A metrological characterization of the Kinect V2 time-of-flight camera

Andrea Corti, Silvio Giancola *, Giacomo Mainetti, Remo Sala

Department of Mechanical Engineering, Politecnico di Milano, Milano, 20156, Italy

A metrological characterization process for time-of-flight (TOF) cameras is proposed in this paper and applied to the Microsoft Kinect V2. Based on the Guide to the Expression of Uncertainty in Measurement (GUM), the uncertainty of a three-dimensional (3D) scene reconstruction is analysed. In particular, the random and the systematic components of the uncertainty are evaluated for the single sensor pixel and for the complete depth camera. The manufacturer declares an uncertainty in the measurement of the central pixel of the sensor of about few millimetres (Kinect for Windows Features, 2015), which is considerably better than the first version of the Microsoft Kinect (Chow et al., 2012 [1]). This work points out that performances are highly influenced by measuring conditions and environmental parameters of the scene; actually the 3D point reconstruction uncertainty can vary from 1.5 to tens of millimetres.

Highlights

- The temperature of the Kinect V2 has an influence in the distance measurement.
- The casual uncertainty increases with the depth and the radial coordinate.
- The systematic uncertainty follows a harmonic trend called ''wiggling error''.
- Different materials and surfaces generate small offset in the depth measurement.
- Multiple reflections generate distortions in concave 3D geometry reconstruction.

1. Introduction

Time-of-flight (TOF) systems are optical devices able to reconstruct 3D scene through the measurement of the elapsed time between the emission of a light ray and its collection after reflection from a target.

TOF cameras are particular TOF systems able to acquire a map of distances through the diffusion of the light on the whole scene, collimating the returning light on a camera matrix sensor and measuring for each single pixel the phase shift between sending and returning light. They are both faster and cheaper than 3D laser scanners: instead of using a single sensor that rotates in spherical coordinates, these cameras use a matrix of sensors and reconstruct a 3D scene with a typical frame rate of 30–60 Hz.

Initially, TOF systems where built for the automotive market, for car blind spot measurements, in order to intercept the proximity of a person or another vehicle [2]. Now, TOF systems are able to help the driver doing manoeuvres and self-driving cars are soon expected to be fully automated: they are able to perceive the world through these 3D sensors [3]. However, 3D depth cameras have had a major success with video games and entertainment. The second version of the Microsoft Kinect (Kinect V2) is an accessory of the Xbox One, that reconstructs a scene and identifies the player's body position for interaction with a virtual reality. It is also one of the most efficient low-cost TOF camera available in the market. The open source driver libfreenect2 [4] gives the possibility to use the Kinect V2 device in Linux distributions and Point Cloud Library applications.

This paper, based on the Guide to the Expression of Uncertainty in Measurement (GUM) [5], presents a metrological characterization of the Kinect V2 time-of-flight camera and a verification of its reliability and technological limits. In order to determine 3D position uncertainty, precision and accuracy have been estimated with different configurations and tests.

Similar works have already been published for the first version of the Kinect, uncertainty reconstruction has been studied in many papers [1,6–8]. The first version of the device opened a large number of possible robotic applications using perception through depth cameras [9–11].
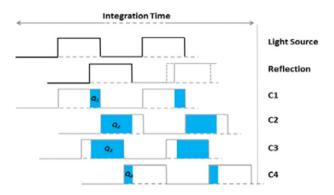
---

* Corresponding author.

E-mail address: silvio.giancola@polimi.it (S. Giancola).

Fig. 1. Four phase-stepped samples to compute phase offset ϕ [23].

Regarding TOF systems, performances has been analysed in [12,13] with previous TOF cameras. Other studies focused on performance comparison between several 3D cameras: Kinect V1 has been compared with other 3D systems in [14,15] as well as Kinect V2 in [16,17].

Reconstruction of human body with Kinect V2 has been evaluated in [18,19] for human interface purposes with the arm and the face, but the depth measurement has not been taken into account. Some applications underline the Kinect V2 characteristics, performances and weaknesses [20–22] but they do not provide any complete metrological evaluation.

In the first part of this paper, the system stability and in particular the drift due to temperature variations have been studied. Then, the analysis has been split into two levels named in the following as pixel-wise and sensor-wise. Regarding the pixel-wise characterization, the measure uncertainty related to the three positioning components, the angle of the incident light on the scene and the target surface characteristics (colour and material) has been evaluated. Successively the sensor-wise characterization based on known geometries reconstruction, mixed-pixels analysis and multiple paths error evaluation, have been performed. Before discussing the results, the TOF camera model and the Microsoft Kinect V2 main features will be presented.

