Evaluation of low-cost depth sensors for outdoor applications

R. Haenel [1, 2] *, Q. Semler[1], E. Semin[1], P. Grussenmeyer [2], S. Tabbone [3]

[1] SYSLOR SAS, 1 Square Camoufle, 57000 Metz, France – (raphael.haenel, quentin.semler, edouard.semin) @ syslor.net
[2] University of Strasbourg, CNRS, INSA Strasbourg, ICUBE Laboratory UMR 7357, Photogramemtry and Geomatics Group, 67000 France – (raphael.haenel, pierre.grussenmeyer) @ insa-strasbourg.fr
[3] University of Lorraine, CNRS LORIA UMR7503, Campus Scientifique, F-54506 Vandoeuvre-lès-Nancy, France – salvatore.tabbone @ loria.fr

Commission II

KEY WORDS: low cost. depth imaging, LiDAR, ToF sensors, 3D scanning, computer vision, sensor systems.

ABSTRACT:

Depth information is a key component that allows a computer to reproduce human vision in plenty of applications from manufacturing, to robotics and autonomous driving. The Microsoft Kinect has brought depth sensing to another level resulting in a large number of low cost, small form factor depth sensors. Although these sensors can efficiently produce data over a wide dynamic range of sensing applications and within different environments, most of them are rather suitable for indoor applications. Operating in outdoor areas is a challenge because of undesired illumination, usually strong sunlight or surface scattering, which degrades measurement accuracy. Therefore, after presenting the different working principle of existing depth cameras, our study aims to evaluate where two very recent sensors, the AD-FXTOF1-EBZ and the flexx2, stand towards the issue of outdoor environment. In particular, measurement tests will be performed on different types of materials subjected to various illumination in order to evaluate the potential accuracy of such sensors.

1. INTRODUCTION

Recent advances in digital technologies have made measurement tools for the creation of 3D models much more accessible. In particular, in photogrammetry, the effective miniaturization of photo sensors currently allows to create very interesting 3D models from a simple smartphone. And in a very similar way, we tend towards the same process of laser scanning as shown by the LiDAR technology integrated in the last devices delivered by Apple.

Since the release of the Microsoft's Kinect for Windows in 2010 and with all the scientific research conducted around it for computer vision purposes, several generations of cost-effective 3D depth sensors have emerged. Indeed, low-cost depth camera represents nowadays an interesting solution to bypass regular image intensity cameras or even scanning 3D laser rangefinders to provide relevant 3D action data. Each pixel of this data map has a specific numeric value representing the distance between the sensor and a target object. The objective of such sensors is to transcribe as faithfully as possible what the human eye could see and offer this information to machines so that they can understand the surrounding world.

This work is part of the research conducted by [Haenel et al., 2022] aiming to integrate depth data from any sources to build close-range point clouds in real time on smartphones. First tests were studied using data provided by the depth API from ARCore [Valentin et al., 2019]. These data are the result of a certainly efficient Machine Learning process, delivering depth information at 30 fps with interesting accuracy and reasonable precision. However, this system is very sensitive to colour and texture change and the computed depth maps suffer from a smoothing effect that tends to distort flat surfaces or create outliers described as flying points.

Therefore, this paper stands as a solution to address the encountered issue. Indeed, one idea following this work was to add an external device alongside the smartphone to bring more relevant depth data in order to improve the point cloud results.

During the following section, we will first discuss why many challenging problems can be solved by using depth cameras. We will illustrate the current applications in which depth sensing is used. This will allow us to describe the real potential of such systems both in research areas or industrial cases. And we will finish by describing the different technologies dissociating the sensors on the market, giving their strengths and weaknesses and to what extent they can be used. Then, in section 3, we will present the two sensors that have been tested. Their characteristics will be presented and put into perspective with respect to other very recent depth sensors. Finally, in section 4, we will describe the tests performed outdoors, on several surfaces, with different lighting conditions. Generally, those sensors are designed to work indoors, but we will assess their performance in outdoor environments to measure global characteristics such as accuracy or precision and evaluate the effect of colour and material on sensor performance.

2. STATE OF THE ART OF DEPTH SENSING

2.1 Depth sensing applications

The first consumer-grade depth camera, the Microsoft Kinect, has clearly revolutionized depth sensing in such a way that nowadays, depth information plays a key role in plenty of applications by solving major problems either in computer vision applications [Han et al., 2013; Lachat et al., 2015] or in human action recognition [Chen et al., 2017]. In fact, either human eye or computer algorithms can be easily fooled by patterns, perspectives or even movements. For example, people detection and tracking has been enhanced with depth sensing

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLVIII-2-W1-2022-101-2022

technology [Bevilacqua et al., 2006]. Indeed, classification of people for automatic door opening under adverse conditions (poor illumination, poor textured scenes, crowded areas, …) has been improved by adding depth information [Slattery and Shiba, 2019]. Another promising application where depth sensors are important is ADAS (*advanced driver assistance system*) and autonomous vehicle. [Hernández-Aceituno et al., 2016] have used a Kinect to map obstacles achieving more accurate results than stereovision techniques.

With KinectFusion [Izadi et al., 2011] has paved the way for innovative low-cost handheld scanning systems using such a sensor. Following this work, several additional works have arisen, aiming to study the contribution of depth sensors in the robotics field [El-laithy et al., 2012, Pinto et al., 2015]. Indeed, the cost-effective real time 3D data flow provided can be integrated into a global scheme of SLAM (*Simultaneous Localisation and Mapping)* coupled with inertial sensors measurements to perform navigation.

