exploration (DIVE) Lab, recently deployed at NIST.

Our prototype makes use of an off-the-shelf Pan-Tilt-Zoom (PTZ) Internet Protocol (IP) Camera that can stream the video feed and be controlled via the Hypertext Transfer Protocol (HTTP) protocol. A *Unity* desktop application receives the video stream and allows the user to send PTZ commands to the camera, via the keyboard, over the network. Simultaneously, *MTConnect* data generated by computer numerical control (CNC) machines is continuously fetched from the SMS Test Bed, as shown in Figure 1. Quick Response (QR) Codes are used to represent different CNC machines and encode their *MTConnect* Universally Unique IDentifier (UUID). *ZXing*, a barcode processing library, is used for QR Code detection and decoding, while *OpenCV* is used for drawing the detection information on the processed frames. When a QR Code is detected and successfully decoded, the current *MTConnect* data for the respective machine is shown on screen, as shown in Figure 2. In this case, two QR Codes are detected in the frame, representing two different machines: *GFAgie01* and *Mazak01*. Timestamped data corresponding to the two machines is pulled from the SMS Test Bed and displayed on the side-panels next to the video feed as long as they are in view. In doing so, an operator is able to remotely identify which machines are currently producing value or are experiencing downtime.

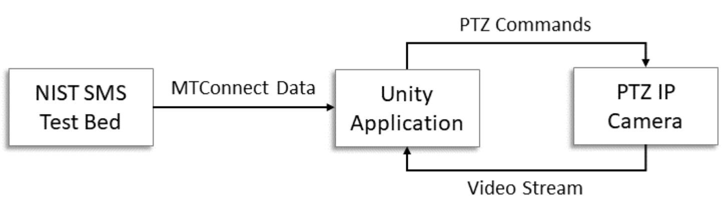

Figure 1: Prototype process diagram.

3.1 Approach Limitations

While our prototype serves as a proof of concept designed to make use of simple ubiquitous technologies, e.g., IP cameras and QR codes, there are some obvious limitations to this approach.

First of all, even more so than AR markers, QR codes are not designed for far-field use, being difficult to detect and especially decode across large distances, unless scaled appropriately, which in itself is often infeasible or impractical. Secondly, there are scalability issues concerning

[3]Access to data generated by the SMS Test Bed can be found here: http://smstestbed.nist.gov/.

the number of QR codes in view at any given time, which in turn affects how well users can access data for which they are interested. In other words, the simultaneous detection of multiple machines would also increase the amount of unwanted data on the screen. Additionally, each machine needs to be physically tagged and the markers need to be in the camera's line of sight. Our design allows users to manipulate the line of sight of the camera by accessing its PTZ capabilities. However, detectable markers still (a) need to be oriented orthogonal (or nearly orthogonal) to the camera and (b) cannot be obstructed by other physical objects. This suggests that multiple cameras would be needed to ensure that no machine is obstructed from view. Lastly, while not necessarily a drawback for some use cases, this approach is limited to displaying 2D data over the video feed, given a lack of 3D spacial understanding of the scene.

3.2 Augmenting the Video Feed with Digital 3D Data

Building on the previously described prototype, we present an additional use case showcasing the potential of replacing QR Codes with AR-ready fiducial markers that are more easily tracked by design. Given that the markers can be used to compute the camera pose, 3D objects can be overlaid onto the video stream with an accurate perspective, in addition to the 2D data of the previous use case. This offers the potential of digital models of machines being superimposed over their physical counterparts or displaying any other spatial information in the scene, perhaps in different layers depending on the use context. This is illustrated in Fig. 3, where three CAD models of CNC machines are overlaid on the video feed using the MAXST AR software development kit[4]. A model of the DIVE Lab was created, where three points of interest (the three tables) were mapped. Using this model, the points of interest can be accurately tracked while moving the camera by having a single AR marker in view, thus mitigating some of the issues highlighted by the previous use case such as individual machine tagging.