2. Preliminary studies

2.1. Time-of-flight model

A time-of-flight (TOF) camera is composed of an infrared (IR) emitter, a matrix of IR sensors and an electronic circuit able to collect the reflected signal and to calculate the round trip distance applying a proper mathematical model. The two main TOF technologies on the market are the pulsed and the continuous wave (CW).

In the first case, the delay between the sent and received pulses is measured with a fast counter synchronized with the emitted signal. The resulting time delay (Δt) is converted into distance (d) by means of:

$$d = c\frac{\Delta t}{2}$$

(1)

where c is the speed of light (in air).

This approach requires a very accurate counting process, not possible to be achieved at room temperature [23].

In the second case, the CW method modulates the emitted light with a sinusoidal or, more frequently, a square wave with frequency (f) in the range of 10–100 MHz [24]. In each period, in order to obtain the phase offset ϕ, 4 samples (phase-stepped by 90°) of the reflected light, captured by the sensor, are taken (Fig. 1).

The distance d can be computed using the following formula:

$$d = \frac{c}{4\pi f}\phi \quad \text{with } \phi = atan\left(\frac{Q3 - Q4}{Q1 - Q2}\right).$$

(2)

In Eq. (2) it can be noticed that the ($Q3 - Q4$) and the ($Q1 - Q2$) terms reduce the effect of constant offset on the measurements, while their ratio reduces the effect of constant gain. Also to resolve the whole 2π range must be taken into account the sign of numerator and denominator, using the "atan2" function.

The phase-based nature of the CW method implies that there is an aliasing distance, that delimits the range of the TOF system, called ambiguity distance [23], defined as:

$$d_{amb} = \frac{c}{2f}.$$

(3)

With a single frequency technique, the ambiguity distance can only be extended reducing the modulation frequency and, as a consequence, reducing accuracy [23].

Advanced TOF systems, with multi-frequencies technology, permit an improvement both in extending the ambiguity range and in maintaining the accuracy.

2.2. Kinect V2 for Windows characteristics

The Kinect V2 is composed of a TOF CW camera with a resolution of 512×424 pixels, an RGB camera with a resolution of 1920×1080 pixels, an array of 4 microphones and some electronics for signal elaboration. The camera is able to acquire at a maximum frequency of 30 Hz. The operating field is defined by a depth range of 0.5–4.5 m and a 70° horizontal and 60° vertical view angle [25].

A Software Development Kit (SDK), provided by Microsoft, gives the possibility to write software to control the device and, in particular, to acquire and save, via USB 3.0 adapter, the depth stream, the IR stream, the colour stream and the audio stream values. The SDK also contains a function, based on the pinhole camera model, called "coordinate mapper", which transforms acquired depth map data from the camera reference frame to the world reference frame, creating a point cloud. In the end, processing the distance data returned by the TOF camera, with proprietary algorithms, the Microsoft Kinect V2 is able to track up to 6 bodies in its field of view and for each of them it calculates the position and the orientation of 25 joints.

2.3. Setup

Data acquisition, for the characterization which will be part of the next sections, was carried out using a dedicated C♯ software developed with Kinect for Windows SDK. After that, data processing and plotting was carried out with Mathworks Matlab scripts.

The most of the test described in the next sections were performed aligning the TOF camera with a planar target. This operation was carried out manually or, in the need of accuracy, with the aid of a robotic arm and an IR camera calibration procedure [26].

In the last two experiments, particular setups, described in detail in Sections 5.2 and 5.3, were employed.

3. Stability of the system

Electronic sensors and signal conditioning circuits are influenced by temperature fluctuations, that often cause output drift. Since the Kinect V2 gets warmer after some minutes of activity, it is important to analyse the stability of the system during static measurements. Moreover, inside the device a thermostat controlled fan is located, so the planned test aims to denote differences between acquisitions shot with or without a continuous air flow.

Fig. 2. Detail and general specifications of the fan used as external cooling system.

