

FIRST EXPERIMENTS WITH THE TANGO TABLET FOR INDOOR SCANNING

Abdoulaye A. Diakité^{a,*}, Sisi Zlatanova^a

^a 3D GeoInformation, Department of Urbanism, Delft University of Technology,
Jaffalaan 9, 2628 BX Delft, The Netherlands - (a.a.diakite, s.zlatanova)@tudelft.nl

Commission VI, WG VI/7

KEY WORDS: Project Tango, 3D scanning, Indoor Building Modeling, 3D Indoor Navigation

ABSTRACT:

During the last two decades, the third dimension took an important place in the heart of every multimedia. While the 3D technologies mainly used to be tools and subject for researchers, they are becoming commercially available to large public. To make it even more accessible, the Project Tango, led by Google, integrates in a simple Android tablet sensors that are able to perform acquisition of the 3D information of a real life scene. This makes it possible for a large number of applications to have access to it, ranging from gaming to indoor navigation, including virtual and augmented reality. In this paper we investigate the ability of the Tango tablet to perform the acquisition of indoor building environment to support application such as indoor navigation. We proceed to several scans in different buildings and we study the characteristics of the output models.

1. INTRODUCTION

Techniques allowing the generation of accurate 3D data and leading to realistic 3D models have been extensively investigated by researchers in computer graphics, photogrammetry and computer vision. Indeed, while the former were focussing on the creation of imaginary virtual worlds that are as close to the reality as possible, the photogrammetry and computer vision fields were mainly working on the acquisition of the information of real life scenes or objects to faithfully build their digitized versions. This gave birth to the first 3D games and 3D designing tools (e.g. CAD¹) in one hand and to acquisition techniques like stereovision and laser scanning on the other hand. More recently, 3D movies and interactive 3D sensors in the game industry (e.g. Nintendo Wii and Microsoft Kinect) considerably contributed to the familiarization of a wide audience to such technologies. At the time being, an entertainment company unable to cope with 3D is hardly competitive. Finally, the last but not least trend concerns the popularization of 3D printers that made all the manufacturers realise the high potential of 3D models. Due to all those popular tools, we assist to what could be called the "democratization of the 3D technologies".

To make them even more accessible, Google Inc. included in a basic and low cost Android tablet few sensors giving to the latter the ability to perceive its environment. The project took the name of Tango and aims at giving to a mobile device the ability to navigate the physical world thanks to spatial perception abilities supplied by advanced computer vision, image processing, and special sensors.

1.1 Related Work

The initial goal of the Project Tango is to provide to researchers and developers the needed devices to come up with new application ideas. Since the project was triggered in June 2014 (Wikipedia, 2016), there is not yet much work about it. Nevertheless, few very recent work involved the device in their investigations. For

example, in (Andersen et al., 2016) the tablet is used (among others) as an augmented reality transparent display to assist surgical telementoring.

Researches in localization seem to show particular interest in the sensors composing the device. (Agarwal et al., 2015) used the mobile version of the Tango project for metrical localization purposes based on the Google Street View. The cameras of the Tango combined to its VIO are used to match features between acquired images and Google Street panoramas. On the other hand, (Sweeney et al., 2015) relied on the Tango tablet's sensors to detect the gravity direction and estimate the absolute pose of the device in a known environment (point cloud) for localization experiments. The authors mentioned the development of a virtual reality application without further details on its availability.

Closer to our work, in (Schops et al., 2015), the authors proposed an approach for dense 3D reconstruction of large outdoor scenes based on monocular motion stereo. The method is image-based and the large field-of-view of the fisheye camera of the Tango tablet makes easy the acquisition of large areas. Interactive time processing is achieved thanks to the GPU of the device used to perform depth maps computation.

In this work, we focus our interest on the detailed specifications of the Tango tablet and its usage for indoor modelling purpose as a support to indoor navigation.

