Six DOF tracking system based on smartphones internal sensors for standalone mobile VR

Fredd A. Duque Farías
Abstract

Nowadays mid-range smartphones have enough computational power to run simultaneous location and mapping (SLAM) algorithms that, together with their onboard inertial sensors makes them capable of position and rotation tracking. Based on this, Google and Apple have released their own respective software development kits (SDKs) that allow smartphones to run augmented reality applications using six degrees of freedom tracking. However, this same approach could be implemented to virtual reality head-mounted-display (HMD) based on smartphones, but current virtual reality SDKs only offer rotational tracking. In this study the positional tracking technology used for augmented reality mobile applications has been implemented in a virtual reality head-mounted-display only powered by a smartphone by combining virtual and augmented reality SDKs. Compatibility issues between SDKs have been faced to develop a working prototype. An objective and controlled measurement study has been conducted that included 34,200 measurements, to test the accuracy, precision and jitter tracking of the prototype against the Oculus Rift, a dedicated virtual reality system. Results show that the developed prototype offers a decent tracking precision and accuracy in optimal conditions. It was concluded to be highly dependent on the camera view. Although, jitter presented the opposite behavior, being dependent to the device used but independent on the camera view. In its optimal conditions, user studies demonstrated that the prototype was capable of offering the same tracking performance feeling as the Oculus Rift although jitter was quite noticeable, and a common user complain. Further studies are proposed that can improve the tracking performance of the prototype by filtering jitter and using two or more cameras with a different angular to correlate feature points and obtain a wider view of the environment were the prototype is used.
Sammanfattning

ABSTRACT

Nowadays mid-range smartphones have enough computational power to run simultaneous location and mapping (SLAM) algorithms that, together with their onboard inertial sensors makes them capable of position and rotation tracking. Based on this, Google and Apple have released their own respective software development kits (SDKs) that allow smartphones to run augmented reality applications using six degrees of freedom tracking. However, this same approach could be implemented to virtual reality head-mounted-display (HMD) based on smartphones, but current virtual reality SDKs only offer rotational tracking. In this study the positional tracking technology used for augmented reality mobile applications has been implemented in a virtual reality head-mounted-display only powered by a smartphone by combining virtual and augmented reality SDKs. Compatibility issues between SDKs have been faced to develop a working prototype. An objective and controlled measurement study has been conducted that included 34,200 measurements, to test the accuracy, precision and jitter tracking of the prototype against the Oculus Rift, a dedicated virtual reality system. Results show that the developed prototype offers a decent tracking precision and accuracy in optimal conditions. It was concluded to be highly dependent on the camera view. Although, jitter presented the opposite behavior, being dependent to the device used but independent on the camera view. In its optimal conditions, user studies demonstrated that the prototype was capable of offering the same tracking performance feeling as the Oculus Rift although jitter was quite noticeable, and a common user complain. Further studies are proposed that can improve the tracking performance of the prototype by filtering jitter and using two or more cameras with a different angular to correlate feature points and obtain a wider view of the environment were the prototype is used.

KEYWORDS

Virtual reality, Positional tracking, Simultaneous location and mapping, Head-mounted-display

1 Introduction

Extended Reality, or XR, is an umbrella term to describe Augmented Reality (AR), Virtual Reality (VR) and Mixed Reality (MR). Old concepts that just recently have gained traction and are on the rise in the tech business. Many companies are investing huge amounts of money into the field to try to gain a head start in this exciting technology which “may be the next largest stepping stone in technological innovation” [6]. Each of these technologies has had different development and slightly different focus. Virtual Reality has been mainly impulsed by game industry where Facebook, HTC and PlayStation are the biggest players. Augmented Reality has been widespread in smartphones since Google and Apple launched their respective SDKs. Finally, Mixed Reality hasn't had the same attention or spread, and it has only shyly appeared in the market by the hand of Microsoft and recently Magic Leap with its Magic Leap One.

Augmented Reality, in particular, has had a huge increase in the last years, the key reason resides in the fact that all new smartphones released are AR compatible which means there is no need of extra hardware to enjoy this exciting technology. This is also really appealing for developers and companies as they always want to target the highest possible audience. Several tech entrepreneurs, included the CEO of Apple, Tim Cook, explained that AR shows the most potential, calling it a “core technology” [25]. So far, AR has applications in, among others, education, design, medicine, architecture, construction and tourism [1-5]. On the other hand, MR and VR require expensive additional hardware which doesn't make these technologies accessible for everyone and harder to widespread them. However, there are VR head-mounted-displays powered by smartphones which offer a compelling approach with an insignificant investment. The only problem resides in the fact that they only offer three Degrees Of Freedom (DOF) which means that the user cannot move from where (s)he is, as it only tracks rotations, but not translations. In other words, that system cannot register movements like crouching down or standing up, moving forward or backward. This highly reduce the natural interactions with the virtual world.