Rated voltage	12 V
Input current	0.04 A
Speed	5000 rpm
Max air volume	0.0026 m³/s

As a matter of fact, it is not possible to manually turn on or off the internal cooling system. So, an external fan has been applied over the original one (Fig. 2) in order to have the total control on the air flow. In fact, the external cooling system continuous rotation sets the internal temperature under the low thermostat threshold, as to prevent the activation of the controller and the simultaneous rotation of the two fans.

A first test has been carried out acquiring 20 000 samples at 30 Hz, placing the sensor at about one metre from a white planar wall. In order to highlight the measure trend, a moving average has been calculated on 500 distance samples returned by the central pixel of the sensor. For the entire duration of this test (10 min), the internal cooling system remained off, because the temperature of the sensor remained below the high level threshold.

A second 20 000-sample acquisition was then performed with the external fan switched on, using the same setup of the previous test. The central pixel distance data and the moving average of the two tests are plotted in Fig. 3.

While the figure on the left shows an increase of the mean value of more than 2 mm in 10 min, the test carried out with a continuous cooling system (on the right), able to maintain a stable temperature value, does not present any significant drift, but only a random spread confined into a band of 0.6 mm.

In order to obtain more information and to validate the previous tests' results, 24-hours acquisitions, with and without the external

fan, have been carried out at 1 Hz sampling frequency. The results (raw data and moving average) are presented in Fig. 4.

Without the external fan, the system presents a 4 h transient, characterized by an oscillating trend, due to an alternating operation of the internal fan. After this phase, the device temperature reaches a higher value that inhibits the shutdown of the fan, which continuously rotates and stabilizes the system until the end of the test.
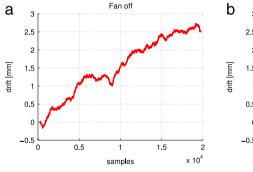
The second test confirms the hypothesis that the constant rotation of the fan can ensure a rapid stabilization of the measurement and a reduction of the transitory phase. In all the tests successively presented, a fan, fixed to the rear of the Kinect V2, will be used and maintained in constant rotation.

Finally, with the same setup of the previous test, 5000 depth samples, returned by the central pixel, have been acquired. The histogram of the distance measurements and one of the cumulative probability distribution function are depicted in Fig. 5. It can be noticed that the sample has a Gaussian distribution. For this reason, in the following part of the article, all the datasets can be considered normally distributed in order to compute mean values and standard deviations.

4. Pixel-wise characterization

4.1. Random error in the space frame

In the first part of this test, precision of the camera at different distances is evaluated. Kinect V2 was screwed on a photographic tripod and aligned with a planar white target mounted on a robotic arm flange able to translate along the camera axe with an accuracy of 0.1 mm. Because of the maximum length (1000 mm) of the completely extended robotic arm, the operating distance (800–4200 mm) was divided into 4 parts, with 200 mm overlap (800–1800 mm/1600–2600 mm/2400–3400 mm/3200–4200 mm). In each of them, the plane was translated with 20 mm steps and 4000 depth samples per pose were acquired. Only the data
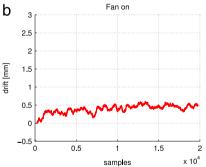

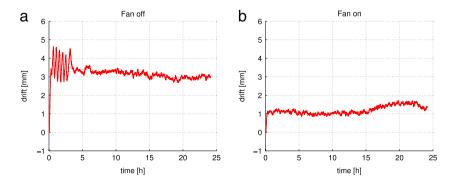**Fig. 3.** Static measurements of a single pixel in time without (a) and with (b) cooling system.

Fig. 4. Static measurements of a single pixel in time during 24 h without (a) and with (b) cooling system.

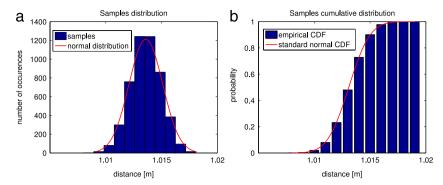**Fig. 5.** Histogram (a) and cumulative probability distribution function (b) of 5000 distance measurements performed at 1000 mm from a planar wall.
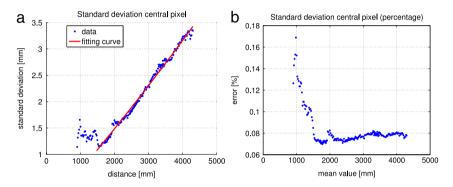

Fig. 6. Absolute (a) and percent (b) casual error in function of the distance.
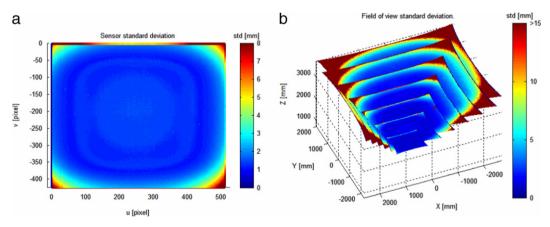

Fig. 7. Casual error in function of the pixel position (a) and along the field of view (b).

returned by the central pixel are analysed and standard deviation versus mean value is plotted in Fig. 6(a).