1.2 Technical Specificities

Technically, the Tango tablet is similar to tablet/phablet available in the market. At their difference, it integrates a motion tracking camera and an infrared (IR) 3D depth sensor allowing to perform scanning and to track the motion of the device, all in 3D (see Fig. 1(b)). To efficiently support the software and allow an interactive time processing of the data acquired by the sensors, it is equipped with a NVIDIA Tegra K1 192 CUDA cores processor supplied by 4 GB of RAM and 128 GB of internal storage. All that is included in a $119.77 \times 196.33 \times 15.36$ mm device with a 7" HD screen and a total weight of 370 g, that makes it very handy to use.

Common sensors that can be found in smart phones and tablets are also present in the Tango, namely accelerometer, gyroscope,

*Corresponding author
¹Computer Aided Design

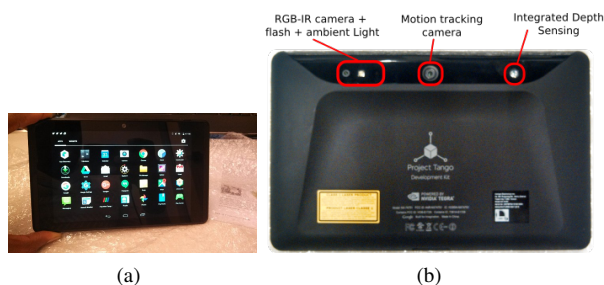


Figure 1: Google Tango tablet. (a) Interface similar to a classic Android tablet. (b) Embedded sensors for 3D perception.

ambient light, barometer, compass and GPS. Similarly, it also has the common ports and connectivity options: Micro HDMI, USB 3.0, Micro SD and furthermore a Nano SIM slot allowing to use 4G LTE network, in case of unavailability of Wi-Fi. Nevertheless, internet connexion is not required to use the tablet's hardware and software.

1.3 Embedded Technologies

Thanks to all the sensors embedded in it, the Tango tablet is able to perceive and interpret information of its surrounding environment. For this purpose, it uses three main technologies: motion tracking, area learning and depth perception.

The motion tracking technology allows the device to know in real time its own position and orientation in the 3D space. For the Project Tango, a visual-inertial odometry (VIO) (Weiss et al., 2012) approach was adopted. It is based on feature tracking in images, combined with the analysis of the data provided by the Inertial Measurement Unit (IMU) (Morrison, 1987) of the tablet. Indeed, images are acquired with the motion tracking camera that benefit of a large field of view (up to 160°) and image processing algorithms are deployed to detect corners and edges features and to study their optical flow across the acquired frames (up to 60 images/s). On the other hand, the gyroscope and the accelerometer, forming the IMU, provide accurate orientation and acceleration of the device. All those information combined give the real pose of the Tango tablet at any time in the space.

To allow the tablet to be aware of the world in which it evolves, area learning skills are made available, based on the well known Simultaneous Localization and Mapping (SLAM) approach (Thrun and Leonard, 2008). By collecting and saving visual features in an Area Description File (ADF) (Google Developers, 2015c) during the motion tracking process, a memory of the environment is made. Such information present many advantages such as making the motion tracking more accurate and performing correction of drifts that occurs due to the loss of accuracy of the motion tracking alone.

The depth perception rely on the IR depth sensor. The latter is based on a time-of-flight measurement system (Gokturk et al., 2004) allowing to accurately measure the distance of the tablet to the surrounding objects. It can handle a range of objects at distance from 0.5 up to 4 meters, and as for IR sensors in general, areas lit with high IR light sources like sunlight or incandescent bulbs, or objects that do not reflect IR light cannot be scanned well (Google Developers, 2015a).

2. SCANNING INDOOR BUILDING ENVIRONMENT WITH THE TANGO

Even though they are less accurate, mobile or hand-held scanners are known to be more suitable than fixed ones to perform indoor

scanning. They are even more interesting for application such as indoor navigation as they provide a better insight of the empty space, that stands as the navigable space. Indeed, due to the important probability of occluding objects in indoor environment, fixed scanners that usually require wide space and several poses are less efficient for such tasks. Thus the manoeuvrability of the Tango tablet is an advantage for indoor scanning.

