Evaluating the suitability of several AR devices and tools for industrial applications

Evaluating the Suitability of Several AR Devices and Tools for Industrial Applications*

Edoardo Battegazzorre[0000−0002−1383−3285], Davide Calandra[0000−0003−0449−5752], Francesco Strada[0000−0001−9197−9100], Andrea Bottino[0000−0002−8894−5089], and Fabrizio Lamberti[0000−0001−7703−1372]

Politecnico di Torino, Dipartimento di Automatica e Informatica, Turin, Italy
{firstname.lastname}@polito.it

Abstract. In recent years, there has been an increasing interest in Industrial Augmented Reality (IAR) due to its prominent role in the ongoing revolution known as Industry 4.0. For companies and industries it is essential to evaluate carefully which of the developed AR-based technologies to adopt, and when, for tasks such as training, maintenance, assistance, and collaborative design. There is also a wide array of hardware and software alternatives on the market, characterized by a significant heterogeneity in terms of functionalities, performance and cost. With this work, our objective is to study and compare some widely available devices and Software Development Kits (SDKs) for AR by leveraging a set of evaluation criteria derived from the actual literature which have been deemed capable to qualify the above assets as suitable for industrial applications. Such criteria include the operative range, robustness, accuracy and stability. Both marker-based and marker-less solutions have been considered, in order to investigate a wide range of possible use cases.

Keywords: Augmented Reality · Industrial Augmented Reality · Evaluation · Marker Detection · Positional Tracking

1 Introduction

The last years have been characterized by a growing interest in Industrial Augmented Reality (IAR) applications, mainly because of their key role in the ongoing developments framed under the Industry 4.0 umbrella [7, 2]. Moreover, the continuous advancements in the field of Augmented Reality (AR) translate into an ever-broader choice of devices characterized by progressively a lower cost, higher performance, and a growing set of capabilities that can be relevant for industrial scenarios. However, these hardware and software solutions are characterized by significant heterogeneity in terms of technical specifications and features offered. Therefore, it is essential to provide methods for comparing them and assessing their behavior in industrial settings. In [3], Duenser et

* This work has been supported by a study funded by SIPAL Spa under the research project titled "Cantiere Tecnologico per infrastrutture militari e civili (Unmanned vehicles and Virtual facilities)", Regione Puglia, and by the VR@POLITO initiative.

al. investigated, through a literature survey, different techniques for performing evaluations in the context of AR, providing an overview of each technique and discussing how they were used in the specific study. They also proposed a taxonomy of such techniques, which classifies them into objective measures, subjective measures, qualitative analyses, usability evaluations, and informal evaluations. In particular, objective measures appeared to be the most adopted evaluation technique over the time-span considered in [3]. As an example, the authors of [1] presented an objective evaluation of different IAR tools and Software Development Kits (SDKs) for a specific industrial setting, namely the shipyard. The research focused on evaluating marker-based detection techniques; the authors intentionally disregarded marker-less 3D model-based tracking approaches, stating that they would not satisfy the minimum requirements for industrial applications in terms of hardware performance, storage, and robustness in highly dynamic environments.

Despite the effectiveness of objective measures, investigating specific characteristics of selected devices and tools requires the use of other evaluation approaches. For example, the authors of [5] leveraged a subjective evaluation protocol for assessing the capabilities of Mixed Reality (MR) in the field of production and logistics. Users involved in the study were first asked to engage in an "item-picking" scenario (using the Microsoft HoloLens™) and then to fill in a questionnaire investigating aspects such as realism of the projection, visibility of the holograms, ease of interaction, comfort, ergonomics, and satisfaction.

The evaluation protocol should also take into account the specific requirements of the industrial context, as done in [10, 8]. Although requirements such as cost-effectiveness, data security, and applicable regulations have to be partially or totally addressed at the software level, the device to use still plays an essential role in terms of set-up time, system reliability, quality of presentation, real-time capabilities and ergonomics. To this end, the possibility to use, in industrial contexts, devices and technologies coming from the entertainment sector (like in [6, 11]) requires careful analyses. If it is true that these devices are generally quite affordable and offer a broad set of general-purpose features, they are also characterized by lower precision, reliability, and employability compared with solutions specifically designed for the industry. For instance, in [9], the authors proved that Lighthouse 1.0[1] (the tracking technology exploited by some HTC® Vive™ devices) is unsuitable for a particular class of scientific tasks involving the accurate visual stimulation of self-motion in a virtual environment.

The present work provides an objective evaluation of three AR devices, namely the Microsoft HoloLens™, the Samsung® Galaxy Tab S4 Android™, and the DreamWorld DreamGlass (each combined with different AR libraries) to assess the viability of such technologies in industrial applications and environments. In particular, we focused on different tracking methods available for each of the considered device-SDK pairs. Hence, we first assessed the performance of the different marker detection techniques, and then we moved to analyze the marker-less sensor-based tracking techniques (either inside-out or outside-in)

[1] https://www.vive.com/us/accessory/base-station/

Table 1. Marker-based configurations (devices and libraries) considered in the study.

Device	Library	Type of marker	Marker sample	Maximum number of markers
HoloLens™ / Android™	Vuforia	ArMarker		2.147.483.648
HoloLens™	SpectatorView Marker Detection	ArUco 6×6		1.000
Android™ Tablet	ARCore	ArMarker		2.147.483.648

Table 2. Marker-less configurations (devices and libraries) considered in the study.

Device	SDK	Tracking technology
DreamGlass + NOLO	DreamWorld + NOLO VR	Outside-in
HoloLens™	Holo-Toolkit	Inside-out

offered by these devices. Table 1 and Table 2 summarize the different configurations (device, SDK, type of marker, etc.) considered in the evaluation. The methodology used to assess marker-based IAR was inspired by the work reported in [1], extending the previous analysis to encompass also different marker types and sizes, more evaluation metrics, and a larger number of devices/SDKs. The sensor-based analysis applied the same methodology exploited in [9] to different positional tracking techniques. It should be noted that, following the approach in [1], we did not consider the 3D model-based tracking techniques because of their poorer suitability for the context of interest.

2 Selected Hardware and Software

This section illustrates the different hardware and software solutions included in the evaluation, which were selected to cover a broad set of characteristics. Thus, the analysis included devices that are both enterprise-grade and consumer-grade, head-mounted and hand-held, tethered and untethered, etc. On the software side, native solutions provided through the official SDKs were considered when available, and the Unity game engine (2018.4) was leveraged to implement the measuring tools required for the experiments.