This test denotes a linear trend (R^2 equal to 0.9916) from 1.2 mm at about 1500 to 3.3 mm at the maximum reliable distance (4200 mm) with an initial noisy baseline. This is due to the fact that the higher is the distance from the target, the lower is the quantity of IR light caught by each pixel, and this is responsible for the degradation of the signal-to-noise ratio. Similar result can be found in [20].

Considering a set of acquisitions shot with Kinect V2 1250 mm far from a planar wall, standard deviation of each pixel is calculated and plotted into an image having the sensor's aspect ratio. Fig. 7(a) denotes a relation between the standard deviation trend and the radial coordinate of the sensor because of the IR light cone, and so the illumination of the scene, is not homogeneous [27]. This leads corner points to be more noisy than the central ones. Finally, in Fig. 7(b), the results obtained analysing seven acquisitions shot in a field of view extended from 750 to 3750 mm along Z camera axis are reported. Corner points (with standard deviation >15 mm) are

probably located outside the illumination cone and so their depth value, returned by Kinect V2, is not reliable.

4.2. Systematic error in the space frame

The aim of this test is to determine the accuracy of the single pixel distance measure. The experiment was conducted using the same setup described and data acquired in the 4.1.

In Fig. 8(a) the difference between real and measured distances is plotted. Four graphs, one for each step, are aligned minimizing the distance of the overlapping points. This harmonic shaped trend, called "wiggling error", is in agreement with literature [28,29], and depends on the non idealities of the electronics and mainly on the non exactly sinusoidal shape of the modulated signal, that contains odd harmonic components.

In Fig. 8(b) a fitting curve (R^2 equal to 0.9915), calculated as a sum of sinusoidal functions, is plotted over the measured points.

Moreover, analysing the data set acquired at 1250 mm, the mean value of each pixel's depth is calculated and a reference best
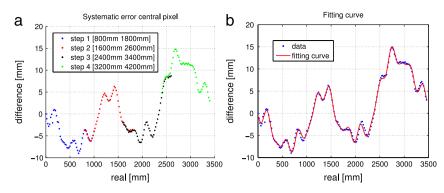
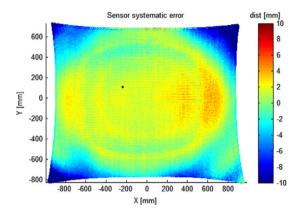**Fig. 8.** Distance systematic error (a) with fitting curve (b).

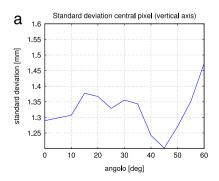**Fig. 9.** Distance systematic error inside the sensor.

fitting surface is computed on the point cloud obtained processing the meant distance data with the intrinsics parameters look up table (LUT) extracted from the SDK ("coordinate mapper"). Fig. 9 denotes a distance error confined into a band of less than 20 mm; central points present a positive displacement, while corner ones have a negative one.

Furthermore a point with an high displacement value can be noticed near the centre of the image, probably due to a defect into the sensor matrix.

4.3. Error due to the incidence angle on the target

4.3.1. Central pixel analysis

In this section, the influence of the incidence angle between IR light and the target plane on the distance measure is evaluated. A similar setup of the 4.1, with a smaller robot, is used in order to guarantee very accurate rotations around axes parallel to sensor's X and Y. From 0° (target parallel to the camera sensor) to 60°, with 5° step, 4000 distance frames are acquired, for each pose. Rotations around both a vertical and horizontal axis are performed.

Standard deviation of central pixel is calculated and depicted in Fig. 10.

From the analysis of the graphs, any particular relation between angle and single pixel distance measurement is pointed out. This test, extended to a central portion of the sensor, will be investigated into the next paragraph.