On the other side, because of hardware limitations, not all the types of indoor environment can be handled by the Tango. Also, as a mobile device with limited resources, the duration of a scanning have to be considered.

We discuss here the output format provided by the Tango tablet, the types of buildings and rooms in which it can perform accurate scanning and we analyse the optimal duration allowing the production of a good 3D model.

2.1 Provided Output

The number of applications (apps) available and taking advantage of the Tango's sensors is currently very limited, as many projects and developments are still ongoing. Few of them, e.g. *ParaView Tango* or *RoomScanner*, exploit its depth sensor to provide a raw uncoloured point cloud, as illustrated in Fig.2(b). Combining the depth sensor with the motion tracking, some other apps provide also the position and trajectory of the tablet, or even more advanced, a triangular mesh directly. It is the case of the *Constructor* app (Google Developers, 2015b) that allows to build a textured 3D mesh of the scanned scene in interactive time (see Fig.2(c)). The raw point clouds acquired are processed on the fly and surfaces are approximated by triangular meshes. Thanks to the RGB camera of the tablet, pictures are also acquired and used for texture mapping. In this work, we use mainly the Constructor app to perform the scanning.

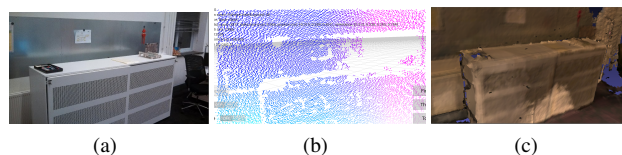


Figure 2: Outputs data provided by the Tango. (a) Real scene. (b) Point cloud generated by the Tango. (c) Textured mesh.

Taking advantage of the motion tracking skills of the tablet, the Constructor app updates parts of the scanned models where there are missing portions and allows to enhance the details as the tablet's holder evolves in the scene. There is no apparent overlapping of scanned surfaces and just like a recording tool, it is possible to pause the scanner and visualize what has been scanned while walking in the room and resume the scanning later. The final model is processed and refined before being saved in a proprietary format (.srb) or in a more general 3D mesh format (.obj). The resulting 3D model is a textured triangular mesh described as a set of vertices and their corresponding RGB color information. Figure 3 illustrates a scan of a waiting room made with the Constructor app. Figures 3 (c) and (d) allow to see the triangulation and texturing applied to the point cloud displayed in Fig.3(b).

2.2 Type of building/room to scan

The tango tablet does not allow to perform highly detailed scanning, but it can handle very well flat surfaces like walls, floors, tables, etc. Big sized furnitures are also usually well represented (bed, chair, couch, etc.). This is visible in Fig.3 where the flat areas are well reconstructed, but the triangulation is not fine enough

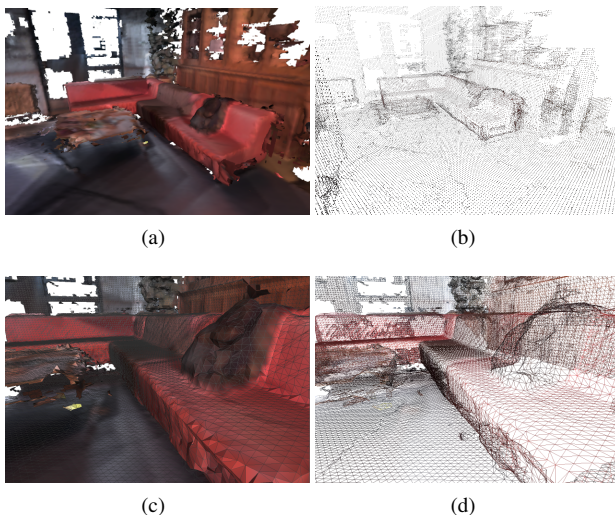


Figure 3: Example of a scan performed with the Constructor app. (a) Global view of the 3D mesh. (b) Point cloud corresponding to the points of the mesh. (c) Zoom on the textured triangulated surface. (d) Triangles coloured wireframe model.

to represent edges and corners with more precision. On the other hand, small objects (e.g. coffee mug) may not even be visually identifiable in the resulting models and may stand as artefacts, similarly to cluttered environment. Figure 4 illustrates the result of the scan of cluttered office desks.