In 2020, alongside TrueDepth technology, depth sensing was brought to another level of accessibility with the advent of Apple's devices equipped with a LiDAR (*Lighting Detection and Ranging*). The miniaturization of the light projectors allowed to integrate active depth cameras into mobile phones more easily. This led to a bunch of new applications from digital face recognition, VR / AR (virtual /augmented reality) experience to large scale mapping.

2.2 Depth sensing technology

There are different types of depth cameras depending on how the distances are calculated, all with different strengths, weaknesses and favourable operating conditions. Depth data can mainly be obtained by structured and coded light, stereoscopy or Time-of-Flight (ToF)/ LiDAR.

Structured and coded light depth camera is based on a specific projected light represented by either dots, stripes or even color-coded pattern. Depth can be retrieved by comparing the difference between the known projected pattern and the same pattern deformed according to the shape of the observed object. Due to the high accuracy depth data achieved by this computation principle, the structured light systems are often privileged for 3D reconstruction [Kamzi et al., 2014]. For example, the first generation of the Kinect and Apple's FaceID rely on this technology. Nevertheless, this method is heavily affected by external light sources, especially outdoors, as sunlight is largely more powerful than the projected light, resulting in hard interference that degrades depth results [Kamzi et al., 2014].

The stereoscopic system used in some sensors aims to create depth by imitating human eyes, represented by two very close cameras. Solving depth involves matching points between the images using epipolar geometry; then the correspondence yields a view-difference encoding representation, also known as disparity, from which the distance to the target object can be deduced. Unlike structured light cameras, stereo cameras can use any source of light to measure depth and are therefore very efficient in changing lighting conditions. However, the scene should be sufficiently textured and contrasted. For example, in front of white walls or in the dark, stereo vision will not work properly. Moreover, this type of systems requires complex processing algorithms and the measurement range is directly related to the spacing between the two cameras. The Intel® RealSense™ D455 is one of the latest depth camera using the stereoscopic system to compute 3D information.

ToF is the most widespread depth sensing technology in commonly available depth sensors because of its simplicity and ease of use. As the name implies, this technology simply relies on measuring the time delay between a light source pulse (either a solid state laser or LED) and the light reflected received by the camera. Compared to structured light, ToF cameras are less affected by ambient light [Kadambi et al., 2014]. However, when the scene becomes too bright, the image sensor will also saturate. This explains why the second generation of Kinect, released in 2013, uses time-of-flight technology rather than structured light to improve the outdoor experience.

There are two distinct approaches used within ToF systems to calculate the light wave travel delay: pulse modulation (PM) depth sensor using direct measurements or continuous wave modulation (CWM) sensors based on indirect measure.

Direct ToF (dToF) or PM are based on a short light pulse of only a few nanoseconds. In this case, the measured delay between the emission and the reflection of the target is used to determine the distance travelled. This technique is particularly well suited for few points ranging.

Indirect ToF (iToF) or CWM relies on a more complex methodology, where a continuous modulated sinusoidal wave light signal is emitted across the scene. The receiver will continuously measure a grid of reflected rays by identifying the phase difference between the transmitted and reflected signals. The phase angle described in this way will determine the distance to the target given the known modulation frequency. These types of sensors are more likely to perform outdoors under adverse conditions because the results are in-pixel reliable.

More details about physical principles, scanning mechanism, and precise equations can be found in [Horaud et al., 2016; Slattery and Shiba, 2019].

Finally, the LiDAR technology is often confused with ToF. Rightly so, the difference between the two techniques is sometimes very thin. In a simplified way, we can say that in the majority of cases, LiDARs are ToFs but not all ToFs are LiDARs. Indeed, inside LiDAR systems, the distance to each object is calculated from the delay between the emission of the laser pulse and the return pulse in the same way as in ToF sensors. But the main difference is that this returning pulse can be composed by several echoes, each describing a particular object encountered within the scene. For example, LiDARs are particularly well suited to drone surveys in wooded areas to recover both the canopy and the ground.

	Structured and coded light	Stereoscopy	ToF / LiDAR
Accuracy	●●●●●	●●●○○	●●●●○
Range	●○○○○	●●●●○	●●●○○
Indoors performance	●●●●○	●●●●○	●●●●○
Outdoors performance	●○○○○	●●●●○	●●●○○
Camera interference	●●●●○	○○○○○	●●○○○
System complexity	●●●○○	●●●●○	●●○○○

Table 1. Comparison of the main methods used in common depth cameras (inspired by [Brading et al., 2013])

3. PRESENTATION OF THE SENSORS

3.1 Choice explanation

On the basis of the study conducted, we ended up with Table 1 which summarizes the strengths and weaknesses of the previously presented methods found in most of the commonly commercially available depth cameras. Sensors based on structured light have been excluded because as already mentioned they are not suited for outdoor areas. In addition the stereoscopy methodology have also been excluded, since we use a smartphone or tablet as the main acquisition mean but also as the user platform. These devices are barely equipped with a dual camera. We therefore decided to further investigate commercially available TOF sensors.

Within the framework of our research, we have acquired two iTOF sensors (Figure 1): the AD-FXTOF1-EBZ, and the flexx2. Our choice was guided in particular by the hardware aspect of the device, the output connectors to an external source (smartphone or tablet), either via USB or Wi-Fi, the size and the weight to ensure an ergonomic solution. In addition, we wanted to have easy access to an API (*Applicative Programming Interface*) or SDK (*Software Development Kit*) for development accompanied by wrappers to libraries such as OpenCV®, ROS®, to be able to adapt the sensor to our use cases.