4.3.2. Centred sensor's crop analysis

This section extends the test described in 4.3.1. With the aid of a robotic arm, a planar white target is rotated into 13 poses around a vertical axis from 0° to 60° with 5° steps. Depth data returned by a symmetric centred part of the sensor are considered. Standard deviation of each pixel is calculated and some significant results are plotted in Fig. 11.

Analysing the images, a band distribution (parallel to the rotation axis) of the computed standard deviation values is well evident and denotes its correlation with the incidence angle of the single ray. The higher is the angle, the lower is the amount of light per area unit returned by the target (and so the signal-to-noise ratio).

A similar test, with same results was conducted rotating around an horizontal axis.

4.4. Error due to the target characteristics

4.4.1. Effect of the colour

Dependency of the measure on target colour is discussed in this section. For this test, different material samples have been used (cardboard, adhesive film and fabric) and for each of them, some different colours have been selected (black, white, yellow, red, blue and green). One sample at a time was attached on a planar surface, aligned with the camera, at a distance of about 775 mm and 4000 acquisitions of depth values were performed. The results are plotted on error bar graphs (Fig. 12). Analysing them, a standard deviation value of about 1 mm, constant in each set of acquisitions,
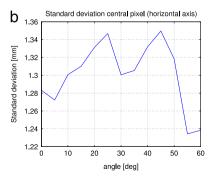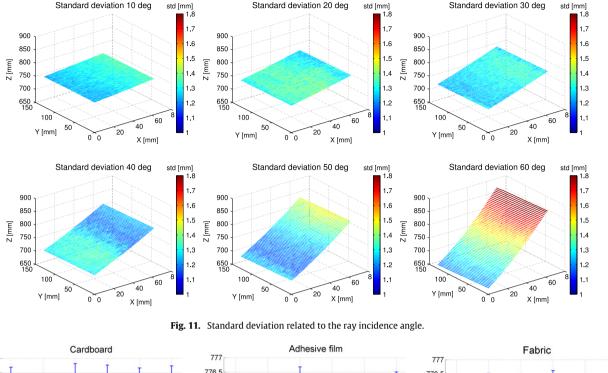

Fig. 10. Standard deviation value depending on target rotation angle around vertical (a) or horizontal (b) axis.

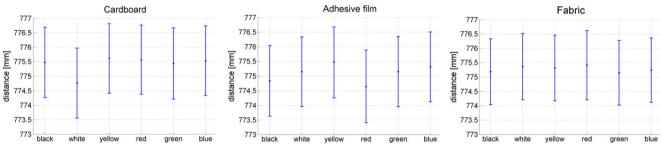Fig. 11. Standard deviation related to the ray incidence angle.

Fig. 12. Error distribution of different colour targets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

can be noticed. So it can be stated that surface colour has no effect on it.

On the contrary these tests denote an evident offset among the mean values of each set of acquired data. In particular, cardboard samples returned a 0.8 mm difference between the nearest and the farthest detected object. A similar error can be extracted from adhesive film data. For the fabric ones instead, the distance offset is confined into a 0.5 mm wide band.

Considering the three sets of results together, it is not possible to find a correlation between a precise offset value and a particular colour, so the analysis was extended to different material objects.

4.4.2. Material

Dependency of the measure on the target material is discussed in this section. For this test, samples of five common materials were used (cardboard, plastic, aluminium, wood and fabric). The test performed is similar to the previous one.

Plotting the results of this test (Fig. 13) it can be observed that the lower distance is associated with the most reflective material (aluminium). On the contrary, the object seen as the farthest by the camera, is the wooden one (characterized by a rough surface). Also in this test, standard deviation is almost constant and so no correlation between its value and the material can be pointed out.

Finally, a circular sample of reflective tissue is acquired, a commonly used material in IR computer vision. Because of its high reflectance, that saturates the camera pixels, this sample is not seen by the Kinect V2 at any distance (Fig. 14).
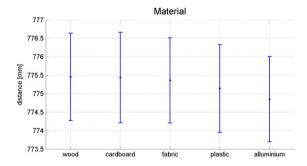

Fig. 13. Error distribution of different material targets.

Considering both the previous analyses it can be deduced that the maximum difference between the mean values is less than a millimetre and so the influence of the colour and the material of the target on the measure can often be neglected. Some measurement problems can derive from the acquisition of reflective surfaces [21].