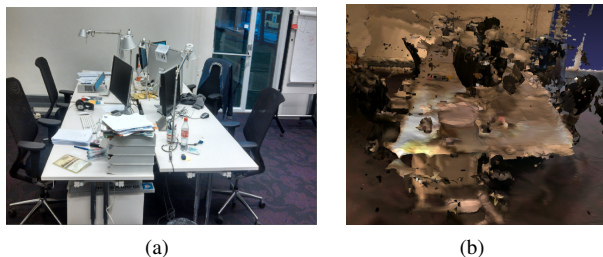


Figure 4: Cluttered desks. (a) Picture of the real scene. (b) Mesh generated by the Tango.

Similarly to Fig. 3, the boundaries of the objects are not precisely reconstructed. Moreover, due to the number of small objects on the desks, the mesh generated in that area is very noisy. This can be explained by the fact that the IR sensor of the tablet cannot reach all parts of the objects, also because of the limited distance range it is hard to update missing part even with several moves. Nevertheless, the biggest objects, when they do not absorb IR rays (e.g. screens), are better approximated and still can be visually identified (see Fig. 4 (b)).

The limited distance range of the IR sensor is also a limitation if a complete scanning of buildings with high ceilings need to be performed. It is typically the case of many public buildings, like train stations, universities, hospitals, etc., since such constructions are mainly designed to contain very large rooms able to host a large number of people. Figure 5 illustrates the case of a room in an university in which the ceiling is too high for the tablet to reach it. Thus neither details of the ceiling, nor a completely closed scan of the room can be obtained (see Fig.5 (b)).

Nevertheless, 4 meters (maximum distance range) is enough for most of the buildings of type residence, apartment, hotel, office, etc., where the ceilings of the rooms are lower, to allow

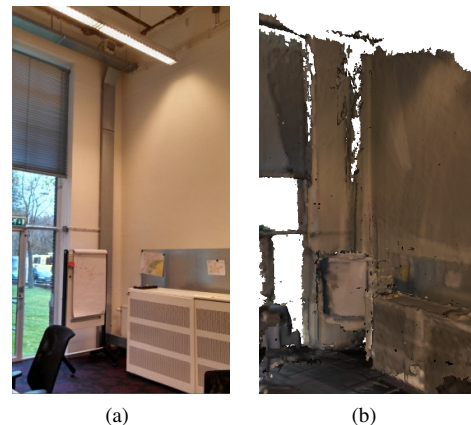


Figure 5: Problem with high ceilings. (a) University office with a high ceiling. (b) Model obtained with the Tango.

usual maintenance (painting, light change, etc.). A possibility to overcome such drawbacks is to consider image-based approaches similar to (Schops et al., 2015). Indeed the authors discussed the advantages of using depth maps obtained from images instead of methods based on the active sensors like the IR of the tablet. They also pointed out the possibility of extending the depth range and handling better sunlight illumination.

2.3 Duration of a scan

The duration of a scan using the Tango tablet is constrained by both hardware and software limitations. At its current version (development kit), the battery of the tablet allows to perform 3D scanning during approximately 1 hour and 20 minutes non stop. This time limit may not be a big problem as the usage of a power supply can be considered to overcome it. The main hardware issue is the overheating of the tablet when the scanning process lasts too long. Indeed, the real time generation of the 3D mesh implies an extensive usage of the sensors and the computation resources of the tablet. The overheating is even accentuated when the user is a human-being, since the temperature of the hands does not help to cool it down. The dilation resulting from the overheating directly affects the sensors and considerably affects their accuracy. Consequently, most of the functionality of the tablet are disabled in such case for security purpose and it is recommended to turn it off and cool it down before reuse. In our experiments, we could notice that after one hour of scanning, the acquired 3D information become too big to be processed by the tablet that also becomes unstable in use, so we could not save the data and lost all.