The higher the immersion the better the experience, but with only three DOF a big part of this illusion is lost as the user experiences lack of freedom of movement. Here is where this thesis takes part, state of art smartphones are equipped with enough sensors to be capable of six DOF tracking and actually AR applications for smartphones works already in six DOF, then why hasn't this technology implemented in VR HMDs powered by smartphones?
2 RESEARCH QUESTION

What are the position tracking accuracy, precision and jitter of a standalone VR HMD powered solely by onboard sensors on a smartphone using SLAM algorithms to correct the inertial data in the context of developing six DOF applications compared to the Oculus Rift, an example dedicated hardware?

3 MOTIVATION

Nowadays, VR tools are more popular than in the past and it has shown its applications beyond entertainment. VR as an immersive tool has demonstrated its capabilities to improve learning [7] and training in a wide variety of situations, e.g. training for health-care professionals [8, 9] or developing soft skills [10]; they also represent valuable, work-related tools for neuroscientists, psychologists, biologists, and other researchers as well [11]. Numerous studies have been conducted in clinical settings by using VR [12] for example, to detect and treat fears as acrophobia and arachnophobia [13, 14, 15]. Indeed, it is also a way to reduce cost and risk by training in a safe environment dangerous jobs as firefighters, skyscraper workers, etc. VR saves the need of having special installations to train them or sending them to specific locations where they can be trained. Nobel Prize winner, Edvard I. Moser talked about the use of VR [16], emphasizing its significance for research and clinical practice. Clearly shown the importance and benefits of VR for the research and non-research community. This paper aims to make this technology accessible for everyone by developing a HMD prototype with six degrees of freedom only based on a smartphone onboard sensors, overcoming the limitation of current VR headsets powered by smartphones with only three DOF which will allow a complete immersion as any VR dedicated hardware like the Oculus Rift without the need of expensive dedicated hardware or powerful computers to run them as some virtual reality systems demand [26]. Instead a modest investment in headset and a mid-range smartphone will be enough.

4 BACKGROUND AND THEORY

In this section, current tracking systems used in VR HMD will be explained. Then, recent work in the field of positional tracking systems based on by smartphones will be introduced to understand where the developed prototype of this paper will land and the technologies it uses.

4.1 Positional tracking systems: Dead Reckoning

It is important to clarify that along this document the term positional tracking is used to define six DOF tracking systems. For three DOF, rotational tracking will be used instead. Positional tracking is distinctively complicated compared to rotational as they use inertial data to estimate the pose. This technique is called dead reckoning which double integrates accelerations to calculate the variation in position, but as a consequence of it, the measurement error increases over time and could drift towards infinity within a matter of seconds [21, 22].

4.2 VR Positional tracking systems

The number of different technologies available for positional tracking in the context of HMD is too wide to explain in depth and out of the purpose of this paper. Although all VR systems are based on inertial sensors, they differentiate themselves in the way the correct these data. However, a brief description will be introduced for a better understanding of the field the developed prototype falls in. The following classification has been chosen as it clarifies the trade-off that six DOF VR HMDs present.

Figure 1: Outside-in on the left and Inside-out on the right [27].

4.2.1 Outside-in tracking: this has been the most common approach during the last years due to its simplicity compared to inside-out tracking, which is based on complex computer vision algorithms. In general terms, a trackable object (VR headset) is placed within the view of external sensors. The tracking will only work if the tracked object is within the sensors view, range and it is not occluded. IR light emitters and IR cameras are popular technologies used in most well-known VR systems as e.g. HTC Vive and Oculus Rift. This design is mainly motivated by the simplicity of the technology, which basically measures the difference in time of regular infrared emission due to difference in distance (HTC Vive) or calculates difference in position according to the change of look of a known pattern (Oculus Rift IR Constellation). Nonetheless, it has the downside that a setup is needed with a computer to do the calculations connected to the headset via cable, plus external sensors that must be placed around it to determine its position relative to the environment, making the tracking system dependent on a specific place.

4.2.2 Inside-out tracking: this second design has embedded all the sensing and computational hardware in the headset. So, there is no need of setting up sensors around the VR user. In this field, inside-out tracking approach will become essential as the market demands more mobility. Facebook Oculus Quest, HTC Vive Focus and Lenovo Mirage Solo are inside-out standalone VR HMDs just released to the market.