5. Sensor-wise characterization

5.1. Known geometry reconstructions

In this section the quality in the reconstruction of known geometries like planes, cylinders and spheres is analysed, measuring the distribution of the distances between the acquired points and a mathematical model.
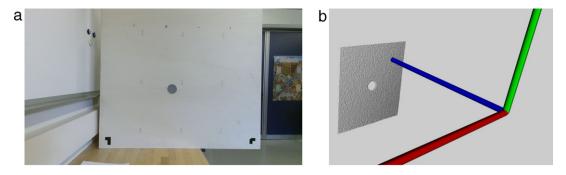
Fig. 14. Reflective tissue test: setup (a) and corresponding point cloud (b).
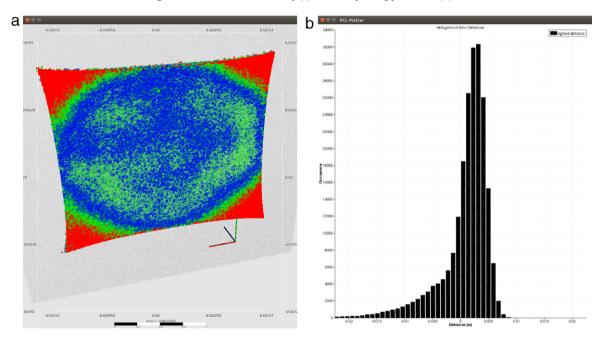

Fig. 15. Plane: point-model distance (a) and error distribution (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.1.1. Plane

A plane parallel to the sensor, about 1000 mm distant, is acquired. The point cloud has an extension of approximately 1.4 m along the X axis and 1.2 m along the Y axis. So this result confirms that the viewing angle of Kinect V2 is 70° horizontal and 60° vertical. The resolution at 1 m, along the direction identified by the X and Y axis, is about of 2.7 mm.

An operation of planar best fitting is applied to the point cloud. Fig. 15(a) shows the obtained plane aligned with that plane. The Reference system on the right of the figure shows the position of the sensor. Each points are painted in function of the distance from the model plane. A blue colour represents a closed point, a green colour an intermediate point and a red colour a far point.

Figs. 15(b) depict the error distribution into the sensor area. The greater distortion, of about 15 mm, can be observed at the corners of the image. Nevertheless, in the central part of the image, the distance is confined into a band of about ±5 mm, while, around the corners, it reaches values of tens of millimetre.

The displacement error of the corner points from the reference plane derives from optical distortions, introduced by the IR camera lens, that are not completely compensated by the mathematical model-based conversion from depth images into point clouds.

5.1.2. Cylinder

The cylinder is a simple object, commonly used to test the quality of 3D reconstruction. A cylindrical shaped object, having a diameter of about 500 mm, was placed at a distance of about 1000 mm from the sensor.

Using the Iterative Closest Point algorithm, implemented in PCL, the acquired points are aligned with a computer-generated model of the cylinder.

Figs. 16(a) and (b) highlights the distance error distribution where most of the acquired data are less than 25 mm distant from the model. Furthermore, as stated in Section 5.1.1, the points with a greater deviation (tens of millimetres) are those corresponding to the upper and lower edge of the sensor.

5.1.3. Sphere

Finally the quality of the reconstruction of a spheric surface is evaluated, scanning a portion of a 105 mm radius sphere-shaped object, 1000 mm far from the sensor. 3D points are aligned with the model with an Iterative Closest Point algorithm (Fig. 17(a)).

Error distribution (Figs. 17(a) and (b)) is almost symmetric, centred in zero and confined in a band of less than 5 mm. Only contour points are tens of millimetres distant from the model surface, probably due to mixed pixel error, described in the next section.

5.2. Mixed pixels error

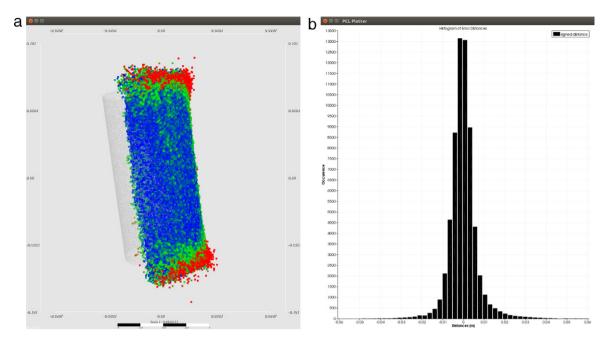This test aims to analyse the quality of edge reconstruction. A 3 mm, planar surface was posed about 350 mm far from a 1 m^2 flat