Another critical problem to consider is the position loss of the Tango tablet during a scan due to motion tracking failures that may happen. Indeed, the motion tracking tools and algorithms are not precise at 100%. The IMU can, for example, be affected by inconstant or sudden movement of the tablet holder, thus small positioning errors are accumulated as the scanning process goes. Because of this, we observe at some point clear displacements of still objects of the scene in the generated 3D model. Another reason of the positioning loss can be the lack of features in the scene. Since the motion tracking approach tracks features on images to estimate the motion of the camera, when a feature free area is scanned, there is no more information about the position of the scanner in the scene. The motion tracking is then considerably affected, leading to very noisy 3D models.

Figure 6 illustrate the type of important distortion that can happen during scanning. Despite the correct 3D reconstruction of

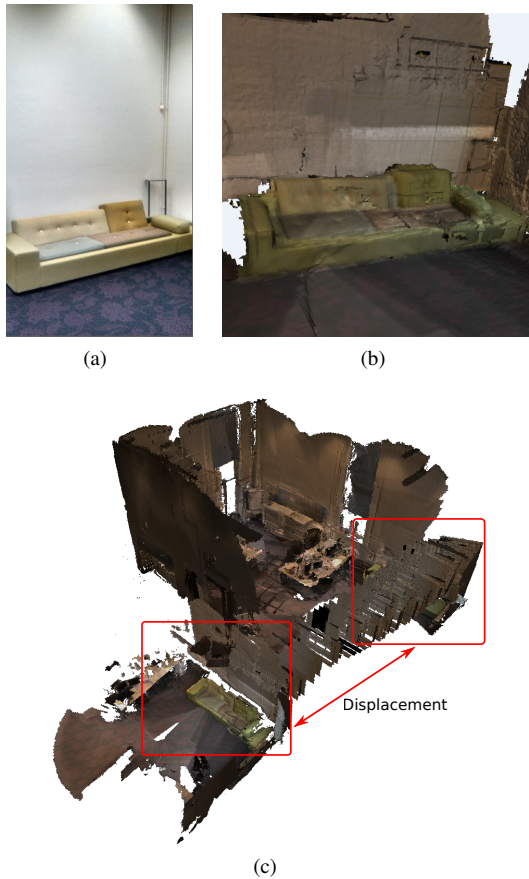


Figure 6: Motion tracking loss. (a) Couch with a wide featureless area above it. (b) 3D model of the couch. (c) Huge displacement of the scanned scene after focussing the tablet above the couch.

the couch (Fig. 6(a) and (b)), keeping the tablet on the empty wall area above the couch for 5 seconds lead to the displacement that can be seen in Fig. 6(c). This corresponds to the loss of the tablet's position in the scene before the couch area containing more features was scanned again. Such displacements alters considerably the quality of the resulting model.

Considering all those limitations, the best way to perform enough precise scanning of large buildings without much distortions is to split the scanning and perform several ones. The splitting can be based on the time of a scan of the size of the rooms. But such approach is obviously not without consequence since a direct problem that it brings is the registration of all the scans to form one full 3D model of the scanned building. Table 1 proposes a summary of the abilities of the Tango tablet for 3D scanning purpose.

3. EVALUATION OF RESULTING MODELS FOR 3D INDOOR NAVIGATION PURPOSE

To be able to perform indoor navigation in a 3D model, there are several types of information required, such as the geometrical, the semantic and the topological information (Isikdag et al., 2013). The geometry is required to spatially describe the place where the navigation should be performed, the semantic informs about the nature and property of the objects present in the scene and the topology describes the spatial relationships linking those objects and helps in defining a graph of possible paths.

