2016 IEEE International Symposium on Mixed and Augmented Reality Adjunct Proceedings

Evaluation of Motion Tracking and Depth Sensing Accuracy of the Tango Tablet

Rafael Roberto*
Voxar Labs, CIn-UFPE, Brazil

João Paulo Lima†
DEINFO-UFRPE, Brazil
Voxar Labs, CIn-UFPE, Brazil

Thúlio Araújo‡
Voxar Labs, CIn-UFPE, Brazil

Veronica Teichrieb§
Voxar Labs, CIn-UFPE, Brazil

Abstract

This paper presents an evaluation of the Tango tablet regarding its motion tracking and depth perception capabilities. A methodology for performing such kind of evaluation is proposed. Motion tracking error is assessed in both small workspace and large environments. In the small workspace scenario, the distances reported by the motion tracking system are compared with values measured with the aid of a graph paper. In the large environment condition, the tracking error consists in the difference between initial and final positions given by the system when the device moves around the environment and returns to the same location. Depth sensing precision is evaluated by comparing the 3D coordinates reported by the system of the inner corners of a chessboard pattern with ground truth values obtained from color camera image processing. The results show that Tango tablet sometimes presents large motion tracking errors, which may harm AR experience. In addition, Tango tablet depth sensing presents average error values similar to desktop depth cameras, but it is more sensitive to infrared reflection properties of the objects to be mapped.

Index Terms: I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Depth Cues, Range Data, Tracking; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities

1 Introduction

Google Tango [2] is a technology platform that allows mobile devices to detect their position and navigate the physical world similar to a human walking around indoor and outdoor environments. Tango uses advanced computer vision, image processing, and special vision sensors to give the device the ability to have a spatial perception of itself.

The first device featuring Google Tango was an Android platform tablet, which was released in June 2014. Since then, several applications have been developed using its motion tracking and depth sensing technologies. For instance, [13] presents a system that uses Tango to previsualize virtual objects in on-set film production in order to help actors interact with them. Another one uses the tablet to reconstruct and locate the user in the environment to provide a redirected walking [6].

Since there are thousands of this device in use, it is natural that more applications will appear using its motion tracking and depth sensing technologies. Moreover, there is a commercial device launched with these technologies by Lenovo in partnership with Google [10], which tends to stimulate the development of even more programs featuring these capabilities. Therefore, it is relevant to evaluate its accuracy. As far as the authors know, so far no study evaluates the Tango tablet. This experiment is important because it provides expertise on how to assess tracking systems and also to find scenarios and situations in which they work well and fail.

Evaluating tracking systems is a challenging task [15]. Many efforts have been made in the past years to provide metrics and standards to analyze the aspects related to this problem [8, 9]. The main reason is the difficulty to find the ground truth to compare with the obtained results. There are several benchmark datasets available aiming computer vision tracking systems evaluation, each one having many purposes and providing several types of input data. For instance, [5] presents an image dataset to evaluate planar model based techniques, [11] describes a benchmark to measure the quality of 3D model based algorithms and [14] provides RGB and depth information aiming SLAM systems. All of them also provide the expected results, which are used to assert the precision of the algorithm. However, it is hard to use these datasets on mobile devices. One reason is the difficulty to extract information from different sensors since they are noisy and the precision varies among devices.

This work aims to evaluate the precision of both motion tracking and depth sensing technologies on Google Tango tablet. An evaluation methodology is thus proposed and employed for fulfilling this goal. This study also discusses the obtained results.

2 Tango Tablet

As mentioned, the Google Tango device has some novel features that provide new ways to navigate in different environments using technologies such as motion tracking and depth perception [2]. This section explains how these technologies work on Tango tablet and describes the characteristics of the sensors used by them.

Motion tracking means that a Tango device can track its own movement and orientation through 3D space. Tango implements motion tracking using visual-inertial odometry to estimate where a device is relative to where it started. Standard visual odometry uses camera images to determine a change in position by looking at the relative position of different features in those images. Visual-inertial supplements visual odometry with inertial motion sensors capable of tracking a device's rotation and acceleration. This allows a Tango device to estimate both its orientation and movement within a 3D space with even greater accuracy. Unlike GPS, motion tracking using visual-inertial odometry works indoors. In addition to the gyroscope and accelerometers, Tango uses a wide-angle motion tracking camera (also known as the "fisheye" lens) to add visual information, which helps to estimate rotation and linear acceleration more accurately. To perform motion tracking, the Tango APIs provide the position and orientation of the user's device in full six degrees of freedom. The data is returned with two main parts: a vector in meters for translation and a quaternion for rotation.