The inside-out design main advantage is mobility with the downside of any mobile system which is limited space. Placing all necessary tracking and computational hardware in the headsets is quite challenging. In addition, to the complexity that inside-out tracking itself introduce as they use computer vision algorithms to track the position of the headset relative to the environment. Besides, a minimum level of illumination is needed as black and white cameras are used to reduce the error of the inertial sensors by tracking features from the user surroundings.
In summary, this design offers mobility with the downside of limited computational power. On the contrary, outside-in VR systems are fixed to a particular location, but they offer higher computational power. It is a trade of high-end systems versus mobility.

5 RELATED WORK

Studies have shown that smartphones inertial sensors are capable of positional tracking in very different magnitude orders. For large distances in the order of kilometers, Global Positional System (GPS) and Inertial Measurement Unit (IMU) in smartphones have shown their capabilities to develop Intelligent Transportation System (ITS) services [17]. For medium distances in the order of meters, a robust Indoor Tracking System has also been developed and tested showing reliable results [18] [29]. Thus, for large and medium distances, depending on the specific application, errors of the magnitude of meters can be assumed, but for VR/MR applications where the tracking must be really precise and centimeters drifting can cause not only a bad user experience, but motion sickness too, makes it not viable for VR/MR applications. Motion sickness is a concerning health issue in VR, this phenomenon causes general discomfort, headache, fatigue among other symptoms [23] due to desynchronize movement of the virtual world and user’s self-movement. For this reason, a good tracking experience is needed for VR. Fortunately, with the increase of performance in smartphones making them capable of running real-time Simultaneous Localization and Mapping (SLAM) algorithms fused with additional sensors data as gyroscopes and accelerometers, highly improve accuracy by correcting the inertial data error using visual clues, called feature points which are visually distinct features in the environment. Featured points are distinctive markers on an image that an algorithm can use to track: corners of a table, high-contrast patterns in a wall, floor or in a picture, high textured-scenarios, etc [28]. On the other hand, single non-textured smooth planes would not produce enough featured points for a good tracking and might end up in drifting errors. Glossy, shiny, transparent or semitransparent objects could introduce error in the tracking system as well.

6 METHOD

A prototype has been developed to test a six DOF tracking system based on smartphones internal sensors for standalone smartphone-based VR HMD and it has been compared to dedicated hardware featuring external sensing such as the Oculus Rift. First, an objective study was carried in a controlled environment to see its capabilities as a positional tracking system; the method of operation and controlled conditions taken were inspired by a similar study by Niehorster DC et.al. that aimed to test the HTC Vive tracking performance for scientific research applications [21]. Finally, after the data from the objective study was gathered and analyzed, a subjective user evaluation was conducted following the conclusions of the first study to see how users perceive the prototype tracking compared to the Oculus Rift.

6.1 Choice of Technologies

Unity 2018.3 has been used as development platform; for positional tracking, Google ARCore 1.7 SDK has been chosen, which uses Concurrent Odometry and Mapping (COM), an algorithm that fuses SLAM and inertial data from the smartphone sensors to determine its pose [20]. For creating the stereo-vision in a smartphone display, Google Virtual Reality 1.2 (GVR) SDK has been used. Compatibility issues within these libraries had to be faced.

Google Daydream View headset and Samsung Galaxy S9 SM-G9600 have been chosen for their compatibility with the libraries mentioned above. The headset had to be modified to uncover the smartphone camera (see figure below).

Finally, Oculus Rift has been used as reference of VR positional tracking system during the experiments.

![Figure 2: Prototype headset (Google Daydream View) and smartphone (Samsung Galaxy S9).](image)

6.2 Controlled Environment for Testing

All measurements were taken in a 3.20 x 3.80 m room with a 2.70 m high ceiling. As SLAM algorithms are sensitive to change in illumination, light conditions were controlled by using 4 fluorescent lights of 60 watts with no natural light intervention. The measurements were taken in a 3 x 3 m cartesian grid with an accumulative error at the extremes of the grid and an average error at each point of the grid below 0.1 cm in both axes. The grid had vertical and perpendicular horizontal lines placed at every 25 cm, which allowed to easily marker 35 points to measure at the intersection of them.

![Figure 4: Measurement setup of the experiment. (a) Center of the tripod. (b) Tripod legs following vertical and horizontal lines. (c) Tripod holding the Oculus Rift and prototype headsets.](image)
The origin of coordinates was defined by the geometrical center of the grid, which exactly coincide with an intersection between a horizontal and a vertical line. Center marked with a $l$ in figure 3, where the axis convention used for the orientations of the measures can be seen too. Therefore, $X$ and $Z$ represented the ground plane and $Y$ the height. For the best performance of the Oculus Rift, three sensors were used mounted on tripods placed within the manufacturer recommendations in the coordinates $(1.25,1)$, $(-1.25,1)$ and $(-1.25,-1)$ at a height of 1.90 m pointing to the center of the grid and downward at an angle of 30 degrees (see fig. 3).