2.1 Microsoft HoloLens™

The Microsoft HoloLens™ (1st generation) is an optical see-through AR Head-Mounted Display (HMD). The device uses an inside-out SLAM (Simultaneous Localization And Mapping) positional tracking to register the virtual contents with the surrounding (real) space. The principal method for aligning virtual and real contents is based on so-called spatial "anchors", i.e., geometric descriptors attached to specific points that help the device to track them over time. Another tracking possibility relies on the front-mounted RGB camera of the HMD for detecting fiducial markers, with different approaches detailed in the following.

ArUco Marker Detection The HoloLens™ supports an official marker tracking solution based on the ArUco markers [12, 4]. The plug-in provides two different detection strategies:

- *stationary*, for markers that are known to be fixed in space; this strategy applies heuristics to filter out noisy detections;
- *moving*, for non-stationary markers; this strategy averages multiple observations and does not apply any noise reduction filter.

In the evaluation, both strategies were separately analyzed using the default algorithms configuration provided by the vendor.

Vuforia This library[2] probably represents the simplest method to implement the marker detection with the HoloLens™, since it is officially supported and allows for the combination of marker detection capabilities with the device's spatial mapping in a modality named Extended Tracking. The Extended Tracking feature was intentionally disabled in our evaluation. Regarding the implementation details, in the analysis, Vuforia 8.3.8 was used, generating the markers with the ArMarker[3] tool.

2.2 Android™

Android™ devices can exploit AR through the native ARCore SDK or a set of third-party libraries (e.g., Vuforia, Wikitude[4], Kudan AR SDK[5]). In this work, the native AR solution was tested against the Vuforia library using an Android-based tablet device, namely, a Samsung® Tab S4).

[2] https://www.ptc.com/en/products/augmented-reality/vuforia
[3] https://github.com/shawnlehner/ARMaker
[4] https://www.wikitude.com/
[5] https://www.xlsoft.com/en/products/kudan/ar-sdk.html

ARCore This library offers a series of tools to implement a marker-based tracking solution through the Augmented Images API[6]. The API can recognize a maximum of 20 concurrent reference images providing as output position, rotation and size of each image. The trackable images must be included in an internal database that has a maximum size of 1000 elements. Similarly to the HoloLens, ARCore can also perform SLAM to infer the device's position in the real environment and, thus, use it to improve the robustness and stability of the marker-based tracking. However, to better characterize the API, this feature was disabled in the evaluation. Again, ArMarker was used for generating ARCore-specific markers.

Vuforia The Vuforia library for Android™ is the same used for the HoloLens™. Thus, the same settings were used (i.e., version 8.3.8, Extended Tracking turned off and markers generated with ArMarker).

2.3 DreamWorld DreamGlass™ and NOLO™

The DreamGlass™ is an optical see-through AR HMD characterized by a native 3-DOF (Degrees of Freedom) tracking. Opposite to the HoloLens™, this is a tethered device and, thus, requires an external computation unit (a personal computer running Windows or an Android™ smartphone). The device is equipped with two front cameras, an IR camera used for gesture recognition, and a RGB one. According to the vendor, these cameras can be enabled simultaneously only in Android™. This fact, along with the lack of support for native marker detection and the impossibility to deploy Vuforia applications as Windows stand-alone executables, make it particularly hard to support marker-based tracking. To cope with this limitation, the vendor suggests to integrate the headset with the NOLO™ 6-DOF optical-inertial motion tracking system. The NOLO™ kit includes a single table-top base station, a wired HMD tracker, and two wireless controllers. In the evaluation, the DreamGlass™ and the NOLO™ were used in combination with a backpack laptop (precisely, the MSI VR One 7RE) in order to eliminate the encumbrance of the wired components (i.e., the HMD and its tracker).

3 Methodology

The definition of the evaluation methodology was preceded by an investigation phase aimed to identify the basic requirements for industrial AR applications. As reported in [13], IAR applications can be subdivided in many use cases: manual assembly, robot programming and operations, maintenance, process monitoring, training, process simulation, quality inspection, picking process, operational setup ergonomics and safety. Some features may be more or less relevant in relation to the various industrial activities. For example, the device field of view

[6] https://developers.google.com/ar/develop/c/augmented-images

(FOV) may be a key factor for the maintenance/assembly of large elements (e.g. vehicles), but it may be less relevant for other activities (e.g. item picking). As a result of the investigation, the need for a proper tracking technology to register virtual contents with real objects proved to be the most common requirement among all the use cases. Based on this consideration, the initial focus was put on marker detection techniques, considering alternative methods when the former were unavailable or not applicable.

3.1 Marker-based AR

We defined the following four metrics for assessing the viability of each device or SDK in an IAR context: *Detection Distance* (DD), *Loss Distance* (LD), *Detection Rate* (DR) and *Stability* (ST). The experimental setup for the evaluation of each metric is illustrated hereinafter.

Detection Distance (DD) and Loss Distance (LD) When focusing on the evaluation of a marker detection technique, one of the first aspects to consider is the maximum distance at which the device is capable of tracking a marker (DD distance). To measure this distance, we placed a marker on a tripod at the same height as the device tested. Then, the experimenter began to walk towards the marker, starting from a distance at which the marker detection was not occurring and stopping when the first detection was registered (Figure 1-2, left). Afterwards, the experimenter measured the LD distance by slowly walking backward while keeping the marker in the device's FOV until the tracking was lost. It is worth noting that we could measure the LD metric only for the Vuforia and ARCore configurations. The reason is that, after detection, both SDKs exploit tracking techniques to maintain marker identification as long as possible even for considerable distances. On the contrary, ArUco Marker Detection is characterized by a completely different behavior since it only provides information about marker detection and lacks a tracking component. Hence, in order to filter out spurious or unrepeatable detection events, the ArUco DD value was measured as the distance at which the detection occurs steadily with at least a frequency of 1Hz. In contrast, LD was measured as the first sampled distance at which the above requirement was not fulfilled.

For all the SDKs, DD and LD measures were repeated with different configurations of three parameters, as summarized below (Figure 1, right):

- four angles of approach (0°, 30°, 45° and 60°);
- three marker sizes (5, 10 and 20 cm);
- two lighting conditions (artificial and natural light).