(a) AD-FXTOF1-EBZ (b) flexx2
Figure 1. Selected sensors

3.2 Sensor characteristics

To highlight the characteristics of our sensors, we compared them to one of the most recent and prominent depth sensors on the market today: the Azure Kinect. Table 2 summarizes these characteristics and puts them into perspective to the Azure Kinect.

The first sensor used is one of the latest products provided by Analog Devices® to build a reliable solution for 3D computer vision applications. It enables the capture of a 640 x 480 depth map of a scene at 30 fps, which is actually one of the highest resolution found on the market. This resolution highly improves the distance measurement on small and thin objects. To work efficiently, this device needs to be coupled with a processor board from the Raspberry Pi or Nvidia family.

The second is the latest device designed by PMD Technologies® to deliver powerful 3D vision to many applications from object detection to precise gesture control. Compared to the AD-FXTOF1-EBZ and the Azure Kinect, the depth sensing performance is not as good with a resolution of only 224 x 172 pixels. As depth accuracy is linked to the resolution of the sensor, this suggests that the geometry of measured objects will be less detailed and perhaps less accurate. However, the flexx2 offers a higher frame rate (x2) and a greater distance range. But first and foremost, this sensor is really small and compact as there is no need for power supply as the AD-FXTOF1-EBZ. The sensor only needs to be connected via USB-C to a device to be used efficiently, which makes it very accessible when used with a simple smartphone.

	AD-FXTOF1-EBZ	Flexx2	Azure Kinect
Technology	iTOF	iTOF	iTOF
Resolution	640 x 480 pixels	224 x 172 pixels	NFOV: 640 x 576 pixels WFOV: 515 x 512 pixels
Wave length	940 nm	940 nm	850 nm
FOV	87° x 67	56 x 44	NFOV: 75° x 65° WFOV: 120° x 120°
Size	30 x 32 x 19,24 mm	71.9 x 19.2 x 10.6 mm	103 x 39 x 126 mm
Range	Mode 1: 25 cm to 80 cm Mode 2: 30 cm to 300 cm	0,1-4 m	NFOV: 0,5 - 3,86 m WFOV: 0,25 - 2,21 m
Frame rate	30 fps	Up to 60 fps	30 fps
Measurement accuracy	< 2% of the measured distance	1 % at 4 m	11 mm + 0,1 %

Table 2. Characteristics comparison between available sensors and the Azure Kinect

One of the main differences between the selected sensors and the Azure Kinect is the wavelength used for the light source emitter.

As previously mentioned, ToF sensors can be affected by solar radiation contamination. But sunlight itself is affected by some well-known factors such as water vapor, oxygen and CO_2, which absorb the light reducing its effective intensity [King, 2019]. Therefore, gaps within the sunlight radiation spectrum (Figure 2) created by those factors are ideal wavelengths for ToF to perform. The wavelength used in both the AD-FXTOF1-EBZ and the flexx2 corresponds to a dip of amplitude in the sunlight spectrum which means there will be less interference from ambient light compared to the 850 nm wavelength of the Azure Kinect.

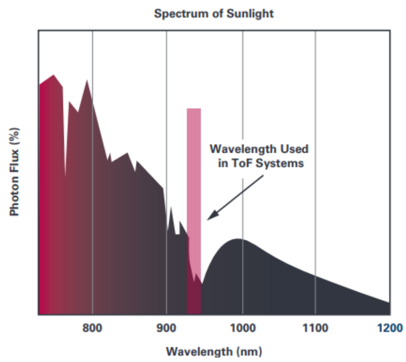


Figure 2. Spectrum of sunlight [Slattery and Shiba, 2019]

4. EXPERIMENTAL TESTS

Usually, these types of sensors are used for indoor applications because the surrounding environment is more limited and can be better controlled. Outdoors, the environment is more extended, more complex and, most importantly, affected by various lighting changes. Indeed, one of the most important factors to consider during measurements is light, because for infrared-based sensors, sunlight can saturate images resulting in poor quality depth maps [Suarez et Murphy, 2012]. In most cases, it is not possible to diffuse the lighting to reduce the error and we have to deal with it.

This potential saturation is directly related to a reflectivity phenomenon that depends on the nature of the studied materials. For example, metal surfaces are more prone to reflection problems than grass. Each one having its own characteristics, the light will be reflected differently (Figure 3). But not only that, some materials can also absorb part of the light source or diffuse it. The global geometry of the target object plays also a key role in ToF measurements.

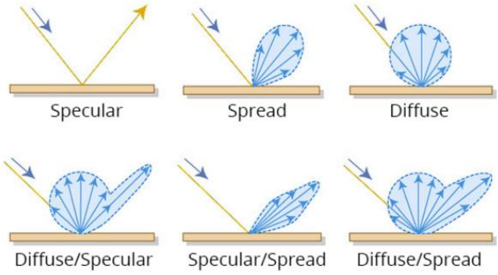


Figure 3. Different possible reflections of a light source [King, 2019]

4.1 Measurement tests on several materials

In order to evaluate the impact of lighting conditions on the sensors, we performed measurement tests on different surfaces with the sensors mounted on a survey pole. In this way, we had an accurate ground-truth distance to rely on for comparison of results, and we ensured good stability during acquisition to minimize measurement bias.