![Figure 3: Chart representing the grid layout used for the experiments. The location of the Measurement points and Oculus Rift sensors in the chart corresponds to the real ones.](image)

The first four experiments kept constant the height ($Y$) while the stand was moved in the $X-Z$ plane along the 35 points marked on the grid. It aimed to evaluate the tracking performance dependence on the camera view as SLAM algorithms rely on the feature points found to estimate the pose, thus four different scenarios were created to see the capabilities of ARCore to keep track of the position according to the different “scenes” presented (see fig. 5). The next 3 tests evaluated, the results independence on the device used, height estimation and shift of the origin of coordinates.

So, seven tests were conducted in total in the same order as follows:

- Test 1 empty wall scenario tested the tracking performance when there are almost no trackable features on the camera view.
- Test 2 simple scenario introduced simple trackable common objects as white A3 paper with some phrases printed on it. An improvement in performance was expected compared to Test 1.
- Test 3 living Room aimed to simulate a real case scenario where the prototype could be used. Worse tracking was expected compared to Test 2 as big reflected objects (plant pot and garbage bin) where introduced in the camera view.
- Test 4 markers evaluated how much can the tracking performance improve in “ideal conditions”. Specially designed markers were used with a high number of trackable features.
- Test 5 second device repeated the scenario with best tracking performance out of the four different first tests, but with completely different hardware, to check that the results were device independent and not due to any fault in any of the equipment. For that purpose, the prototype was replaced with another Samsung Galaxy S9. For the Oculus Rift, a new headset and three new sensors were used instead.
- Test 6 again used the scenario with best tracking performance out of the initial tests. $X-Z$ were kept constant while $Y$ was moved between two different heights. $Y$ accuracy and precision evaluation was targeted here.

Before each test, the room had to be scanned with the prototype to find trackable features. The scanning procedure took between 30 seconds to 60 seconds. Every test had around the same duration, from 70 to 80 minutes.

To ensure that every measure taken was done at a fully static state, after the stand was moved to a new location, it was left for 30 seconds before to take a new measure. Next, 1 second of data was collected providing 30 samples per device.

### 6.3 Objective study: Tests 1-7

As the innovation introduced is smartphone-based VR HMD is positional tracking, all the experiments carried focus on testing translations, no rotations are evaluated. The first four experiments kept constant the height ($Y$) while the stand was moved in the $X-Z$ plane along the 35 points marked on the grid. It aimed to evaluate the tracking performance dependence on the camera view as SLAM algorithms rely on the feature points found to estimate the pose, thus four different scenarios were created to see the capabilities of ARCore to keep track of the position according to the different “scenes” presented (see fig. 5). The next 3 tests evaluated, the results independence on the device used, height estimation and shift of the origin of coordinates.

The Oculus Rift headset and the developed prototype were mounted on a fourth tripod at a height of 1.80 m and 1.65 m respectively, both always facing the same direction. At every measure taken the center of the tripod was aligned with the measuring point and the legs with the corresponding parallel and vertical lines (see picture below), making the headsets to face the same wall with no angle variation.
• Test 7 aimed to check the drifting error of pose of the initial system of coordinates after an agitated free movement of the headsets in X-Y-Z axes.

Tests number 5, 6 and 7 used the scenario were the prototype had the best performance, since it was aimed to compare both systems and the Oculus Rift was in an optimal configuration as well (three sensors instead of two were used and they were positioned according to the manufacturer instructions for solid tracking).

Figure 6: Scenarios: (a) empty wall, (b) simple, (c) living room and (d) markers.

6.4 Data analysis

Positional tracking precision, accuracy and jitter of the prototype is evaluated against the Oculus Rift at every measurement point. For jitter, at each measure point, the size of reported sample-to-sample jumps were analyzed together with the X-Z spatial distribution of the reported data in a X-Z chart. Thus, Root Mean Square (RMS) of variation in position between samples will be used to study sample-to-sample jitter, as it displays the velocity of noise in the reported data. The higher the RMS, the higher the jitter [21].

\[ RMS_m = \sqrt{\frac{1}{n} \sum_{i=1}^{n-1} (m_i - m_{i+1})^2} \]

Where \( m \) is the number of different measure points and \( n \) the number of samples taken per measure point.

For precision and accuracy, the same X-Z chart were used. Aforesaid chart plots the average measured X-Z position for each measurement points. The average of height is also represented in a X-Z chart where height is coded in colors (heatmaps).