Fig. 16. Cylinder: point-model distance (a) and error distribution (b).
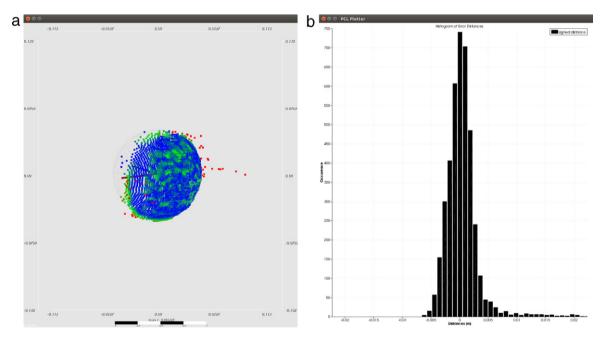

Fig. 17. Sphere: point-model distance (a) and error distribution (b).

background. Point clouds are acquired, with the Kinect V2 device manually aligned with the setup, at the distance of about 1050 mm from the target. Also the background is acquired.

The acquisition denotes some spurious points in the jump from foreground to background, as shown in red in Fig. 18(a). We used a sample consensus segmentation algorithm available in PCL for the identification of the 2 planes, as shown in Fig. 18(b). The remaining points, not belonging to any plane, correspond to the mixed pixels.

This phenomenon, also called jump edge [30], is a consequence of a weighted average operation [28], in the depth map, applied to neighbouring pixels, in order to reduce noise. As a consequence of this, an intermediate value is attributed to contiguous pixels belonging to the transition areas.

This test was extended to a situation in which the background plane is too far to be acquired (distance > 4500 mm). In this case the mixed pixel error is not present.

5.3. Multiple path error

A TOF camera well-known measure problem is the multipath error. This phenomenon originates from multiple reflections, generated into concave geometries in which part of IR light is reflected from one surface to the other, before returning to the camera sensor [24,31].

The path covered by the signal, longer than the ideal case, generates a distorted point cloud. In order to analyse the error entity, a specific setup was built with two surfaces (a white wall
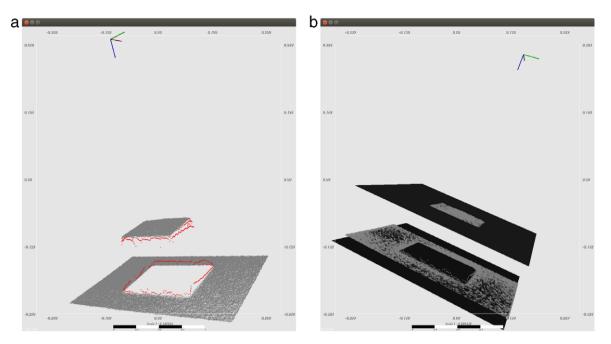
Fig. 18. Point cloud (a) where red dots denote mixed pixel and the 2 plane recognized with sample consensus algorithm (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
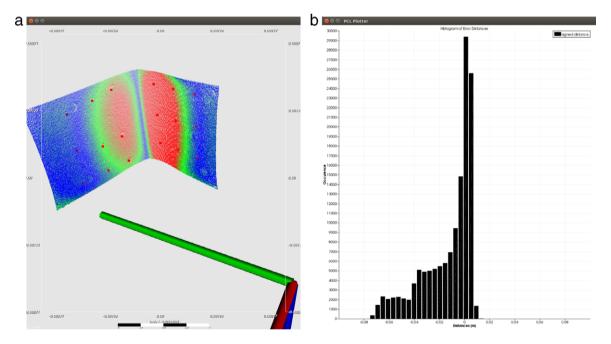

Fig. 19. Point cloud aligned with the reference markers in red (a) and error distribution (b). The wall denotes less distortion than the wooden panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and a wooden panel), opportunely aligned, in order to create a dihedral.

First of all, a computer model of the surfaces was reconstructed acquiring nine points on each plane with a Leica TS06 total station, useful to compare the point cloud returned by Kinect V2.

Because the total station and the Kinect V2 devices present a different pose and reference frame, three circular black markers are pasted on the furthest part (from the plane–plane intersection line) of each surface. Measuring their centre coordinates, with the total station, and extracting their barycentre position, with Kinect IR image analysis, it was possible to obtain the mutual transformation matrix and to align the point cloud into the total station reference frame.