The measured distance is computed inside a 10x10 square pixel area around the center of the depth map. Although this vertical measurement is significant in determining the accuracy of the sensor, it is not sufficient in itself. Indeed, it often happens that the images are subject to distortions especially present at the edge of the image that distorts the real object. Thus, during each of the measurements, we made sure that the computed point clouds covering the field of view of the camera correspond to an average plane without major distortion (Figure 4).

The AD-FXTOF1-EBZ sensor showed very poor results described by a rather important instability of the distances (Table 3). The observed differences are less important on measurements made in the shadow than in the sun, but the oscillation is present in all cases. To address this issue, we made some test indoors to see if this problem still occurs. The tests showed that a systematic bias is always present independently of the lightning conditions. We suspect that a sensor calibration error could explain this phenomenon. This defect being major, and having only few means to compensate this error, we decided very quickly not to follow up this sensor.

Concerning the flexx2 sensor, more reliable distances were obtained (Table 4). Although occasionally, we encountered an odd result, here for gravel, in general, the observed accuracy correctly matches the officially stated values of 1 % error at 2 m. The flexx2 provides better outdoor results thanks to a patented background illumination suppression (SBI) methodology, which consists of an in-pixel circuitry that subtracts ambient light to minimize saturation effects. SBI enables a pixel-fine expansion of the dynamic range by up to a factor of 20 and as such avoiding early saturation of the pixels in case of strong sunlight. This methodology is very relevant towards our objectives because the system is self-sufficient, meaning it can outperform other state-of-the art depth cameras for outdoor applications without the need of specific optical filters or coated lenses to reduce the influence of bright sunlight. Theoretically, it was estimated that in full sunlight (100K Lumens), the measurements should lose around 10% of the maximum range compared to indoors situation, which is very reasonable.

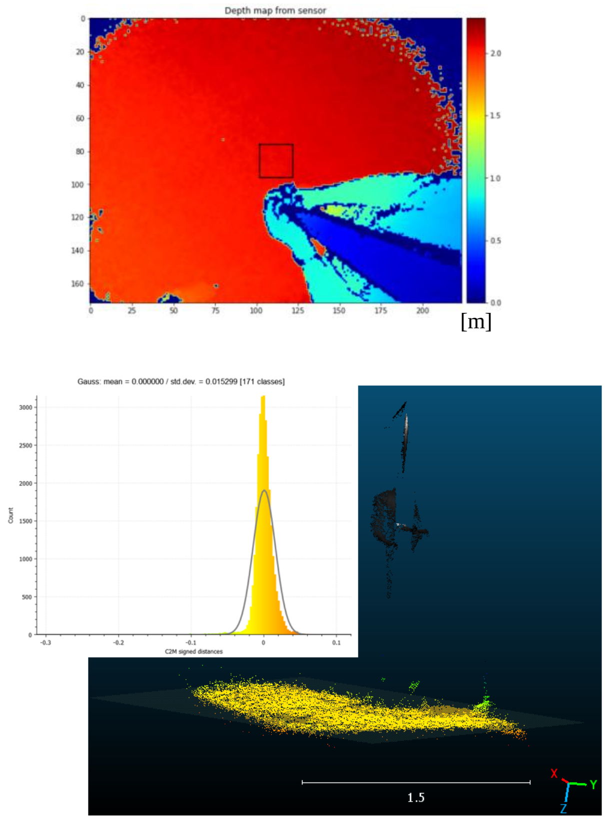


Figure 4. Provided data studied during the tests

Surface type	Measured distance		Theoric distance
	Shadow	Light	
Manhole cover	1,63 – 1,67 m	1,60 - 2,50 m	1,73 m
Macadam	1,58 – 1,73 m	1,60 – 2,50 m	
Gravel	1,71 – 1,73 m	1,69 – 1,76 m	
Grass	1,71 m	1,67 – 1,71 m	

Table 3. Accuracy analysis on different surfaces for the AD-FXTOF1-EBZ

Surface type	Measured distance		Theoric distance
	Shadow	Light	
Manhole cover	2,02 m	2,08 m	2,037 m
Macadam	2,03 m	2,12 m	
Grass	2,03 m	2,00 m	
Gravel	1,71 m	2,01 m	
Metallic surfaces	2,08 m	2,05 m	
Floor markings	2,04 m	2,04 m	

Table 4. Accuracy analysis on different surfaces for the flexx2

4.2 Relative depth error analysis

The flexx2 datasheet provides experiments on depth performance in a laboratory environment with a high reflectivity coated wall. Especially, the relative depth error has been studied. It represents the 90% quantile of the relative deviation from the ground truth according to the distance. It mainly results from a combination of statistical errors and systematic uncertainties.

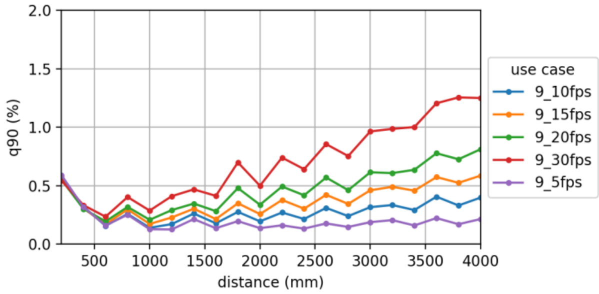


Figure 5. Relative depth error according to the distance

As represented by Figure 5, the relative depth error is expected to increase with the distance as the signal-to-noise ratio gets worse. This is due to the fact that the amplitude of the detected signal decreases with the distance and, therefore, even a small amount of noise results in a high deviation of the measurement [Oprisescu et al., 2007]. We can also see in Figure 5 that low framerates are less affected by deviation than high framerates. Indeed, during outdoor measurements we studied the effect of framerate on the provided depth maps (Figure 6).