Depth Perception gives an application the ability to understand the distance to objects in the real world [2]. Tango tablet implements depth perception with time-of-flight (ToF) technology, which requires the use of an infrared projector and an infrared sensor.

Tango tablet depth sensor is designed to work best indoors at moderate distances (0.5 to 4 meters). This configuration gives good depth at a distance while balancing power requirements for infrared illumination and depth processing. It may not be ideal for close-

*e-mail: rar3@cin.ufpe.br
†e-mail: joao.mlima@ufrpe.br, jpsml@cin.ufpe.br
‡e-mail: tsla@cin.ufpe.br
§e-mail: vt@cin.ufpe.br

978-1-5090-3740-7/16 $31.00 © 2016 IEEE
DOI 10.1109/ISMAR-Adjunct.2016.73

range object scanning or gesture detection.

The Tango APIs provide a function to get depth data in the form of a point cloud. This format gives (x,y,z) coordinates for as many points in the scene as are possible to calculate. Each dimension is a floating point value recording the position of each point in meters in the coordinate frame of the depth-sensing camera.

3 Evaluation Methodology

The following subsections describe the methodology used to evaluate the motion tracking and depth sensing functionalities of the Tango tablet.

3.1 Motion Tracking

The chosen evaluation method is based on moving the Google Tango tablet between two known positions in the real world and comparing the distance between these positions computed by the device with the ground truth value. This way, it is possible to evaluate the error the system accumulates during motion tracking from a starting point to an ending position. It was used a graph paper with the precision of one millimeter to ensure the experiment accuracy. The paper was pasted in a table so it does not move during the tests. A needle was attached to the base of the Tango tablet in order to have the exact position of the device over the paper. Figure 1 shows the setup. It was designed two different experiments based on this setup, one to evaluate a small AR workspace and another one for large environments.

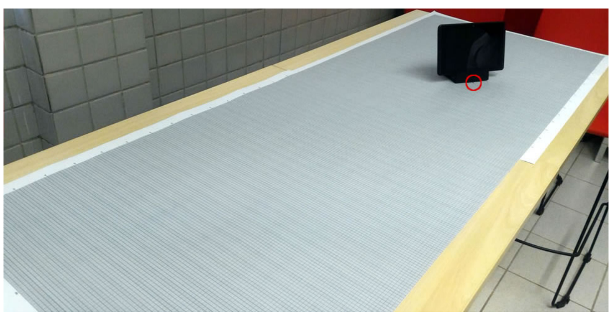

Figure 1: Evaluation setup consists of a graph paper with precision of one millimeter and measuring 1.5 x 0.55 meters. Red circle highlights the needle used to get the exact position on the paper.

For the first one, the idea is to evaluate how the Tango tablet works on a small workspace, which for this study is a table with an area up to one square meter. Therefore, the device is positioned with the needle on the origin of the graph paper and their axis are aligned. The tablet will be moved freely and placed on any other position on the graph paper. The error is the difference of the Euclidean distances between the origin and the final position computed on the Tango device and measured with the graph paper.

Regarding the large environment, the goal is to measure how the Tango device behaves when performing motion tracking on places such as a regular office or outdoors. The office is a closed room that has an approximate area of 50 square meters and artificial illumination. As for the outdoor experiment, it was placed in two different courtyards measuring around 100 square meters each. It was also located in the corridor of a building that is open to the outside. All the outdoor measurements were collected using natural illumination during daylight.

It is not possible to have a graph paper that is large enough to cover the entire office or the courtyards. Thus, for this experiment, the device is also positioned with the needle on the origin of the graph paper, their axis are aligned and it is moved freely in these environments. The difference from the previous experiment is that the tablet is returned to the same position where it started. The error is also the Euclidean distance between the position computed on the Tango device after finishing the movement and the initial position.

3.2 Depth Sensing

Regarding the depth sensor, it was evaluated the accuracy of the 3D points positions obtained from it. The process consists in calculating the Euclidean distance between 3D points reconstructed from the color camera and corresponding ones from the depth camera. The registration between color and depth cameras is provided by the Tango API. The intrinsic parameters used are the ones from the manufacturer calibration. The 3D points of the color camera are the inner corners of a detected chessboard pattern whose pose is estimated using the Direct Linear Transformation (DLT) method and refined by minimization of reprojection error [3].

Since the depth image generated by the Tango device has a lower resolution when compared to the color image, it has to be upsampled when obtaining the corresponding depth measure of a chessboard corner. Both nearest-neighbor and bilateral interpolation [16] were evaluated for performing this task.