6.5 Subjective study: user evaluation

Based on the data from the objective study, the scenario that gave the best prototype tracking performance was used for a simple user study aimed to compare the tracking of the prototype and the Oculus in a subjective matter. Five participants, three males and two females between 21 to 32, all with experience in VR were asked to explore a Virtual Reality Room for 5 minutes with no further instructions, then do it again with a different device. The order of which device was used first, Oculus or prototype, was alternated between participants. The think aloud method was used during each test and after them a final semi-structured interview was conducted.

7 RESULTS

Data quality was a main priority, the stability and positioning of the four tripods, the accuracy of the grid, the stability of the sensors, controlled environment light conditions etc. Therefore, it was expected to obtain a similar performance of the Oculus Rift across the tests as the sensors where never moved and all the tests were performed in the same way. In the case of the prototype, variations in performance according to the scenario were expected. Finally, no tracking loss was experienced at any measurement point, tests or device.

7.1 Oculus as a Reference

Nowadays Oculus Rift is one of biggest players in VR technology competition and as a commercial product it is expected to be a good rule to measure the developed prototype capabilities. A brief view of figure 7a shows the precision and accuracy of the device. In fact, due to its high accuracy and precision all the reported measurements along the different tests are positioned almost in the exact same location, making them to occlude themselves. Regarding accuracy, the only noticeable issue is the slightly rotated grid of reported measure points. It could possibly be due to an error of calibration in the setup, yet the error is consistent across all the tests even though the system was recalibrated after each test. Also consistent across tests, heatmaps reported a slightly tilted X-Z plane with the same pattern, an increasing estimation of the height from top left bottom corner to top right corner, respect to their true physical ground plane (see fig. 7b). Interestingly the error was consistent even when a different headset and sensors were used (Appendix C part b of figures 20 to 25). (Further explorations could investigate if the reported rotational angles are related).

Finally, looking at figure 7c it is noticed that the lowest error was obtained in focused around the center of the grid where the distance from the three sensors to the headset is about the same. Closely in focused around the center of the grid where the distance from the three sensors to the headset is about the same. Curiously, the highest error was reported at point 30 in the top right corner instead of the bottom left corner. This could mean that trackable object was positioned too close to the sensors at the extremes of the grid.

Jitter was stable across the grid and tests with the exception when it was situated close the right corner of the grid. The farthest positions from the two left sensors, position where only one sensor
was close to the headset. Jitter was especially stable in Y axes compared to X and Z axis, both presented a slightly higher average error specially in Test 4, where interestingly measurement point 18 in X and Z coordinates presents a jitter 30 and 25 times higher than the average error in the grid respectively. However, this huge increase in jitter is not reflected in the reported average measurement error in the corresponding point, as it can be seen in figure 23 of the Appendix C where point 18 has an error in distance of 1.5 cm, below the overall average across Tests 1 to 5, 1.69 cm. It could be thought that it was due to an error in the measurement procedure, however the prototype jitter was between the normal values for the same measuring points.

7.2 Prototype X-Z plane: Tests 1 to 5

Although the room scanning was performed without any problem despite the lack of features in Test 1 empty wall, it can be clearly seen in the figure 8a the significant offset presented by the measurements. The grid formed by the reported measurements is significantly rotated clockwise, where specially points 22 to 27 had a big misestimation of the X coordinate. The reported estimation of each measuring point was closer to the left measuring point than to the proper one, causing confusion at first glance. Black squares in figure 15c represent out of scale values being three to four times the average Euclidean distance error of the Test.

For the second test, the introduction of simple high-contrasts trackable features as letters was expected to improve the tracking and as figure 10a shows a significant improvement of tracking was experienced, especially in bottom left part of the chart that presented the highest error in Test 1, but an overall improvement in X-Z plane is easily visible, performing quite similarly to the Oculus Rift. Highlight the Z tracking, that obtain an average Euclidean distance error lower that the Oculus Rift.

Test 3 living room added reflective surfaces to the camera view that worsen the performance considerably. The positional tracking average error was similar to the Test 1 empty wall, but double of the Test 2 simple.

As expected Test 4 presented a good performance. It had the best Y performance across all the Tests carried and the second-best X tracking, only outperformance by the Test 5 second device which was also using markers scenario. Surprisingly, the overall distance error of Test 2 was slightly lower mainly due to impressive Z performance. Looking closely to the figure 10b, the first two rows of the grid which are closer to the wall, have a better tracking estimation than the last two, farther away from the markers. This was suspected to be caused by the size of the marker’s features, too small to be recognized by the camera when it is positioned far away.

Test 5 aimed to check that the results obtained in the prototype and Oculus were independent of any particular defect on the hardware used, by repeating the test with the best tracking which was expected to be Test 4 markers but replacing all the hardware. Aiming to test the lower performance of Test 4 than Test 2 due to the markers’ features size too. Markers scenario was used but increasing its feature size a 25%. This time X-Z chart presented a better performance in the last rows of the grid but still not as good as the first ones. Nonetheless, the average distance error was lower than the Test 2 simple, having the best performance out of all the tests carried.