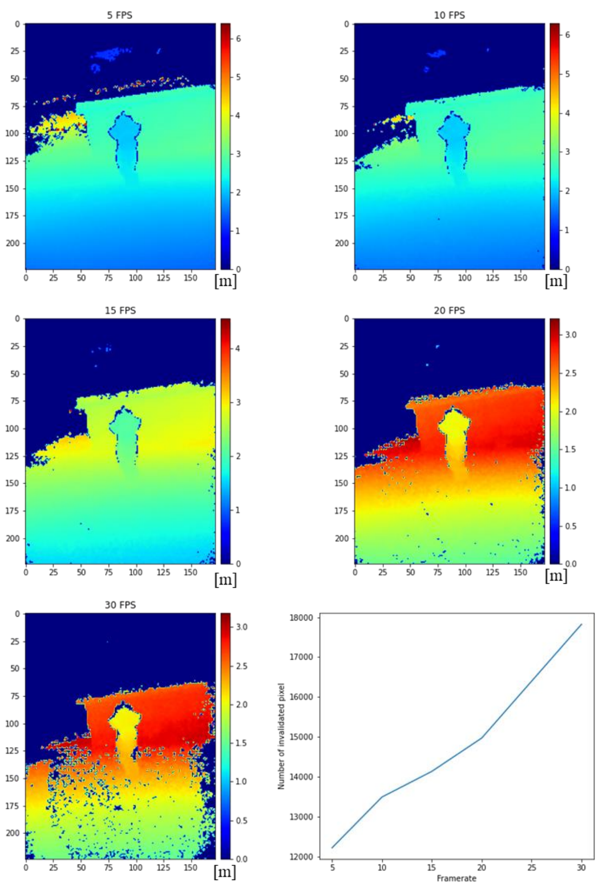


Figure 6. Influence of framerate on distance estimation

By increasing the framerate, the standard deviation related to the distance measured for each pixel increases and the confidence of estimation decreases. Then, when the confidence is lower than a specific threshold, the pixels are invalidated, which means that the related distance is going to be zero. These pixels are mainly related to far distances at low framerates (10 to 15), but this starts to affect pixels of near distances (20 to 30) causing a loss of information of the order of 46% at 30 fps compared to 5 fps. Therefore, in some cases, a high framerate is not as interesting as we think, as it induces noise multiplied by the number of frames, and without specific processing to reduce this factor, the measurements provided will be strongly degraded.

4.3 Comparison with ARCore for point cloud processing

The extraction of 3D information from a scene can be done in an active or passive way. The first method relies on a system that can directly measure the object in the scene, while the second is often the result of a complex algorithm. Very intuitively, it is logical that the active system provides more accurate results than the passive one, as errors are less propagated. But these algorithms are theoretically built to bypass problems found for active system such as the saturation for example. ARCore's Depth API is an algorithm that relies heavily on Machine Learning techniques to retrieve depth information from motion. Despite being very fast (60 fps) and effective on smartphones for creating point clouds, the acquisition scheme is really affected by this motion procedure, as the less you move around the scene the less accurate the results will be.

The main drawback of the flexx2 sensor is that there is no colour information from the measurements taken compared to the Azure Kinect. Since ARCore computes depth data based on what an RGB camera sees, a colour image can be directly linked to the provided depth map, which is missing for the flexx2 (Figure 7). The depth sensor only provides a grey-scale image resulting from the amplitude measurement of the reflected light. The perception of the environment is less good than a basic colour image, and the information present in this image strongly depends on the spectral signature of the elements in the scene Therefore, in order to recover this colour information, a calibration must be implemented to link the smartphone camera and the sensor camera.

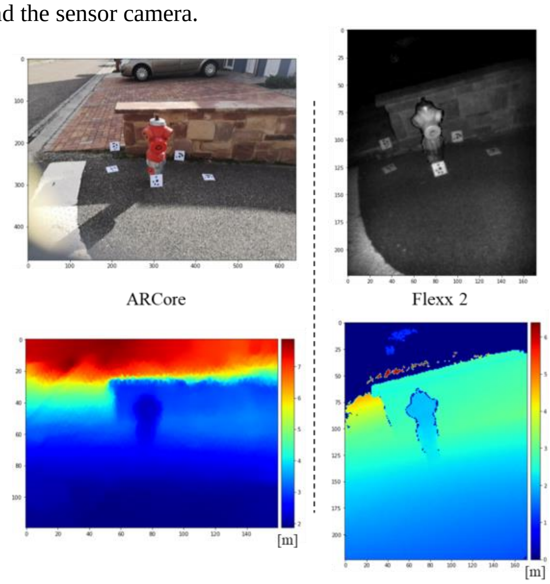
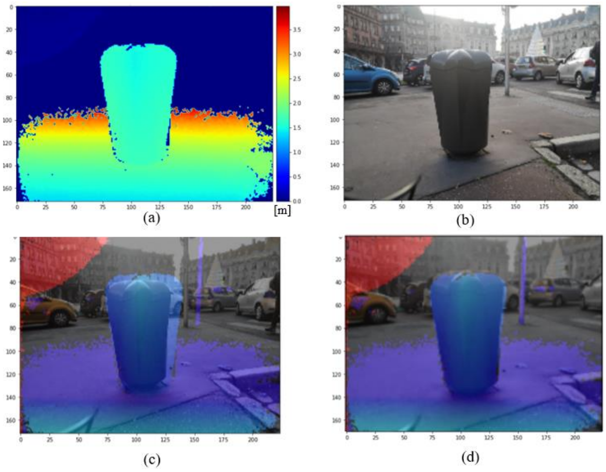


Figure 7. Data provided by ARCore's Depth API on one side and the flexx2 on the other side

https://doi.org/10.5194/isprs-archives-XLVIII-2-W1-2022-101-2022 | © Author(s) 2022. CC BY 4.0 License.