4 Results

In order to evaluate the motion tracking capability available on Google Tango, it was used a sample application available on the project GitHub[1] called "C++ Augmented Reality Example". This application uses the fisheye camera and the device gyroscope and accelerometer to compute its pose relative to its initial position.

4.1 Motion Tracking

For the experiment on the small workspace, the Tango tablet was moved freely to any other position over the graph paper. To have statistical power, the sample size for this experiment was calculated aiming 95% confidence within 1 centimeter precision [4]. Therefore, these measurements were repeated 67 times to ensure that.

Figure 2 shows the error dispersion, in which the smallest was 0.0087 meters and the largest was 0.1808 meters. On average, the error was 0.0666 ± 0.0399 meters.

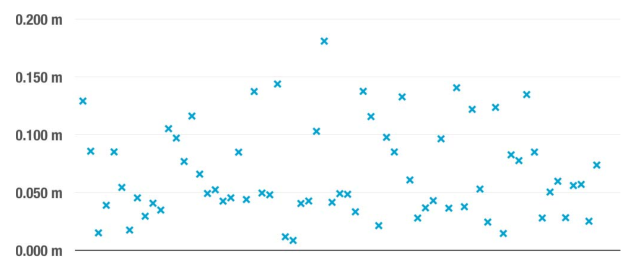

Figure 2: Error dispersion for the small workspace experiment.

Figure 3 shows the device's position distribution over the graph paper during the experiment.

Regarding the evaluation of the motion tracking on large environments, the error is the Euclidean distance between the initial and the final position calculated by the device after moving it freely in this environment and returning to the same location. After a few measurements, it was noted a large difference between the errors from the indoor and outdoor environments. Therefore, it was decided to perform two different evaluations. The number of samples to ensure statistical power emphasize this decision. For the indoor experiments, it was necessary to have 31 samples to have 95% confidence within 2 centimeters precision, which is the double of the small workspace because the covered area was much larger. On the other hand, it was not possible to have such confidence in the outdoor experience. The reason is that error variation is so high that it

[1] https://goo.gl/DYwBPM

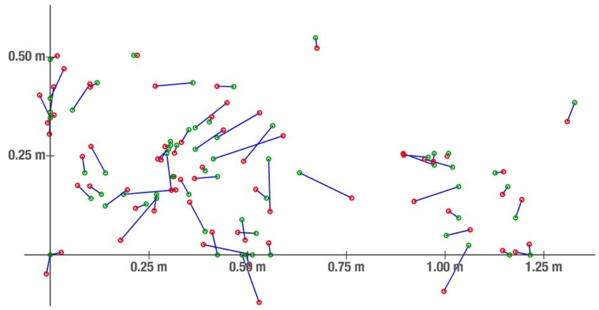

Figure 3: Distribution of the device positions on the graph paper (green) and their correspondent positions calculated by the Tango tablet (red).

would be necessary to have more than 5000 samples to have 95% of confidence within 2 centimeters precision.

Figure 4 shows the error dispersion in the large indoor scenario. The smallest one was 0.0487 meters while the largest was 0.2607 meters. On average, the error of the 45 samples measured was 0.1421 ± 0.0563 meters. Also, the average distance walked with the Tango tablet was 23.6082 ± 7.8920 meters.

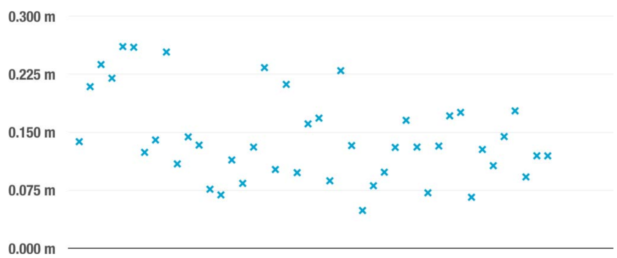

Figure 4: Error dispersion for the large indoor environment experiment.

Figure 5 illustrates one of the paths walked with the Tango tablet and the difference between the initial and final positions.

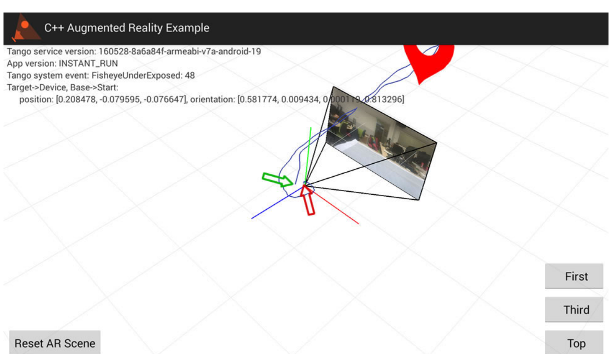

Figure 5: Screenshot of one of the paths computed using the Tango tablet. The green arrow points to the initial place and the red one to the final position calculated after a free walk. The error is the average Euclidean distance between them.