Note that these prototypes have been implemented and tested in a typical room scale laboratory setting. As described in Section 2, while the prototypes work at this scale, they might not immediately scale appropriately for the desired use case, i.e., a large, crowded, and complex

Figure 2: Real-time MTConnect data is shown for the detected machines.

[4]For more information, visit http://maxst.com.

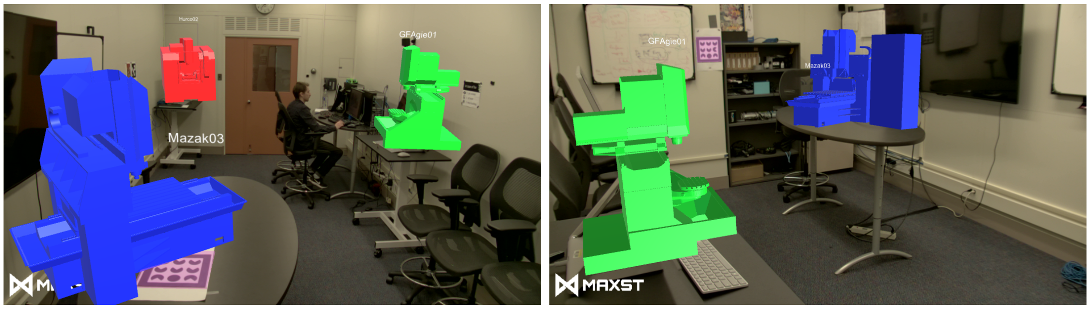

Figure 3: CAD models of CNC machines overlaid on the video feed.

physical environment. Further work and experimentation is required to implement prototype iterations at a larger scale to better understand the full scope of challenges.

4 Future Directions & Opportunities

In this paper, we presented two deployed prototypes that leverage the NIST SMS Test Bed and the DIVE Lab to explore issues related to far-field object tracking in the context of production systems. Based on our preliminary findings, we discuss research directions, including opportunities for standards development.

AR framework developers are pushing towards markerless detection, yet for prototyping and testing purposes, marker-based tracking is still very prevalent throughout industrial implementations. In the case of production systems, where the primary manufacturing assets, such as cranes, industrial robots, and CNC machines, are affixed to a particular location, geospatial definitions can offer more precise data to anchor critical objects in a scene. In other words, rather than relying on techniques such as SLAM to build a feature-map of large area such as a factory floor, we believe that building an as-planned indoor representation might provide additional benefits, such as the ability to include semantics related to the tracked elements and incorporate domain-specific information akin to our MTConnect data streams. Additionally, this approach would alleviate some of the challenges discussed earlier related to far-field marker tracking by removing the need of individual machine tagging and minimizing the number of markers required for the whole scene. This would have the potential for reducing burdens, in terms of both cost, time, and equipment, for testing industrial AR-based prototyping iterations. Furthermore, opportunities exist for the development of measurements methods for the efficiency and appropriateness of as-planned indoor representations.

Moving forward, we plan to explore how richer, pre-defined geospatial representations can influence industrial AR implementation. Based on our early findings, a geospatial representation of a room simplifies the implementation of far-field tracking systems for indoor use. In our prototype, we plan to leverage IndoorGML [8], a standard data format from the Open Geospatial Consortium (OGC)[5]. Specifically designed for formally describing scenarios that require positional data of physical entities inside buildings, IndoorGML[6] provides a framework for geospatial information that relates properties and features of indoor spaces within a flexible framework. We chose IndoorGML for implementation due to existing available tools built around the technology, including an editor for generating IndoorGML documents [7] and

[5]For more information, refer to https://www.opengeospatial.org.
[6]For more information, refer to http://www.indoorgml.net.

an existing link to scene generation in Unity[7], a 3D development platform. Leveraging such geospatial definitions, we hypothesize that (1) less markers would be required for tracking a set of objects, (2) the burden of introducing additional spatially-aware sensors into the predefined environment would be lessened, and (3) such representations coupled with vision-based tracking, like SLAM, would provide more robust object tracking solutions.