Detection Rate (DR) A further metric considered in the evaluation refers to the marker identification capabilities of the considered devices and SDKs within a given time interval measured at predetermined distances (Table 3). The DR metric is computed as:

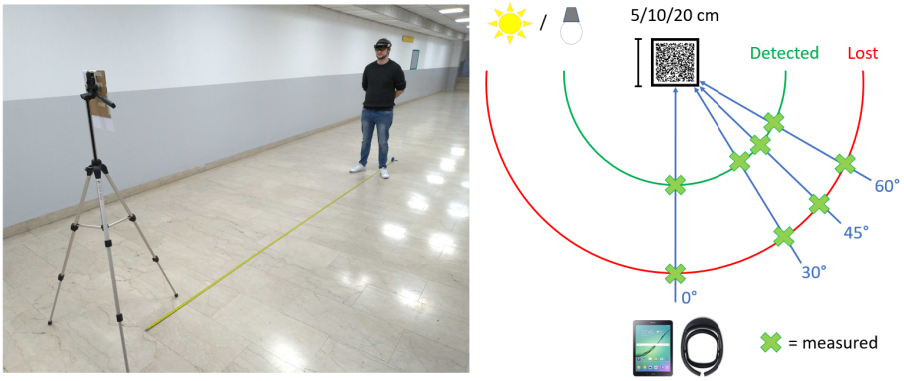

Fig. 1. Setting for DD/LD/DR/ST with artificial light using the HoloLens™ (left) and experimental setup for DD/LD (right)

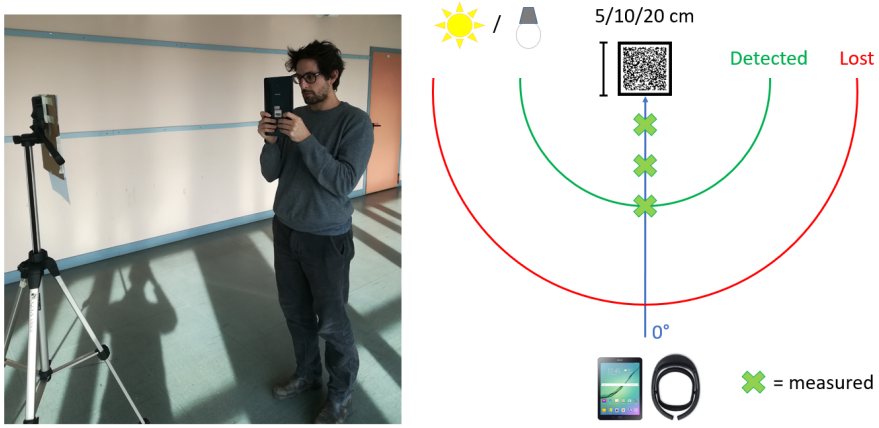

Fig. 2. Setting for DD/LD/DR/ST with natural light using an Android tablet (left) and experimental setup for DR/ST (right)

$$DR(\%) = \frac{N_D}{N_T} \tag{1}$$

where N_D is the number of times the detection occurred in less than 5 seconds, and N_T is the number of detection attempts (10 in the evaluation). The setup of this experiment is similar to the first one, but using a single angle of approach (namely, 0°, see Figure 2, right).

Stability This metric represents the capability to detect the same position and rotation of a static marker over time. The setup used for this experiment was the same used for the DR metric (Figure 2, right). After detecting the marker with an angle of approach of 0°, the detected positions and rotations along all

Table 3. Distances for DR and ST measures: ✓ indicates a distance included in the analysis, (✓) a distance included only if necessary (DD < 0.625 m or DD > 15 m).

Measure \ Distance (m)	<0.625	0.625	1.25	2.5	3.75	5	7.5	10	12.5	15	>15
DR	(✓)		✓	✓		✓	✓	✓			(✓)
ST	(✓)	✓	✓	✓	✓	✓	✓	✓	✓	✓	(✓)

the three axes were sampled for 5 seconds at 50Hz. Measurements were repeated at multiple fixed distances (Table 3) for the three marker sizes (5, 10 and 20 cm), and with both natural and artificial lighting. Positional Stability (PS) is computed as the pooled standard deviation of all samples on each axis as:

$$PS = \sqrt{\frac{\theta_x^2 + \theta_y^2 + \theta_z^2}{3}} \tag{2}$$

where θ_x, θ_y and θ_z represent the samples' variances for each axis. Assuming a completely static marker, the ideal PS should be 0, meaning that all registered values are equal. Angular Stability (AS) is measured as the average of the angular dispersion on all the three axes which, for a given axis, is computed as:

$$R = \sqrt{\left(\frac{\sum cos\alpha}{n}\right)^2 + \left(\frac{\sum sin\alpha}{n}\right)^2} \tag{3}$$

where AS ∈ [0,1]. The upper bound indicates the concentration of all the samples in the same direction and represents the optimal AS behavior, whereas the lower bound indicates uniform dispersion.

3.2 Marker-less AR

When marker detection is not available (or not suitable) for a given application, other techniques have to be considered. If the device supports positional tracking, this feature can be exploited to align the virtual content with its real counterpart over time. Hence, *Tracking Accuracy* (TA) becomes an essential metric for studying the suitability of a given configuration for the applications of interest.

Tracking Accuracy The evaluation of the TA for HoloLens™ and NOLO™ was based on collecting positional and rotational data on a grid of known reference points and comparing them with a ground truth.

The grid for the NOLO™ devices was composed of cells of size 60×60 cm included in the active tracking area specified by the vendor. This area has a range of 5 m and a FOV of 100°from the base station, resulting in a total of 60 valid points for the evaluation (Figure 4, left). The base station was positioned on a stand at 53.3 cm from the floor. During the experiment, the tracker under

Fig. 3. Measurements of TA for the NOLO^{TM} HMD (left) and controller (right).

testing was placed in each point of the grid, and its position was recorded for 1 second at a 50Hz frequency to minimize possible jittering effects. This procedure was repeated separately for the HMD tracker (installed on the DreamGlass^{TM}) and for one of its controllers in three different configurations:

- *front*: the controller is pointed towards the base station, parallel to the bisector of its FOV;
- *top*: the controller is pointed towards the ceiling;
- *side*: the controller is pointed sideways to the right (from the point of view of the base station).