The models provided by the Constructor app of the Tango contain only geometrical and color information. Thus, with the output

mesh as it is, navigation within the open space could be possible but only using metric notations, i.e. coordinates and it will be difficult to provide human understandable instructions such as "go through a given door to the next room". Nevertheless few operations can be performed to enrich the Tango models in information for such applications. Indeed, even if the 3D mesh do not provide enough information to semantically identify all the objects with precision, there are at least three crucial features that can still be easily identified: the planar surfaces, the openings and the occupied/free space. Those information could be exploited to generate an exit path in situation of emergency in a building, for example. Most available indoor navigation methods rely on 2D models or generate 2D paths. Ongoing research such as the SIMs3D Project (SIMs3D, 2015) investigate the possibility of generating 3D paths in crisis situation to provide support to involved services (e.g. fire brigades). To evaluate the utility of tools like the Tango tablet in such research, we discuss in this section the possible ways to exploits the 3D models generated by the Constructor app to produce indoor navigation paths.

3.1 Detection of Empty Space

The empty space is a crucial information in a navigation process as it allows to know where the moving subject can explore. In the case of the models provided by the Constructor app, the extraction of such space is trivial: it simply corresponds to the spaces where there is no point. It allows to have a first subdivision of the scene in a set of free space and non free space. An octree based approach can be performed to structure better the information and provide a support to differentiate the two space categories. A problem that will be encountered in such case is the non free leaves of the octree corresponding to noises in the scene. Thus denoising processes on the data will be needed to reduce them to the maximum. Identifying the free space is not enough to pretend to navigation, and more key features are required.

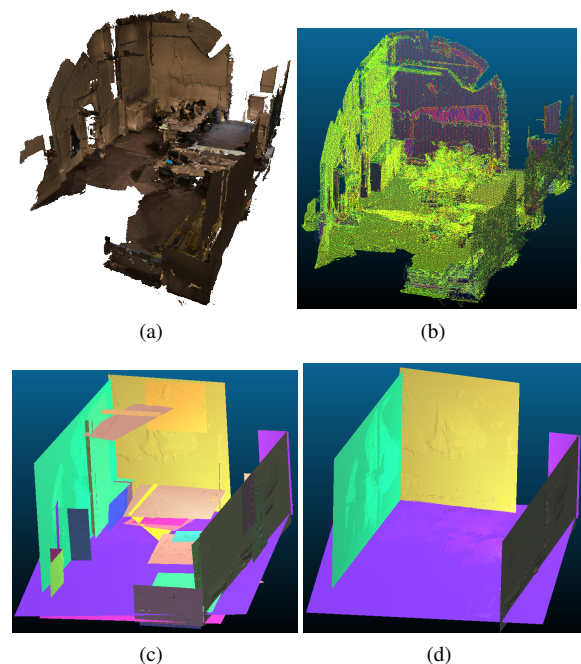


Figure 7: Planar patch detection using RANSAC. (a) 3D mesh provided by the Tango. (b) Equivalent point cloud. (c) Extracted patches. (d) Patches corresponding to walls and floor.

Range	0.5 up to 4 meters
Accuracy	Good (metric measures can be performed)
Resolution	Average (small objects are not clear)
Apps	Constructor, Paraview Tango, RoomScanner
Output	Wavefront .obj
Best scanned objects	Big, flat, with features
Worst scanned objects	Small, thin, sharp, featureless
Impossible surfaces	Deep dark, reflective, shiny, sunny
Sensitivity to light	Average
Time of continuous scanning	~ 80min
Size of a scan	~ $1.9 \times 10^3 M$ faces per 30min
Processing of data	Conventional 3D reconstruction approaches

Table 1: Summarizing table of the capabilities of the Tango Tablet.

3.2 Extraction of Planar Patches

The Tango provides a mesh dense enough to extract planar patches with classic methods reported in the literature. Figure 7 illustrates the result of a RANSAC shape detection applied on a sample model of the same office at fig. 6, without motion tracking loss. The shape detection was performed using the approach of (Schnabel et al., 2007) embedded in the free software CloudCompare (CloudCompare, 2015).