7.3 Prototype jitter

Jitter didn’t show any particular pattern in X, Y or Z axis. No clear relation could be found between the different scenarios and

![Figure 7: (a) Oculus Rift tracking error in X-Z plane, (b) average height error and (c) average Euclidean distance error along Tests 1 to 5.](image-url)
Figure 8: (a) Test 1 empty wall prototype tracking error in X-Z plane in comparing to the Oculus average tracking error along Tests 1 to 5. (b) Prototype height tracking error and (c) prototype Euclidean distance tracking error. Black squares represent out of scale values.

Figure 9: (a) Test 1 living room prototype tracking error in X-Z plane in comparing to the Oculus average tracking error along Tests 1 to 5. (b) Prototype height tracking error and (c) prototype Euclidean distance tracking error. Black squares represent out of scale values.

Figure 10: (a) Test 2 simple prototype tracking error in X-Z plane, (b) Test markers tracking error in X-Z plane and (c) Test 5 second device tracking error in X-Z plane.
the obtained performance. However, in the scenario empty wall, simple and living room more stable data was obtained when the prototype was closer to the front wall, but this could also be a coincidence, more tests are needed to clearly state a conclusion. On the contrary to the similar performance in terms of positional tracking error in Test 4 and Test 5, jitter had a completely different behavior in the same tests. Heatmaps in figure 11 show that the second Samsung Galaxy S9 had almost 3 times higher jitter along the whole grid.

No probable cause was found for the prototype and the Oculus out of scale RMS values (from 2 to 30 times higher than the average error of the corresponding grid). These values were reported in different experiments and axis with no relation (black squares in appendices D and E). It could be considered a measurement error if both devices, Oculus and prototype presented the out of range values in the same measuring point, maybe due to the tripod was still not stabilized when the samples were taken, but only one device at the time reported these out of range values.

7.4 Test 6: Height tracking
All the measures possible were taken to only move the prototype in the Y axes and don’t affect X and Z coordinates. The tripod legs positions were checked after each new measurement ensuring no changes in X-Z position. However, the Oculus and the prototype presented an offset in the X and Z coordinates. Below 1 cm in both axis for the Oculus. The prototype presented an offset of 2,15 cm and 1,45cm in the X and Z coordinates respectively. Surprisingly, the prototype X coordinate estimation almost double the error of Y coordinate.

7.5 Test 7: Free movement
Agitated free movement of the tripod produced a lower shift of the origin of coordinates than expected for the two tracking systems. Even though the tripod holding both devices was shaken and moved around fast and abruptly for 90 seconds the reported shift of the origin of coordinates was below one centime in both cases.

7.6 Axis tracking differences
Looking at table 1, the prototype Y and Z pose estimation were similar along all tests. However, the reported X measures presented the highest error at each of the tests and at Test 1 and 3 doubles the Y and Z average Euclidean distance error.

7.7 Subjective study: user evaluation
All participants expressed confusion about what to do. Wandering around looking for differences was the general response. This led to move around the whole room, looking at the walls, ceiling, etc. Encouraging the participant to pay attention to any detail while moving along the whole room several times. During the tests all the participants clearly stated complains about the jitter in the prototype, but no complaints about the tracking accuracy or precision were mentioned. One of the participants, an AR/VR developer, experimented with the tracking. The participant experienced a loss of tracking after s(he) moved the head from left to right at an unnatural velocity, the participant reported that when s(he) stopped the “world” was not in the right position but in a fraction of second it moved again to the right coordinates. In the interviews, participants were asked about tracking accuracy, precision and jitter. No relevant comments for the purpose of the thesis were extracted.

8 DISCUSSION
The prototype has shown that it is capable of six DOF tracking, although it cannot be compared to dedicated hardware as the Oculus Rift in terms of accuracy and precision focusing, especially when the scenario is not optimal for the camera view. On the one side, lack of features points or the inclusion of reflective surfaces have shown to reduce the tracking performance in the same magnitude. On the other, specific designed markers or simple high-contrast patterns as plain text printed on paper improve the track-

![Figure 11: (a) Test 4 prototype jitter and (b) Test 5 prototype jitter, both using the scenario markers](image-url)
ing performance in the same degree. This clearly stated the tracking performance dependence of the camera view, but independent on the device used as Test 5 demonstrated. However, the jitter presented exactly the opposite behavior, being dependent on the smartphone used, but independent on the scenario used. Considering that jitter varied substantially even though in Test 4 and Test 5, the prototype was powered by two equal smartphones (two Samsung Galaxy S9), this could imply that jitter might vary even more between models and brands.