To ensure the stability of the trackers during measurements, the HMD was placed on the ground (Figure 3, left), whereas the controllers were mounted on an adjustable tripod at 1.5 m from the floor (Figure 3, right), which is the height suggested by NOLO^{TM} for the calibration of the device. The mean positional and rotational errors for each axis (X, Y, and Z) were computed for each grid point by first collecting measurements over one second and then subtracting the ground truth. Finally, these differences were averaged again to obtain the mean positional and rotational errors for each configuration. We underline that, as detailed in Section 4.5, the controllers seem to suffer from a notable rotational drift during regular use (probably due to issues with sensor fusion between optical data from the base station and inertial data coming from the gyros/accelerometers). Thus, we tried to minimize this effect by re-calibrating the controllers every three measurements following the procedure suggested by the vendor.

A similar procedure was then adopted for the TA evaluation of the HoloLens^{TM} tracking. Since this tracking does not require any external sensors, the user could theoretically move and operate in an arbitrarily large area. With this fact in mind, in order to guarantee the same experimental conditions used to evaluate the NOLO^{TM} TA, a rectangular reference grid of 77 points superimposed to the previous one was drafted (Figure 4, right). An experimenter wearing the

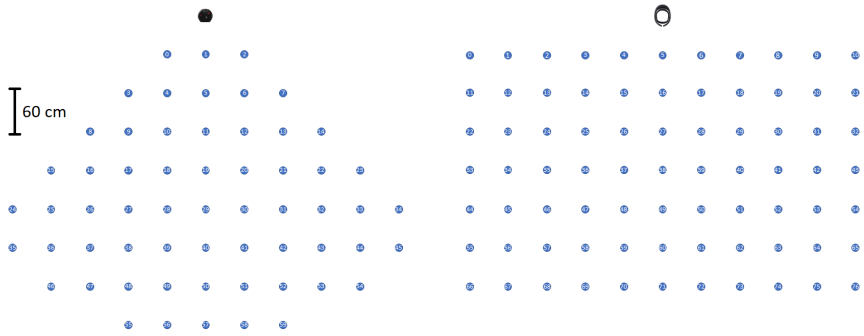

Fig. 4. Grids used to evaluate the TA for the NOLO™ (left) and HoloLens™ (right).

HoloLens™ carefully stepped onto each of the grid points, where the position and orientation of the headset were sampled, like before, at 50Hz for 1 second.

4 Experimental Results

In this section, the results of the experimental activities are presented, firstly focusing on the marker-based evaluation metrics described in Section 3.1, then concluding with the marker-less metric introduced in Section 3.2.

4.1 Detection Distance

Plots in Figures 5–8 summarize the data collected for four of the pairs device/SDK under analysis. Data concerning Android™/Vuforia have not been reported since the behavior was substantially similar to the HoloLens™ counterpart (Figure 5) in both lighting conditions. Results obtained with artificial lighting will be firstly exposed, followed by a discussion of the meaningful differences in comparison with natural lighting. In general, the plots show fairly predictable trends, in which the DD increases with the marker size and decreases with the angle of approach.

The pair showing the best performance was the HoloLens™/ArUco Moving (Figure 6), which reaches a DD of 16.85 m under the most favorable conditions (20 cm marker, angle of approach of 0°) and 1.96 m in the least favorable one (5 cm marker, angle of approach of 60°). The worst performing pair was Android™/ARCore (Figure 8), with a DD of 0.86 m under the most favorable conditions, and a complete detection failure in every test performed at 60°. All the other pairs showed maximum DD values between 2 and 3 m.

HoloLens™/ArUco Moving (Figure 6) had the most linear trend, while its Stationary counterpart (Figure 7) behaved very erratically as, counter-intuitively, the 0° angle of approach showed the worst performance. This result could be attributed to how the tracking algorithm works. One could speculate that, due to the filtering behavior of the given detection strategy, the framing of the full

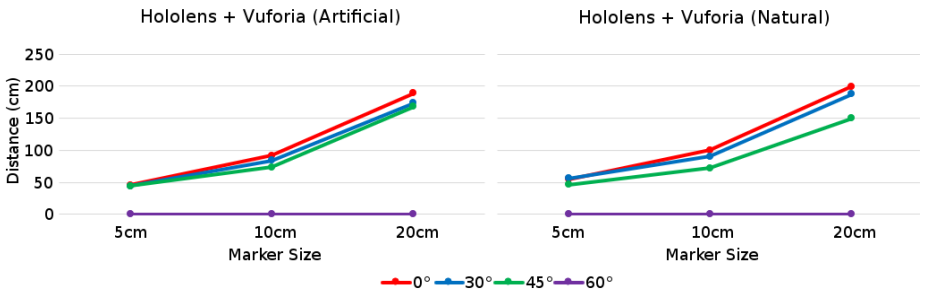

Fig. 5. DD for HoloLens™/Vuforia with artificial (left) and natural lighting (right).

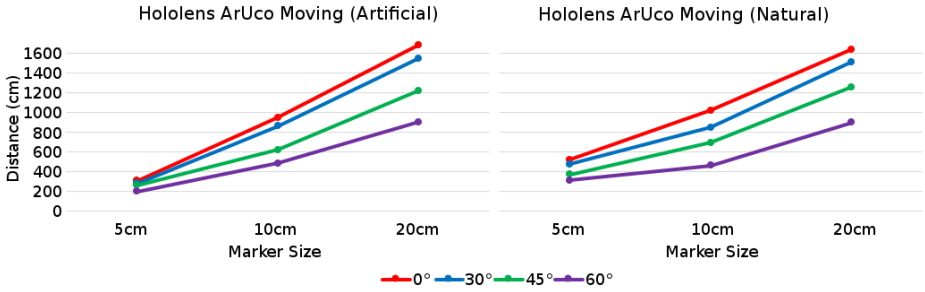

Fig. 6. DD for ArUco Moving with artificial (left) and natural lighting (right).

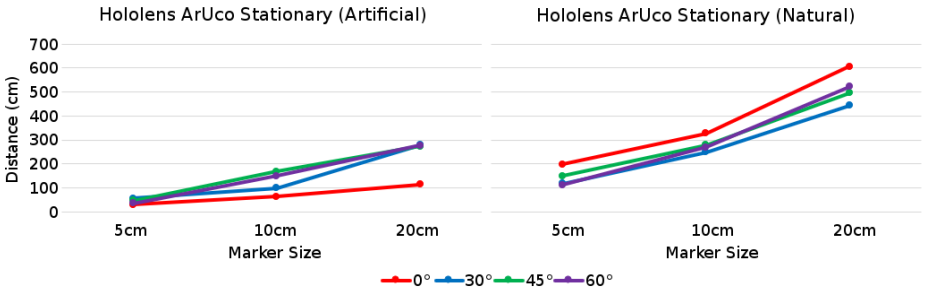

Fig. 7. DD for ArUco Stationary with artificial (left) and natural lighting (right).

marker at 0° yields a substantial number of noisy samples, which are then discarded when filtering is applied on position and rotation. When the experimenter approaches the marker at a wider angle, actual detections are less frequent so that they could appear as more precise from the heuristics' perspective and, thus, less prone to be discarded.