Regarding large outdoor environments, it was performed 21 repetitions. However, this amount was not enough to have statistical power. Even though, the average error of 0.9052 ± 0.7526 meters indicates that precision of the Tango tablet is much smaller when it is dealing with natural illumination and wide spaces.

4.2 Depth Sensing

In the first experiments, the chessboard pattern was printed on a paper using black ink. However, the dark squares did not reflect infrared light in a way that would allow robustly estimating the depth of the chessboard corners. Due to this, it was used a mix of cyan, magenta and yellow ink in order to have dark squares that are correctly scanned by the depth camera. This aspect is illustrated in Figure 6. The chessboard was printed on A4 paper with a square side of 28 millimeters.

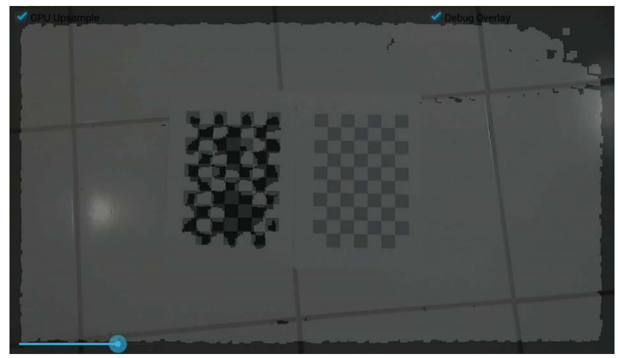

Figure 6: Screenshot of the depth estimation of two chessboards printed on a paper. On the right, the one printed with a mix of cyan, magenta and yellow inks. On the left, the same pattern printed with a black ink. Note that the sensor is not able to estimate depth on the black squares of the left paper.

Figure 7 shows the mean depth estimation error considering different distances between the device and the chessboard pattern and different depth interpolation strategies. In order to obtain error values accurate within 0.1 millimeters at 95% confidence, 150 samples were collected for each configuration. The average execution times of the nearest-neighbor and bilateral depth interpolation procedures for each point were 0.6956 ± 0.1275 and 1.8811 ± 0.3118 milliseconds, respectively.

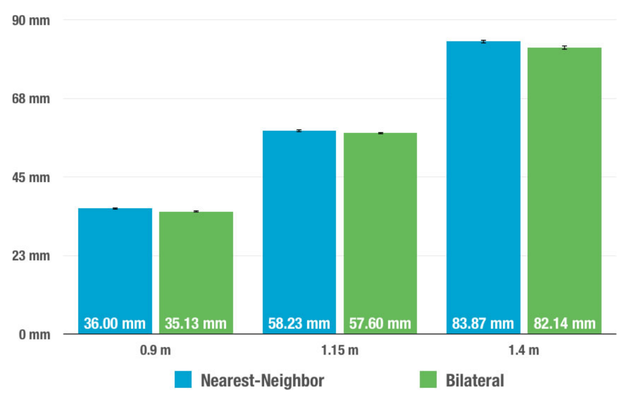

Figure 7: Mean depth estimation error with respect to distance between Tango and chessboard pattern using different depth interpolation methods.

5 Discussion

The results showed that the motion tracking of the Tango tablet is 2.3 times more precise on a small workspace than on large indoor environments. However, the presented errors can have an impact on the user experience. An average error of 6 centimeters is often noticed in a small workspace.

On the other hand, even having a bigger error when dealing with large environments, the precision of the motion tracking on indoor spaces is suitable to provide a good user experience for several kinds of AR applications. However, for scenarios in which it is necessary to have accuracy, this error can harm user experience. Figure 8 (left) shows an example where Tango tablet measure tool is calculating the width of a 0.7 meter door. After moving the device for a few steps away from the door and back, the ruler is placed in a different position, as seen in the right side.

Figure 8: Door width estimation using Tango tablet. Left side shows the initial measurement and the right side shows the ruler position after moving the device. Door actual width is 0.7 meters.

During the experiment, it was noticed that the algorithm that Tango uses for the motion tracking seems to mistrust the sensors it uses regarding their precision. There is an indication that it has a much stronger confidence on the information provided by the fisheye camera. For instance, sometimes the device was left standing still over the table and when some object moves in front of the camera, the motion tracking algorithm calculates that the tablet was moving in the opposite direction.