Tracking performance in the best-case scenario where markers or simple high-contrast patterns were used, although still not close to the 1.7 cm average distance error reported by the Oculus, the results obtained are quite promising being the average distance error around 3 cm, which is difficulty notable for the VR world, as it was concluded from the subjective studies. Per contra, jitter couldn’t be ignored, being the only complain about the prototype by the participants in said study, but nevertheless a simple filter of the reported coordinates could completely solve this problem. A straightforward self-experiment was carried to proof this argument. Two simple VR applications were created. The first one only using GVR SDK which can deliver three DOF by using the accelerometer and gyroscope as ARCore. The second one, was based on the prototype app, but this time the reported translations were ignored and only rotations were parsed to the virtual camera, in other words limiting its tracking only to rotations. The same virtual reality room used along the whole research was used in these experiments as well. A quick self-evaluation made clear that no jitter was experienced at all in the app developed solely with GVR SDK, but it was clearly noticeable in the app using ARCore SDK. Notice that the same hardware was used for both experiments.

Mobility is the reason of being of inside-out VR tracking systems and if any modification of the environment is needed to obtain a stable and acceptable tracking, the prototype will lose part of its attractive and even though that in the worst-case scenario the overall tracking average error was 7 cm that wouldn’t represent a major issue, but the fact that in punctual moments the error went up to 20 cm, makes this system inconsistent. However, it could be argued that ARCore is a released SDK by Google and many commercial apps are based on this technology, therefore it grants quality tracking enough for a good user experience, but it has to be taken into account that 20cm shift in coordinates in a smartphone screen doesn’t produce the same experience in the user when it is held in the hands or when it is maximized by lenses, centimeters away from the eyes. Loss of tracking or unexpected shift of coordinates could prejudice the user producing motion sickness or loss of balance.

On the other hand, nowadays smartphones include two or more cameras with different angular. This could definitely help to improve the tracking. Wide angle cameras will be able to give a broader view of the environment which means that big reflective objects as the bin and plant pot in the Test 3 living room will introduce less error compared to regular lens as these objects will be a small part of the camera view and not what it will be mainly be seeing, helping to keep stable tracking all the time. Then a second camera with different angular could be used to correlate points with the first camera to improve the precision and accuracy.

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.91</td>
<td>2.12</td>
<td>2.69</td>
<td>7.31</td>
</tr>
<tr>
<td>2.31</td>
<td>1.54</td>
<td>0.63</td>
<td>3.14</td>
</tr>
<tr>
<td>4.87</td>
<td>2.76</td>
<td>2.17</td>
<td>6.86</td>
</tr>
<tr>
<td>1.92</td>
<td>1.44</td>
<td>1.68</td>
<td>3.31</td>
</tr>
<tr>
<td>1.80</td>
<td>1.62</td>
<td>1.28</td>
<td>3.02</td>
</tr>
<tr>
<td>0.55</td>
<td>0.84</td>
<td>0.80</td>
<td>1.69</td>
</tr>
</tbody>
</table>

Table 1: Summary of average positional tracking error in each axes and Euclidean distance.

<table>
<thead>
<tr>
<th>RMS</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.06</td>
<td>0.99</td>
<td>1.35</td>
<td>1.43</td>
<td></td>
</tr>
<tr>
<td>1.18</td>
<td>1.05</td>
<td>1.37</td>
<td>2.27</td>
<td></td>
</tr>
<tr>
<td>2.06</td>
<td>1.45</td>
<td>1.70</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td>1.68</td>
<td>1.28</td>
<td>0.98</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>1.52</td>
<td>1.37</td>
<td>4.44</td>
<td>3.55</td>
<td></td>
</tr>
<tr>
<td>0.44</td>
<td>0.45</td>
<td>0.45</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Summary of root mean square in each axes and Euclidean distance.

Another aspect to take into consideration that affects the tracking performance is the video/image stabilization. Human movements are not steady or smooth, which produces jitter. Such jitter will affect the inertial sensors but also the camera view, given a blurry image as a result. As it was argued inertial sensors data could be corrected by implementing filters but fixing blurry images in real time is a challenge. However, smartphones manufactures have achieved more stable images or videos by either software, hardware or combined solutions. Software solutions called Electronic Image Stabilization (EIS) are in general lines post-processing techniques, which makes this unsuitable for VR where the information gather by the camera must be sent in real time to the algorithm in charge to correct the inertial data. Nevertheless, hardware solutions named Optical Image Stabilization (OIS) use floating lens and gyroscopes to stabilize the image projected on the camera sensor even before this is converted into digital information. This would reduce the error on the digital image that the SLAM algorithm would have to deal with, although still it cannot correct large jitter. The smartphone used in these studies did not feature OIS, but many common smartphones in the market nowadays do, such as Google Pixel 3, Huawei P20 Pro, iPhone XS, etc. being no longer a top range feature.
9 FUTURE WORK