Vuforia (Figure 5) showed difficulties in detecting markers at the steepest angle (60°) in almost every case except for a few detections obtained on Android™ in artificial lighting conditions. The different specifications of the cameras on the various devices (including their distinct distortion parameters and sensitivity to light) could explain this behavior.

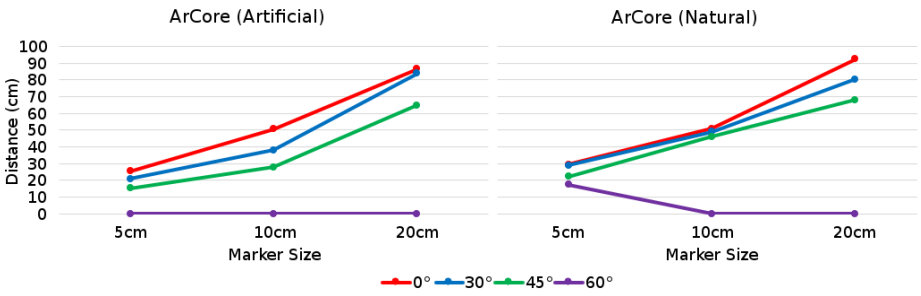

Fig. 8. DD for Android™/ArCore with artificial (left) and natural lighting (right).

With natural lighting, the trends remained the same for all the pairs. Again, the best performance was obtained by HoloLens™/ArUco Moving with DD values of 16.38 m and 3.1 m under the most/least favorable conditions. Android™ with ArCore still scored the lowest distances with 0.92 m under the best conditions and a complete failure under the worst ones. However, it can be observed that, under these conditions, only the smallest 5 cm marker was detected at 60°. This result can be explained by the fact that distances were so short (15.5 cm) that camera distortion could play a significant role, making it impossible to recognize larger markers. The only pair that showed improvements with the brighter, natural lighting was the HoloLens™/ArUco Stationary pair, having its DD values doubled (3 to 6 m in the best case scenario). Still, this SDK behaved oddly as the DD does not seem to be inversely proportional to the angle of approach. Again, this result underlines the sensibility of the algorithm to minimal lighting variations, possibly tied to the behavior of the filter that discards noisier detections.

4.2 Loss Distance

As previously mentioned, only Vuforia and ARCore were able to preserve the tracking farther than the initial DD thresholds, as shown in Figures 9–11. The blue part of each glyph represents the initial DD, whereas the orange part shows the threshold at which the tracking is lost (LD). In all the figures, data collected with both artificial (left) and natural lighting (right) are reported.

From the plots, it is evident that Vuforia outperforms ARCore, showing improvements that go from 92% to 338% of the initial DD. For ARCore, the improvements are more marginal, ranging from 15% to 96% of the initial DD. It is worth pointing out that natural lighting has a negative effect on the average range extension, especially for Vuforia, for which the average improvement decreases from 261% to 162%. It should be noted, however, that there is a trade-off between tracking distance and tracking stability, as illustrated in the following sections.

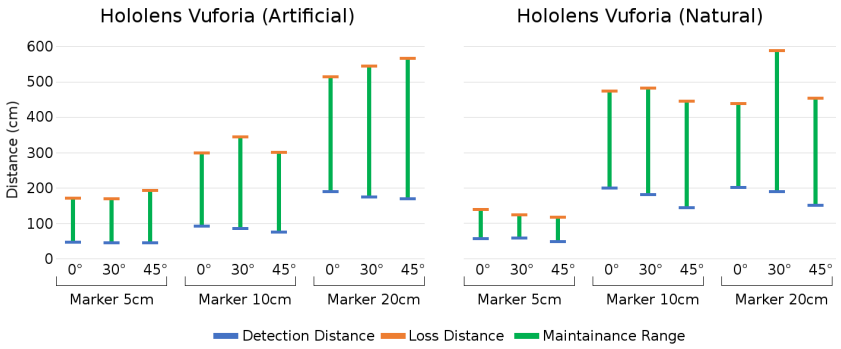

Fig. 9. DD and LD for HoloLens™/Vuforia with artificial (left) and natural (right) lighting.

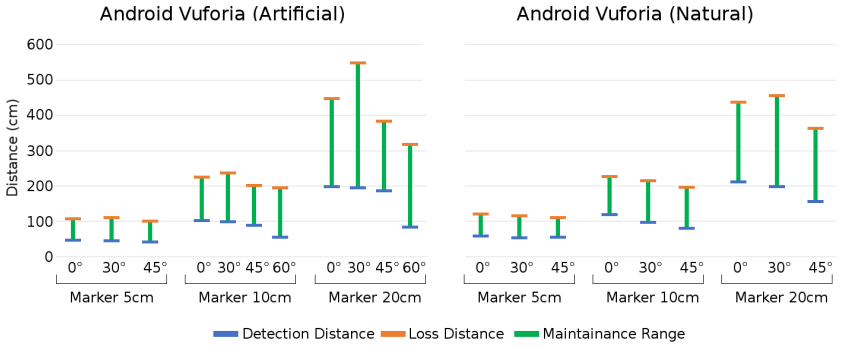

Fig. 10. DD and LD for Android™/Vuforia with artificial (left) and natural (right) lighting.

4.3 Detection Rate

The results for this metric are reported in Figure 12. Two very distinct behaviors can be observed. DR for Vuforia and ARCore immediately go from 100% to 0% as soon as the measurement is performed at distances larger than the DD threshold (Section 4.1). On the contrary, all the ArUco detections were successful (100%) when attempted at distances smaller than the DD, and DR continuously decreases as the HoloLens™ moves away from the DD threshold. This degradation is more evident under natural lighting conditions.

4.4 Stability

Overall, both PS and RS showed no remarkable differences between artificial and natural lighting. For this reason, the discussion will focus on the behavior of PS and AS in artificial lighting conditions. In Figure 13, PS (top) and AS (bottom) are presented as a function of distance for all the five configurations.