One idea could be to build an homography matrix based on the relative position of the depth sensor with respect to the smartphone camera transformed in a rotation matrix accompanied by both intrinsic matrices (Figure 8). This suggestion can be compared to a hand-eye calibration [Hartley Zisserman, 2013]. Another idea, more computational, would be to match key characteristics between a grey image from the camera and the infrared image (IR) provided by the depth sensor. A major problem with this suggestion is to ensure that the IR image provides sufficiently visually interpretable information to be used.

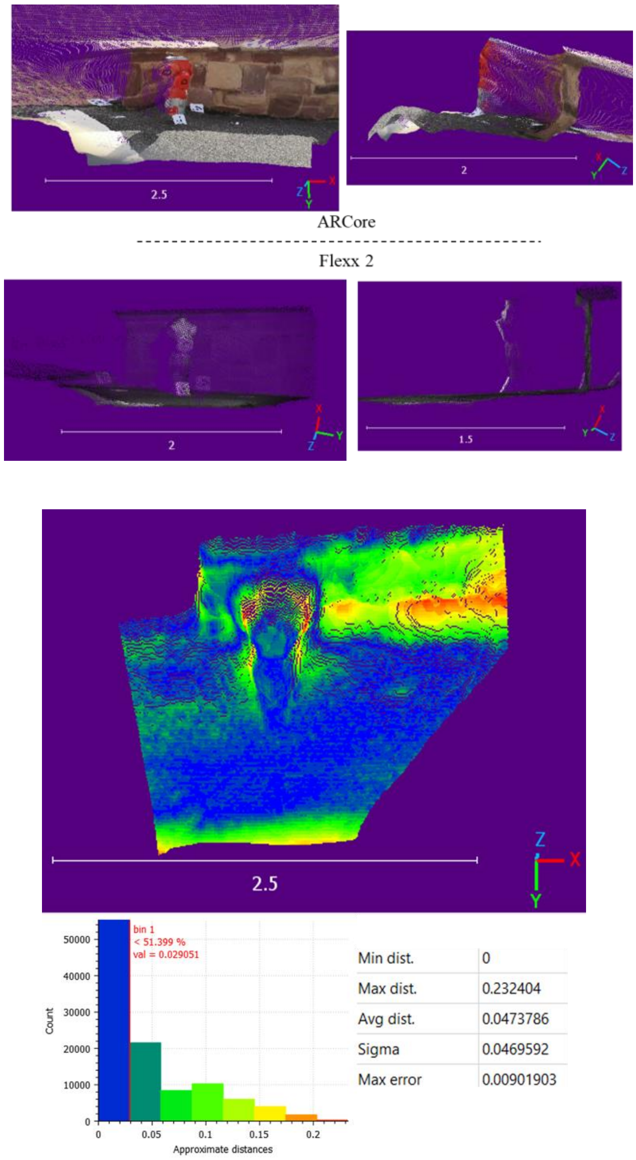


(a) Depth map (b) RGB Image (c) Overlay (d) corrected overlay

Figure 8. Colour information added on top of the depth map thanks to an homography matrix

Figure 9 illustrates the comparison between the ARCore and flexx2 data. As already mentioned, the gradient of distance in the depth map tends to deform flat surfaces especially when the algorithm failed to match characteristics because not enough key points were detected, which is the case on the low wall or on the floor, in contrast to the flexx2 point cloud where both are quite perfectly flat. The point cloud is also affected by a smoothing effect which tends to create flying points around the edges of an object. A flying point is the result of a pixel being at the corner of an object. Therefore, the distinction between foreground and background can sometimes be subject to error, resulting in false distance measurement. This is a major problem present within ARCore depth data that must be resolved by image processing [Haenel et al., 2022]. In the flexx2 framework, distances are calculated out of 5 to 9 raw images, therefore flying pixels can be detected because the reflected light is a mixture of the object and background distance, so these points can be effectively be filtered out. However, the point cloud provided by the flexx2 has an obvious lack of detail compared to the point cloud given by ARCore, resulting from the difference in resolution of the input depth map.

The final comparison between the two consolidated point clouds shows that the difference between active and passive measurements can be significant. Here, we can see a difference of 4.7 +/- 4.6 cm resulting from the different flat surfaces and the absence of flying points for the flexx2 point cloud. It proves that the addition of depth information from a specific active sensor improves results through a better understanding of the geometry of a scene without adding additional acquisition constraints such as motion tracking.