Apart from all the possibilities that dual or triple cameras with different angular could introduce, this paper has focused on the tracking capabilities of VR systems solely based on smartphones sensors, but no controllers were developed or studied. In this field the same problem is encountered. Smartphone-based VR HMD either come without controller or with controllers that only support rotational tracking. This had led to develop interesting new ways to interact with the systems, as gaze [24]. One of the reasons why the Google Daydream View headset was chosen in this research was that it comes with a three DOF wireless controller. This was done having in mind the possibility of future studies of the interactive capabilities of the prototype, as the headset features positional tracking, but the controller only offers rotational tracking. Furthermore, the possibility of developing a positional tracking controller could be study as well, the Google Daydream View controller features a gyroscope and an accelerometer, so dead reckoning technique could be applied, although a way to correct the error should be researched.

Filters for the jitter could be also studied and implemented to the current prototype. The reported coordinates given by ARCore could be filtered by code using i.e., the Kalman filter [22].

10 CONCLUSION

This paper exposed smartphones capabilities to track position thanks to inertial data corrected by SLAM algorithms in the context of VR HMD only powered by smartphones. Objective studies have demonstrated its high dependency on the camera view directly affecting the precision and accuracy of the device tracking, which makes the tracking system inconsistent and non-reliable, specially where there are large flat surfaces without any pattern or large reflective objects. In general lines across all the tests, the prototype presents a modest accuracy, but low precision (Appendix A figure 12). Jitter was found to be noticeable and independent of the camera view, but highly dependent of the device used.

Measures can be taken to improve the tracking performance to a quite promising level, where the tracking is comparable to the Oculus Rift, by removing from the camera view said reflective objects and adding markers. But this act, would conflict with the concept itself of standalone inside-out tracking VR HMD, which basically stands out for not needing any kind of modification in the environment where it is meant to be used, neither a setup.

However, despite all, with a simple filtering of the jitter and small measures to be considered for an optimal tracking, it could perform similarly to the Oculus Rift as the user evaluation has pointed out that no difference was found in both tracking systems, even though the average distance error in optimal condition for the Oculus was 1.7 cm and slightly above 3 cm for the prototype.

REFERENCES


[27] Oculus Ready PCs: are high-performance computer systems that are optimized for the Oculus Rift https://www.oculus.com/rift/oculus-ready-pcs/


APPENDIX A - Overview positional tracking error

Figure 12: X-Z plane showing the tracking performance of the prototype from Test 1 to Test 5 in comparison with the Oculus average tracking performance along the same tests.
Figure 13: Height tracking error of the prototype from Test 1 to Test 5 in comparison with the Oculus average height tracking performance along the same tests.

Figure 14: Euclidean distance tracking error of the prototype from Test 1 to Test 5 in comparison with the Oculus average Euclidian distance error along the same tests.
APPENDIX B - Prototype positional tracking error

Figure 15: Test 1 empty wall tracking performance in empty wall scenario. Black squares represent out of scale values.

Figure 16: Test 2 simple tracking performance in simple scenario.

Figure 17: Test 3 living room tracking performance in living room scenario. Black squares represent out of scale values.
Figure 18: Test 4 markers tracking performance in markers scenario.

Figure 19: Test 5 second device tracking performance in markers scenario.
APPENDIX C - Oculus Rift positional tracking error

Figure 20: Test 1 empty wall tracking performance.

Figure 21: Test 2 simple tracking performance.

Figure 22: Test 3 living room tracking performance.
Figure 23: Test 4 markers tracking performance.

Figure 24: Test 5 second device tracking performance.
APPENDIX D - Prototype jitter

Figure 25: Test 1 empty wall, jitter performance measured by RMS.

Figure 26: Test 1 simple, jitter performance measured by RMS.

Figure 27: Test 3 living room, jitter performance measured by RMS.
Figure 28: Test 4 markers, jitter performance measured by RMS.

Figure 29: Test 5 second device, jitter performance measured by RMS. Black squares represent out of scale values.
APPENDIX E – Oculus Rift jitter

Figure 25: Test 1 empty wall, jitter performance measured by RMS. Black squares represent out of scale values.

Figure 26: Test 2 simple, jitter performance measured by RMS.

Figure 27: Test 3 living room, jitter performance measured by RMS.
Figure 28: Test 4 markers, jitter performance measured by RMS. Black squares represent out of scale values.

Figure 29: Test 5 second device, jitter performance measured by RMS.