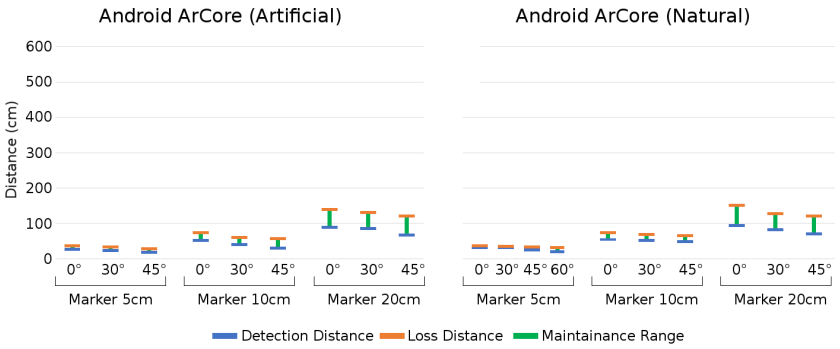

Fig. 11. DD and LD for Android™/ArCore with artificial (left) and natural (right) lighting.

Measurements were performed at fixed distances (Table 3), not exceeding the measured LD for the specific device/SDK pair; for this reason, some lines in the plots fall straight to zero over a certain distance.

PS is significantly low, generally under 0.02, and presents a stable behavior independent of the distance at which it was measured. Nonetheless, HoloLens™ ArUco Moving was the only pair capable of detecting markers at a distances larger than 5 m. Over this threshold, PS starts to increase notably, deteriorating up to 30 times (at 12.5 m) compared to the closest point (62.5 cm). Finally, PS values are comparable also across different marker sizes. However, the number of measurable PS is smaller, given the shorter range in which the marker was detectable (Sections 4.1 and 4.2).

Similarly, AS results were very stable when observed at close distances, with an average angular dispersion close to 1 for all devices/SDKs, except for Android™ ARCore, which never reached a value higher than 0.7. The deterioration of AS at increasing distances follows, in general, a linear behavior with a 20 cm marker. Contrary to PS, degradation is more evident reducing the marker size. However, AS for HoloLens™ ArUco Moving remains higher than 0.5, whereas the AS values of the other pairs fall in the [0.2, 0.4] range.

A general exception regards HoloLens™ ArUco Stationary with artificial lighting, for which a true range does not exist (DD values were always below 1.25 m). For this reason, it was possible to sample PS and AS only at a single point, at which the device/SDK showed again the best performance among all the five configurations.

4.5 Tracking Accuracy

The plots in Figures 14-18 present the differences between the average measured positions and rotations and the ground truth. Both the single points and their interpolation surfaces are shown, providing an overview of the topology of the plane reconstructed by the tracking devices. An illustration of the rotational errors is also provided to the right.

Configuration			Distance (m) < 1.25	1.25	2.5	5	7.5	10	> 10
HoloLens ArUco STAT	5	N	100%	30%	0%	0%	0%	0%	0%
		A	100% (29.3)	0%	0%	0%	0%	0%	0%
	10	N	100%	100%	50%	20%	0%	0%	0%
		A	100% (62.5)	0%	0%	0%	0%	0%	0%
	20	N	100%	100%	100%	80%	60%	30%	0%
		A	100% (114)	0%	0%	0%	0%	0%	0%
HoloLens ArUco MOV	5	N	100%	100%	100%	100%	0%	0%	0%
		A	100%	100%	100%	0%	0%	0%	0%
	10	N	100%	100%	100%	100%	90%	80%	0%
		A	100%	100%	100%	100%	100%	90%	0%
	20	N	100%	100%	100%	100%	100%	100%	50% (18)
		A	100%	100%	100%	100%	100%	100%	100% (17)
HoloLens Vuforia	5	N	100% (0.55)	0%	0%	0%	0%	0%	0%
		A	100% (46)	0%	0%	0%	0%	0%	0%
	10	N	100% (1.0)	0%	0%	0%	0%	0%	0%
		A	100% (92)	0%	0%	0%	0%	0%	0%
	20	N	100%	100%	0%	0%	0%	0%	0%
		A	100% (189)	0%	0%	0%	0%	0%	0%
Android Vuforia	5	N	100% (0.56)	0%	0%	0%	0%	0%	0%
		A	100% (44.5)	0%	0%	0%	0%	0%	0%
	10	N	100% (1.17)	0%	0%	0%	0%	0%	0%
		A	100% (50)	0%	0%	0%	0%	0%	0%
	20	N	100%	100%	0%	0%	0%	0%	0%
		A	100% (86)	0%	0%	0%	0%	0%	0%
Android ARCore	5	N	100% (0.29)	0%	0%	0%	0%	0%	0%
		A	100% (25.4)	0%	0%	0%	0%	0%	0%
	10	N	100% (0.51)	0%	0%	0%	0%	0%	0%
		A	100% (50.6)	0%	0%	0%	0%	0%	0%
	20	N	100% (0.92)	0%	0%	0%	0%	0%	0%
		A	100% (196)	0%	0%	0%	0%	0%	0%

Fig. 12. DR values for all the marker sizes (5, 10 and 20 cm) and both lighting conditions (natural and artificial). In brackets, under the DR value, the distance at which DR has been additionally sampled (cm below the minimum, m above the maximum).