Starting from the point cloud corresponding to the vertices of the mesh triangles (fig. 7(b)), the performed RANSAC shape detection considered a minimum of 500 vertices per planar primitive, with a maximum distance of 0.01 m between the points and the patches and a maximum normal deviation of 25° . As visible in fig. 7(c), main planar surfaces of the model could be extracted. Simple heuristics based on the size, orientation and position of the patches can allow to identify the floor and the surrounding walls of the scene (see fig. 7(d)). Those particular planes and the points they contain stand as a first set of key features that bring crucial semantic information.

3.3 Detection of Openings

The information of walls and floors patches can be used to detect openings in the model also. Indeed, at the places where there are openings, there is no points or faces in the mesh created by the Tango Constructor, due to the presence of window glasses for example or just the emptiness of the area (e.g. opened door). Figure 8 illustrates a simple feature detection that allows to identify an opened door in the wall patch containing the opening.

Since there are several empty areas in the set of points fitting to the planar wall patch (see fig. 8(c)), another criterion is needed to identify the door with more precision. We know that the shape of a door is mostly rectangular, thus that is the first discrimination criterion. Hence here again, simple heuristics based on size and the distance to the floor patch of the rectangular empty areas can be used to find the proper rectangle describing the door (see fig. 8(d)). Similar approach can be used to detect windows as well and allows to gain new key semantic feature in the model.

3.4 Discussion

Despite the fact that the native models produced by the Tango tablet are not rich in information enough to directly perform applications such as advanced indoor navigation, we could see that simple approaches well known in the literature, like RANSAC shape detection, least square plane fitting and feature detection, can allow to enrich those models with basic semantic and topological information. Indeed, knowing the surfaces corresponding to the walls and the floor, the openings and the empty / non-empty

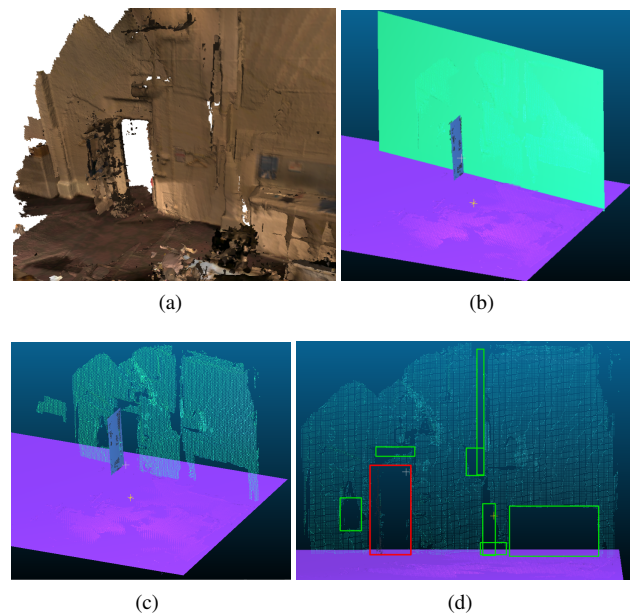


Figure 8: Door detection. (a) 3D mesh provided by the Tango. (b) Corresponding patches of the floor, the door and the wall containing it. (c) Points describing the wall in the mesh. (d) Empty areas (green) in the set of coplanar points including the door (red).

spaces in the model, it makes it possible to compute a path for a moving subject to go from one point of the model to an opening that can stand as an emergency exit for example. Even if this is still limited amount of information, the rapidity of the Tango tablet to generate such 3D model is an interesting asset in situation of emergency where the time is a critical aspect. Thus we believe that further investigation of the usage of such tool can reveal even more interesting features that will be of great benefit to several applications.