Similar to other visualization-driven technologies, industrial AR must overcome a divergence of two traditionally separated standards development communities: (i) the primarily gaming-driven AR frameworks contributed by standards developments organizations (SDOs) such as the Khronos Group and OGC and (ii) data interoperability solutions from SDOs focused on smart manufacturing systems such as the MTConnect Institute[8] and the Open Platform Communications (OPC) Foundation[9]. We believe that our work will provide more guidance and direction for the revision or extension of existing standards and/or opportunities for new standards development. For example, we plan to test how data elements standards by the AR-focused SDOs for affixed and mobile objects (e.g., load-bearing columns and furniture, respectively) relate to analogous manufacturing assets, such as CNC machining centers (affixed) and tooling carts (mobile). We believe that such exploratory tasks will pave the way for the conformance mappings between the two standards communities.

Disclaimer

Acknowledgements

We thank Dr. Moneer Helu, Frank Riddick, Dr. Goudong Shao, Dr. Jeremy Marvel, and the anonymous reviewers of the MBE Summit for their valuable feedback to improve this paper.

References

[1] Mark Billinghurst, Adrian Clark, Gun Lee, et al. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015.

[2] Robert O Castle, Georg Klein, and David W Murray. Wide-area augmented reality using camera tracking and mapping in multiple regions. *Computer Vision and Image Understanding*, 115(6):854–867, 2011.

[3] Youngkwan Cho and Ulrich Neumann. Multiring fiducial systems for scalable fiducial-tracking augmented reality. *Presence*, 10(6):599–612, 2001.

[4] David Claus and Andrew W Fitzgibbon. Reliable fiducial detection in natural scenes. In *European Conference on Computer Vision*, pages 469–480. Springer, 2004.

[5] Fabian Doil, W Schreiber, T Alt, and C Patron. Augmented reality for manufacturing planning. In *Proceedings of the Workshop on Virtual Environments*, pages 71–76. ACM, 2003.

[7] For more information, refer to https://unity.com.
[8] For more information, refer to https://www.mtconnect.org.
[9] For more information, refer to https://opcfoundation.org.

[6] Steven Henderson and Steven Feiner. Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1355–1368, 2010.

[7] Jung-Rae Hwang, Hye-Young Kang, and Jin-won Choi. Development of an editor and a viewer for indoorgml. In *ISA*, pages 37–40, 2012.

[8] Hae-Kyong Kang and Ki-Joune Li. A standard indoor spatial data model—OGC IndoorGML and implementation approaches. *ISPRS International Journal of Geo-Information*, 6(4):116, 2017.

[9] Christian Koch, Matthias Neges, Markus König, and Michael Abramovici. Natural markers for augmented reality-based indoor navigation and facility maintenance. *Automation in Construction*, 48:18–30, 2014.

[10] Pierre Malbezin, Wayne Piekarski, and Bruce H Thomas. Measuring artootkit accuracy in long distance tracking experiments. In *The First IEEE International Workshop Agumented Reality Toolkit,*, pages 2–pp. IEEE, 2002.

[11] SK Ong, ML Yuan, and AYC Nee. Augmented reality applications in manufacturing: a survey. *International Journal of Production Research*, 46(10):2707–2742, 2008.

[12] Gerhard Reitmayr, Tobias Langlotz, Daniel Wagner, Alessandro Mulloni, Gerhard Schall, Dieter Schmalstieg, and Qi Pan. Simultaneous localization and mapping for augmented reality. In *2010 International Symposium on Ubiquitous Virtual Reality*, pages 5–8. IEEE, 2010.

[13] DWF Van Krevelen and Ronald Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1–20, 2010.

[14] ML Yuan, SK Ong, and AYC Nee. Augmented reality for assembly guidance using a virtual interactive tool. *International Journal of Production Research*, 46(7):1745–1767, 2008.

[15] Xiang Zhang, Stephan Fronz, and Nassir Navab. Visual marker detection and decoding in ar systems: A comparative study. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, page 97. IEEE Computer Society, 2002.