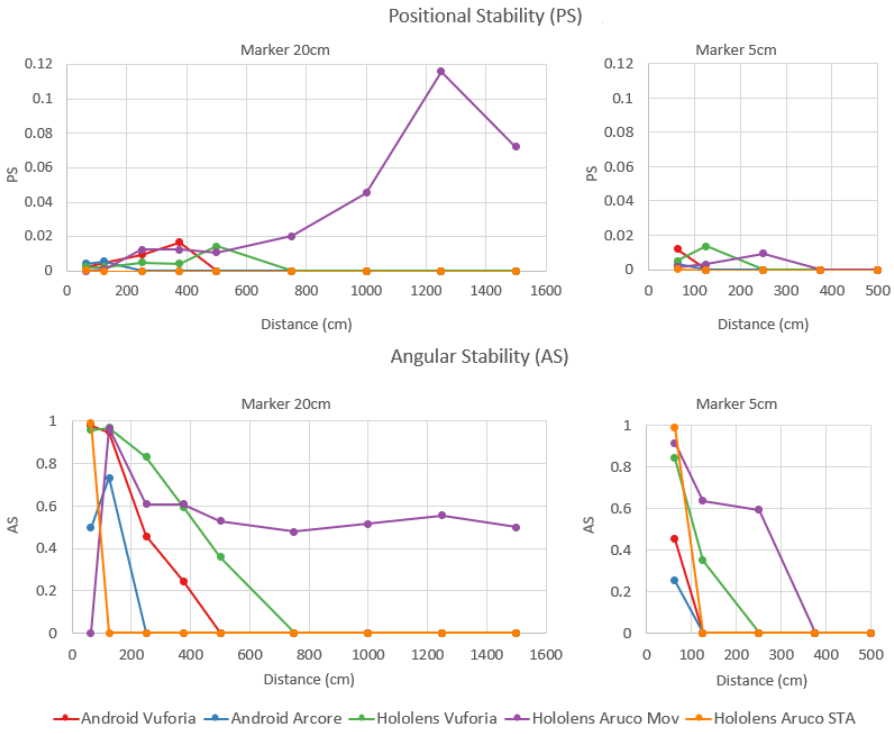

Fig. 13. PS (top) and AS (bottom) for markers sized 20 cm (left) and 5 cm (right) under artificial lighting conditions.

NOLO™ For the NOLO™ HMD, the calibration procedure suggested by the vendor was strictly followed. However, this calibration method did not take into account the height of the DreamGlass™, resulting in a collection of samples having a constant offset on both height (Y axis) and pitch (rotation around the Z axis). Thus, raw data were later compensated using these known offsets (13 cm and 23.9°). Figure 14 shows a variable positional error (especially in the area to the left of the base station close to the limits of the FOV), which generally gets worse moving away from the base station. The average positional error is 9.5 cm and 6.9 cm on the X and Z axis, respectively; the Y axis is the most critical, with an average error of 16.8 cm. The rotational tracking is relatively accurate (especially compared to the NOLO™ controllers, as discussed in the following paragraphs) with an average error of 1.6°, 3.46° and 0.76° on the X, Y and Z axis respectively.

Before examining the data gathered with the controller (Figure 15–17), it is worth recalling that the controller was tested in three different configurations. The average positional errors for these configurations were:

- *front*: 6.7 cm (X), 4.6 cm (Y), 12.2 cm (Z);
- *top*: 5.1 cm (X), 8.0 cm (Y), 7.3 cm (Z);

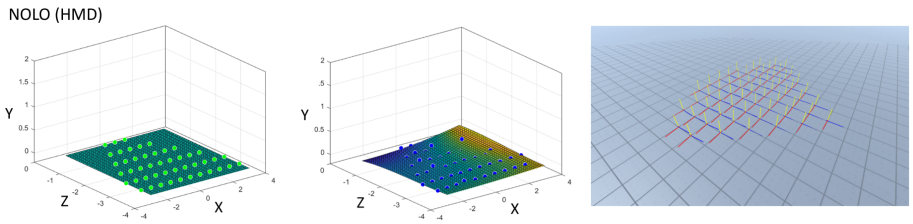

Fig. 14. TA results: actual positions on the grid (left) for the NOLO^{TM} HMD, measured positions (center), and measured rotations (right).

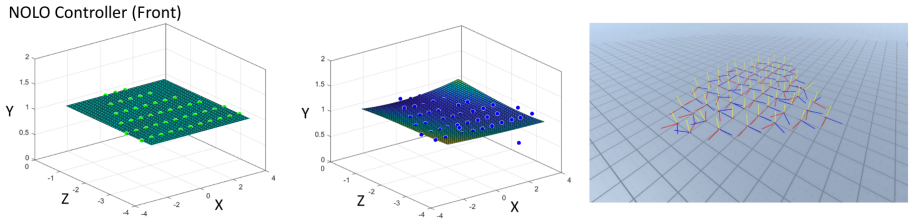

Fig. 15. TA results: actual positions on the grid (left) for the NOLO^{TM} controller (front), measured positions (center), and measured rotations (right).

– *side*: 24.6 cm (X), 14.6 cm (Y), 66.2 cm (Z).

While these errors are comparable for the first two configurations, there is a substantial decay in accuracy for the last one, especially on the Z axis with an average error of 66.2 cm (more than one whole square on the grid).

Data also indicate considerable errors for the measured rotations. In some specific cases, the errors are almost 180° away from the actual rotation. The average rotational errors were:

– *front*: 4.3°(X), 99.9°(Y), 5.3°(Z);
– *top*: 6.9°(X), 52.2°(Y), 23.6°(Z);
– *side*: 4.5°(X), 29.6°(Y), 3.7°(Z).

It can be noticed that the performance of the controller appears to be dramatically worse than that of the HMD. These errors could be ascribed to a drift caused by issues in sensor fusion of optical and inertial data. This can justify the frequent necessity of a controller re-calibration, in order to keep it aligned to the virtual counterpart.

HoloLens^{TM} As ut can be observed from the interpolation surface (Figure 18, center), the HoloLens^{TM} appears to be more accurate than the NOLO^{TM}; the average error is 2.7 cm on the X axis, 0.9 cm, on the Y axis and 2.8 cm on the Z axis. The same consideration applies to rotations, whereby discrepancies from the real values are contained (1.1°, 0.6°, Z: 0.5°for the X, Y and Z axis, respectively).

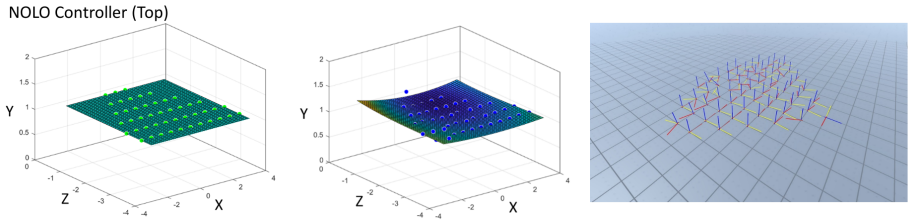

Fig. 16. TA results: actual positions on the grid (left) for the NOLO^{TM} controller (top), measured positions (center), and measured rotations (right).