From top to bottom: computed point clouds and point cloud comparison

Figure 9. Comparison between point clouds

4.4 Preliminary results of depth sensing for underground infrastructure

In the stake of our study, we wish to bring depth sensing into a new dimension, the dimension of underground infrastructures. By underground infrastructures, we mean all urban networks such as water, electricity, gas and so on. We think that depth sensing can be really adopted to such case because all the equipment are localized inside an open trench no more than 1 to 2 metres deep. Therefore, the standard range of common depth camera such as the flexx2 are really adapted to these situations. Moreover, the configuration of pipes can be very complex and a methodology such as ARCore's depth API cannot faithfully recover the geometry. Preliminary use cases were studied using the flexx2 (Figure 10). Interesting results showed that we can easily recover the different depth of the trench while extracting the geometry of thin pipes or count the number of rubber sheath. Further tests are going to be conducted with larger trench and compared to assess to quality of such low-cost point clouds.

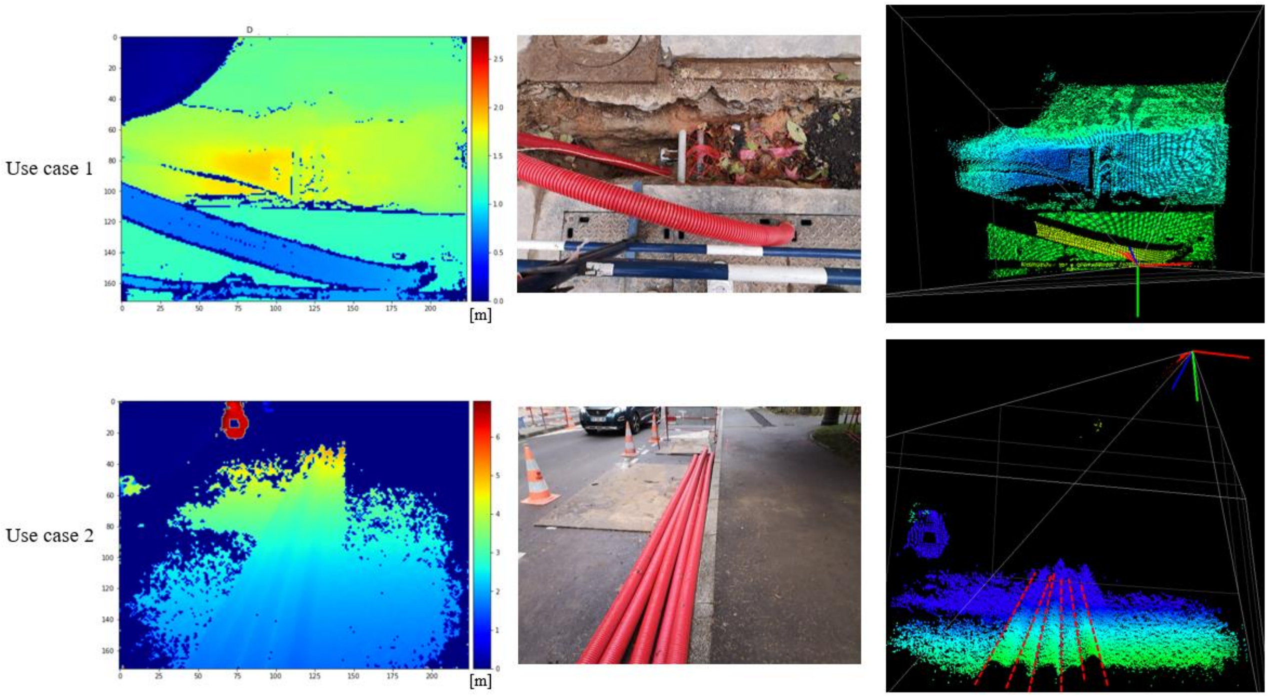


Figure 10. Two preliminary use cases of underground infrastructures

5. CONCLUSION

Over the years, depth sensing has proven to be an essential value-added feature for solving many challenging problems in various computer vision applications. By adding a third dimension to digital imaging, it is possible to construct 3D information and thus better understand the surrounding environment.

With the advancement of technology miniaturization, commodity depth sensors with small form factor and small prices have become much more accessible. This work is part of the scope of a research project aiming to integrate depth data to build point clouds in real time on a common smartphone. The idea was to find out if commercially available depth sensors could achieve interesting results outdoors under possibly adverse conditions.

First, we described the different technologies for depth sensing purpose, presenting their strengths, weaknesses and in which manners they should be used. Indirect ToF is the technology that seems to fit our needs the most as it seems to be well suited for accurate 3D imaging. Therefore, we tested two depth cameras: the AD-FXTOF1-EBZ, and the flexx2. We reviewed their characteristics while explaining how the resolution of the provided depth map plays a role in the accuracy of the geometry estimation, but also how the wavelength used for the light source can influence the robustness of data in ambient light.

Then, we tested both sensors with outdoor measurements on different materials to measure how they react to challenging surfaces. The results showed that the flexx2 is much better than the AD-FXTOF1-EBZ in terms of measurement but also in terms of accuracy compared to officially stated values. Although the AD-FXTOF1-EBZ sensor has a development architecture that is very easy to use and appropriate, the important errors encountered does not match our needs. Therefore, despite having some small defects, the flexx2 seems to be a much more relevant sensor considering its measurements robustness and small form factor.

6. FUTURE WORK

Now that we have tested the flexx2, the goal is to connect it to a smartphone to transfer the data and integrate them into the global scheme of processing point clouds. At the same time, we are going to stay aware of the new innovations to come, because the evolution of the sensors is very fast so it is not to exclude that a new powerful sensor can answer even more favourably our expectations.