4. CONCLUSION

We presented a first investigation of the Google Tango tablet that stands as an easy hand-held scanning tool allowing to perform rapid 3D acquisition of a real scene. We reviewed the technical specificities of the tool and investigated its advantages and limits in the scope of indoor environment scanning. Furthermore we studied the output 3D model provided by the Constructor app, that is freely available and straightforward in usage. Finally we studied the possibility of exploiting the 3D models generated by the Constructor app to perform applications such as indoor navigation. And as we believe that a near future will confirm, the Tango tablet turn out to be a very interesting and promising tool

that can be a support for several applications dealing with 3D indoor models.

ACKNOWLEDGEMENTS

This research is funded by the NWO/STW M4S project Smart 3D indoor models to support crisis management in large public buildings (www.sims3d.net). We thank the Fire Brigade Rotterdam Rijnmond for making one of their buildings available for the tests.

REFERENCES

- Agarwal, P., Burgard, W. and Spinello, L., 2015. Metric localization using google street view. In: *Intelligent Robots and Systems (IROS)*, 2015 IEEE/RSJ International Conference on, IEEE, pp. 3111–3118.
- Andersen, D., Popescu, V., Cabrera, M. E., Shanghavi, A., Gomez, G., Marley, S., Mullis, B. and Wachs, J., 2016. Avoiding focus shifts in surgical telementoring using an augmented reality transparent display. *Studies in Health Technology and Informatics* 220, pp. 9.
- Gokturk, S. B., Yalcin, H. and Bamji, C., 2004. A time-of-flight depth sensor-system description, issues and solutions. In: *Computer Vision and Pattern Recognition Workshop*, 2004. CVPRW'04. Conference on, IEEE, pp. 35–43.
- Isikdag, U., Zlatanova, S. and Underwood, J., 2013. A BIM-oriented model for supporting indoor navigation requirements. *Computers, Environment and Urban Systems* 41, pp. 112–123.
- Morrison, M. M., 1987. Inertial measurement unit. US Patent 4,711,125.
- Schnabel, R., Wahl, R. and Klein, R., 2007. Efficient RANSAC for point-cloud shape detection. *Computer Graphics Forum* 26(2), pp. 214–226.
- Schops, T., Sattler, T., Hane, C. and Pollefeys, M., 2015. 3D modeling on the go: Interactive 3d reconstruction of large-scale scenes on mobile devices. In: *3D Vision (3DV)*, 2015 International Conference on, IEEE, pp. 291–299.
- Sweeney, C., Flynn, J., Nuernberger, B., Turk, M. and Hollerer, T., 2015. Efficient computation of absolute pose for gravity-aware augmented reality. In: *Mixed and Augmented Reality (ISMAR)*, 2015 IEEE International Symposium on, IEEE, pp. 19–24.
- Thrun, S. and Leonard, J. J., 2008. Simultaneous localization and mapping. In: *Springer handbook of robotics*, Springer, pp. 871–889.
- Weiss, S., Achtelik, M. W., Lynen, S., Chli, M. and Siegwart, R., 2012. Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments. In: *Robotics and Automation (ICRA)*, 2012 IEEE International Conference on, IEEE, pp. 957–964.
- Wikipedia, 2016. Project tango — wikipedia, the free encyclopedia. [Online; accessed 16-April-2016].
- Google Developers, 2015a. Overview of the concept of "Depth Perception", USA. <https://developers.google.com/project-tango/overview/depth-perception> (16 Nov. 2015).
- Google Developers, 2015b. Make 3D meshes with Constructor, USA. <https://developers.google.com/project-tango/tools/constructor> (17 Nov. 2015).
- Google Developers, 2015c. Overview of the concept of "Area Learning", USA. <https://developers.google.com/project-tango/overview/area-learning> (19 Nov. 2015).
- CloudCompare, 2015. 3D point cloud and mesh processing software. Open Source Project. <http://www.danielgm.net/cc/> (13 Dec. 2015).
- SIMS3D, 2015. Smart 3D indoor models to support crisis management in large public buildings. <http://www.sims3d.net> (13 Dec. 2015).