The tests on large outdoor environments were not enough to provide results with statistical power because there was a significant variation on the error measured on every sample. However, this disparity suggests that the Tango tablet has some issues to deal with outdoor illumination and wide spaces. It is emphasized by the fact that there were not a single result on the outdoor measurements that presented an error smaller than any indoor sample. Moreover, in some cases, the error was greater than 1.0 meter and in the worst case it reached more than 3.0 meters.

Regarding depth sensing, the mean errors presented by Tango tablet were similar to the ones obtained with a desktop depth sensor. The results also suggest that the depth estimation error increases linearly with respect to the distance between the device and the object. Bilateral depth interpolation provided an average precision improvement of 1.82% (1.08 millimeters) with respect to the nearest-neighbor approach, but it was more than 2.5 times (1.18 milliseconds) slower on average for each point.

6 Conclusion and Future Work

It was presented an evaluation of motion tracking and depth sensing functionalities of the Tango tablet. A graph paper and a chessboard pattern were used as tools for assessing motion tracking and depth estimation errors, respectively. Tango tablet motion tracking errors were around 6 and 14 centimeters for small and large indoor scenarios, respectively, which may not be suitable for some AR applications. In addition, while Tango tablet depth sensing presented average errors compatible with the ones obtained with desktop depth cameras (ranging from around 3.5 to 8 centimeters for distances between 0.9 and 1.4 meters), it presented some issues when scanning objects with low infrared reflectivity.

As future work, a motion capture system, such as the one from ART [1], may be used to generate ground truth for evaluating Tango 3D trajectories. Improvements for Tango motion tracking may also be proposed, such as also exploiting depth information for odometry estimation [7]. Regarding the depth sensor problems when dealing with objects that do not reflect infrared light well, depth map hole-filling methods that take color information into account such as [12] may be investigated.

Acknowledgements

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (processes 456332/2013-8, 456800/2014-0 and 140898/2014-0) for partially funding this research.

References

[1] Advanced Realtime Tracking GmbH. Art advanced realtime tracking. http://www.ar-tracking.com/, June 2016. [Online; last access: 22-Jun-2016].

[2] Google Inc. Atap project tango google. http://goo.gl/xQIkob, Feb 2014. [Online; last access: 09-July-2016].

[3] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[4] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. Wiley, 1991.

[5] S. Lieberknecht, S. Benhimane, P. Meier, and N. Navab. A dataset and evaluation methodology for template-based tracking algorithms. In *Mixed and Augmented Reality (ISMAR). 8th IEEE International Symposium on*, pages 145–151, Oct 2009.

[6] T. Nescher, M. Zanky, and A. Kunzz. Simultaneous mapping and redirected walking for ad hoc free walking in virtual environments. In *IEEE Virtual Reality (VR)*, March 2016.

[7] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and Augmented Reality (ISMAR), 10th IEEE International Symposium on*, pages 127–136, Oct 2011.

[8] A. Petit, G. Caron, H. Uchiyama, and E. Marchand. Evaluation of Model based Tracking with TrakMark Dataset. In *2nd Int. Workshop on AR/MR Registration, Tracking and Benchmarking*, Basel, Switzerland, Switzerland, 2011.

[9] A. Roy, X. Zhang, N. Wolleb, C. P. Quintero, and M. Jägersand. Tracking benchmark and evaluation for manipulation tasks. In *Robotics and Automation (ICRA), IEEE International Conference on*, pages 2448–2453, May 2015.

[10] D. Seifert. The first phone with google tango is a 6.4-inch monster from lenovo — the verge. http://goo.gl/ZAJXzi, June 2016. [Online; last access: 19-June-2016].

[11] F. Shibata, S. Ikeda, T. Kurata, and H. Uchiyama. An intermediate report of trakmark wg international voluntary activities on establishing benchmark test schemes for ar/mr geometric registration and tracking methods. In *Mixed and Augmented Reality (ISMAR), 9th IEEE International Symposium on*, pages 298–302, Oct 2010.

[12] W. Song, A. V. Le, S. Yun, S.-W. Jung, and C. S. Won. Depth completion for kinect v2 sensor. *Multimedia Tools and Applications*, pages 1–24, 2016.

[13] A. Stamm, P. Teall, and G. B. Benedicto. Augmented virtuality in real time for pre-visualization in film. In *IEEE Symposium on 3D User Interfaces (3DUI)*, pages 183–186, March 2016.

[14] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d slam systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[15] H. Tamura and H. Kato. Proposal of international voluntary activities on establishing benchmark test schemes for ar/mr geometric registration and tracking methods. In *Mixed and Augmented Reality (ISMAR). 8th IEEE International Symposium on*, pages 233–236, Oct 2009.

[16] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Computer Vision. Sixth International Conference on*, pages 839–846, Jan 1998.