REFERENCES

Bevilacqua, A., Stefano, L.D. and Azzari, P. (2006) 'People Tracking Using a Time-of-Flight Depth Sensor', in *2006 IEEE International Conference on Video and Signal Based Surveillance*. *2006 IEEE International Conference on Video and Signal Based Surveillance*, pp. 89–89. Available at: https://doi.org/10.1109/AVSS.2006.92.

Brading, M., Salsman, K., Somayaji, M., Droz, T., Van Nieuwenhove, D., Gelabert, P., Dipert, B., (2013) *Using 3D sensors to bring depth discernment to embedded vision apps*, *Embedded.com*.Available at: https://www.embedded.com/using-3d-sensors-to-bring-depth-discernment-to-embedded-vision-apps/ (Accessed: 10 October 2022).

Chen, C., Jafari, R. and Kehtarnavaz, N. (2017) 'A survey of depth and inertial sensor fusion for human action recognition', *Multimedia Tools and Applications*, 76(3), pp. 4405–4425. Available at: https://doi.org/10.1007/s11042-015-3177-1.

El-laithy, R.A., Huang, J. and Yeh, M. (2012) 'Study on the use of Microsoft Kinect for robotics applications', in *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*. *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, pp. 1280–1288. Available at: https://doi.org/10.1109/PLANS.2012.6236985.

Haenel, R., Semler, Q., Semin, E., Grussenmeyer, P., Alby, E., (2022) 'Integration of Depth Maps from ARCore to Process Point Clouds in Real Time on a Smartphone', *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2022, pp. 201–208. https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-201-2022.

Han, J., Shao, L., Xu, D., Shotton, J., (2013) 'Enhanced Computer Vision With Microsoft Kinect Sensor: A Review', *IEEE Transactions on Cybernetics*, 43(5), pp. 1318–1334. Available at: https://doi.org/10.1109/TCYB.2013.2265378.

Hartley, R. and Zisserman, A. (2003) *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN : 0521623049, 2000.

Hernández-Aceituno, J., Arnay, R., Toledo, J., Acosta, L., (2016) 'Using Kinect on an Autonomous Vehicle for Outdoors Obstacle Detection', *IEEE Sensors Journal*, 16(10), pp. 3603–3610.Available at: https://doi.org/10.1109/JSEN.2016.2531122.

Horaud, R., Hansard, M., Evangelidis, G., Ménier, C., (2016) 'An overview of depth cameras and range scanners based on time-of-flight technologies', *Machine Vision and Applications*, 27(7), pp. 1005–1020. https://doi.org/10.1007/s00138-016-0784-4.

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A., (2011) 'KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera', in *Proceedings of the 24th annual ACM symposium on User interface software and technology*. New York, NY, USA: Association for Computing Machinery (UIST '11), pp. 559–568. Available at: https://doi.org/10.1145/2047196.2047270.

Kamarudin, K., Mamduh, S.M., Shakaff, A.Y.M., Saad, S.M., Zakaria, A., Abdullah, A.H., Kamarudin, L.M., (2013) 'Method to convert Kinect's 3D depth data to a 2D map for indoor SLAM'. *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*, pp. 247–251. Available at: https://doi.org/10.1109/CSPA.2013.6530050.

Kazmi, W., Foix, S., Alenyà, G., Andersen, H.J., (2014) 'Indoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: Analysis and comparison', *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, pp. 128–146. Available at: https://doi.org/10.1016/j.isprsjprs.2013.11.012.

King, A. (2019) 'A brief introduction to Time-of-Flight sensing. Part 1 The Basics', *Terabee*. Available at: https://www.terabee.com/a-brief-introduction-to-time-of-flight-sensing-part-1-the-basics/ (Accessed: 10 October 2022).

Lachat, E., Macher, H., Mittet, M.A., Landes, T., Grussenmeyer, P. (2015). First experiences with Kinect V2 Sensor for Close Range 3D Modelling. *Proceedings of the 6th 3D-ARCH International Conference on "3D Virtual Reconstruction and Visualization of Complex Architectures"*, Avila (Spain), 25-27 February 2015, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-5/W4, 93-100, doi:10.5194/isprsarchives-XL-5-W4-93-2015, 2015

Oprisescu, S., Falie, D., Ciuc, M., Buzuloiu, V., (2007) 'Measurements with ToF Cameras and Their Necessary Corrections'. *2007 International Symposium on Signals, Circuits and Systems*, pp. 1–4. Available at: https://doi.org/10.1109/ISSCS.2007.4292691.

Pinto, A.M., Costa, P., Moreira, A.P., Rocha, L.F., Veiga, G., Moreira, E., (2015) 'Evaluation of Depth Sensors for Robotic Applications'. *2015 IEEE International Conference on Autonomous Robot Systems and Competitions*, pp. 139–143. Available at: https://doi.org/10.1109/ICARSC.2015.24.

Slattery, C., Shida, Y., (2019). 'ADI ToF Depth Sensing Technology: New and Emerging Applications in Industrial, Automotive Markets, and More'. *Analog Devices*.

Spreafico, A., Chiabrando, F., Losè, L.T., Tonolo, F.G., (2021) 'The Ipad Pro Built-in Lidar Sensor: 3d Rapid Mapping Tests and Quality Assessment'. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B1-2021, pp. 63–69. Available at: https://doi.org/10.5194/isprs-archives-XLIII-B1-2021-63-2021.

Suarez, J. and Murphy, R.R. (2012) 'Using the Kinect for search and rescue robotics', in *2012 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. *2012 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pp. 1–2. https://doi.org/10.1109/SSRR.2012.6523918.