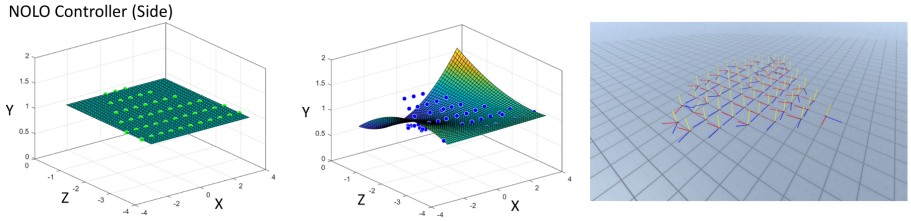

Fig. 17. TA results: actual positions on the grid (left) for the NOLO^{TM} controller (side), measured positions (center), and measured rotations (right).

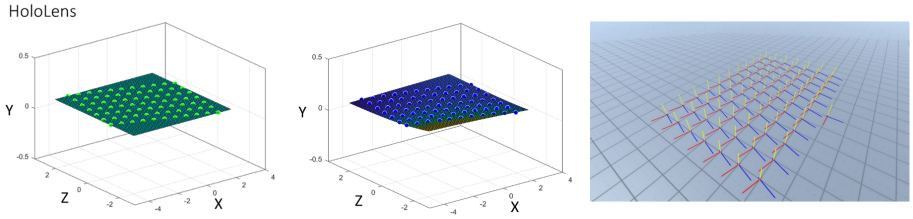

Fig. 18. The real positions on the grid (left) for the HoloLens^{TM}, the measured positions (center) and rotations (right)

5 Conclusions

The objective evaluation described in this work highlighted a series of different behaviors across the considered devices and SDKs. These differences can help in guiding the choice of the specific hardware or library with the aim to ensure a reliable and usable AR experience in industrial environments.

When the capability to detect markers from large distances is required, the ArUco library proved to outperform the alternatives (detection up to 16 m). However, at such distances, tracking stability is quite low. On the contrary, Vuforia requires first to detect the marker at a close distance (always below 2 m), but after detection, it can maintain the tracking up to four times that distance. Moreover, for distances between DD and DL, only the angular tracking stability deteriorates as the device gets farther from the marker, whereas the positional

tracking remains stable. This behavior is similar to the one observed for the Hololens™ ArUco Moving configuration.

Overall, no remarkable differences were detected in the two analyzed lighting conditions, with the exception of the ArUco Stationary configuration. In this case, the tracking algorithm seems to suffer from artificial lighting. Under such conditions, marker detection was affected by a strange behavior, showing better performance with a wider angle of approach, probably ascribable to the effects of the noise detection filter.

Regarding detection rate, results did not provide particularly relevant information, except that Vuforia and ArCore probably hide spurious detections, signaling a detection event only when it happens with a given degree of reliability. For this reason, measured values were always equal to 100% when sampled below the DD threshold, and to 0% above. Instead, for ArUco, a progressive degradation for the metric was observed.

Concluding, considering the investigated configurations, only HoloLens™ can provide effective support for both long-range (ArUco Moving) and short-range (Vuforia/ArUco Stationary) scenarios. The ArUco Moving configuration is characterized by the widest ranges, but at the expense of tracking stability. Vuforia is probably the best trade-off between range and stability, whereas ArUco Stationary offers the best stability performance but in very short ranges. ARCore, instead, resulted as unsuitable for the investigated scenarios.

When marker-based tracking is not feasible, other solutions have to be considered. Regarding the comparison between the HoloLens™'s SLAM and the NOLO™'s outside-in tracking, the results highlighted that the first one achieves the best performance in terms of positional tracking. The DreamGlass™ with NOLO™, on the other hand, proved to be dramatically less accurate (especially in terms of the rotational errors of the controllers), thus resulting in an impractical solution for tracking real objects in AR scenarios.

Future development will focus on the creation of a test scenario covering all those aspects that can be investigated only through a subjective evaluation, like, e.g., presentation quality, usability, and comfort. Moreover, further devices and libraries will be included in the evaluation, together with other tracking techniques like the vision-based, marker-less ones.

References

1. Blanco-Novoa, O., Fernández-Caramés, T.M., Fraga-Lamas, P., Vilar-Montesinos, M.A.: A practical evaluation of commercial industrial augmented reality systems in an industry 4.0 shipyard. IEEE Access **6**, 8201–8218 (2018)
2. De Pace, F., Manuri, F., Sanna, A.: Augmented reality in industry 4.0. American Journal of Computer Science and Information Technology **06** (01 2018)
3. Duenser, A., Grasset, R., Billinghurst, M.: A survey of evaluation techniques used in augmented reality studies. ACM SIGGRAPH ASIA 2008 (01 2008)
4. Garrido-Jurado, S., Muñoz Salinas, R., Madrid-Cuevas, F., Medina-Carnicer, R.: Generation of fiducial marker dictionaries using mixed integer linear programming. Pattern Recognition **51** (10 2015)

5. Lang, S., Kota, M.S.S.D., Weigert, D., Behrendt, F.: Mixed reality in production and logistics: Discussing the application potentials of microsoft hololens^{TM}. Procedia Computer Science **149**, 118 – 129 (2019)
6. Liagkou, V., Salmas, D., Stylios, C.: Realizing virtual reality learning environment for industry 4.0. Procedia CIRP **79**, 712 – 717 (2019), 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 18-20 July 2018, Gulf of Naples, Italy
7. Masood, T., Egger, J.: Adopting augmented reality in the age of industrial digitalisation. Computers in Industry **115**, 103112 (2020)
8. Navab, N.: Developing killer apps for industrial augmented reality. IEEE Computer Graphics and Applications **24**(3), 16–20 (May 2004)
9. Niehorster, D., Li, L., Lappe, M.: The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. i-Perception **8** (06 2017)
10. Quandt, M., Knoke, B., Gorldt, C., Freitag, M., Thoben, K.D.: General requirements for industrial augmented reality applications. Procedia CIRP **72**, 1130 – 1135 (2018), 51st CIRP Conference on Manufacturing Systems
11. Roldán, J.J., Crespo, E., Martín-Barrio, A., Peña Tapia, E., Barrientos, A.: A training system for industry 4.0 operators in complex assemblies based on virtual reality and process mining. Robotics and Computer-Integrated Manufacturing **59**, 305 – 316 (2019)
12. Romero-Ramirez, F., Muñoz Salinas, R., Medina-Carnicer, R.: Speeded up detection of squared fiducial markers. Image and Vision Computing **76** (06 2018)
13. de Souza Cardoso, L.F., Mariano, F.C.M.Q., Zorzal, E.R.: A survey of industrial augmented reality. Computers & Industrial Engineering **139**, 106159 (2020)