Fig. 19(a) depicts the alignment between the point cloud and the reference planes described by the 9 red dots for each wall.

An error of some tens of millimetre (up to 80 mm) affects the pixels less than 250 mm distant from the plane–plane intersection line (Fig. 19(a)).

6. Conclusions

In this paper, a metrological characterization of the Kinect V2 TOF camera has been presented in order to evaluate its measurement performances and limits in 3D reconstruction. The study of the system stability has pointed out the influence of the sensor's temperature on the returned depth values. Using

an external fan, in constant rotation, this problem has been considerably reduced, obtaining more stable data.

In the pixel-wise analysis, the central pixel casual error evaluation shows that measurement uncertainty varies linearly with the displacement, up to a maximum value of 3.3 mm at 4200 mm. Moreover, analysing the depth data acquired by the whole sensor, it has been noticed that the distance measure uncertainty increases with the radial coordinate of the frame and the acquisitions shot farther than 1500 mm from the camera (along Z axis), present not reliable corners' depth values, due to a lack of illumination of the peripheral portion of the scene. The systematic error of the central pixel has a harmonic-shaped trend and can assume values into a range of $[-10; 15 \text{ mm}]$. Tests performed with a tilted target denote a correlation between the computed standard deviation value and the light-plane incidence angle. The depth measurement offset due to surfaces of different materials and colours can be neglected, reflectivity is more relevant. If the target is very reflective, the Kinect V2 is not able to perform a reliable measure.

Regarding the sensor-wise level analysis, some known geometries have been reconstructed (plane, sphere, cylinder). The maximum observed error between the mathematical model and the Kinect V2 reconstruction is about 10 mm. The mixed-pixel problem has been observed only if the target is acquired in front of a background located in the camera field of view. Finally, the multiple path error, one of the major technological issue of the TOF cameras, has been observed in Kinect V2 acquisitions generating distance errors up to 80 mm.

Future works will be focused on the analysis of measurement uncertainty in different illumination conditions and on a deep study of the measurement casual error in function of light incidence angle and NIR reflectivity of the objects surface. Comparison with currently available TOF cameras will be done in future works. The characterization will also be extended to the body joints' position measurements and tracking, as well as scene reconstruction with the Kinect V2 device (Kinect Fusion, SLAM).

References

[1] J. Chow, K. Ang, D. Lichti, W. Teskey, Performance analysis of a low-cost triangulation-based 3d camera: Microsoft kinect system, in: International Society for Photogrammetry and Remote Sensing Congress, ISPRS, vol. 39, 2012, pp. 175–180.

[2] R. Benenson, Cars perception, state of the art.

[3] M. Bertozzi, A. Broggi, A. Fascioli, Vislab and the evolution of vision-based ugvs, Computer 39 (12) (2006) 31–38.

[4] Open source drivers for the kinect for windows v2 device, https://github.com/OpenKinect/libfreenect2 (accessed: 1.07.15).

[5] B.I. des Poids et Mesures, C. électrotechnique Internationale, O. internationale de normalisation, Guide to the Expression of Uncertainty in Measurement, International Organization for Standardization, 1995.

[6] K. Khoshelham, S.O. Elberink, Accuracy and resolution of kinect depth data for indoor mapping applications, Sensors 12 (2) (2012) 1437–1454.

[7] H. Gonzalez-Jorge, B. Riveiro, E. Vazquez-Fernandez, J. Martínez-Sánchez, P. Arias, Metrological evaluation of microsoft kinect and asus xtion sensors, Measurement 46 (6) (2013) 1800–1806.

[8] S. Kahn, U. Bockholt, A. Kuijper, D.W. Fellner, Towards precise real-time 3d difference detection for industrial applications, Comput. Ind. 64 (9) (2013) 1115–1128.

[9] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al., 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, ACM, 2011, pp. 559–568.

[10] A.D. Wilson, Using a depth camera as a touch sensor, in: ACM International Conference on Interactive Tabletops and Surfaces, ACM, 2010, pp. 69–72.

[11] M. Labbe, F. Michaud, Online global loop closure detection for large-scale multi-session graph-based slam, in: Intelligent Robots and Systems, IROS 2014, 2014 IEEE/RSJ International Conference on, IEEE, 2014, pp. 2661–2666.

[12] F. Chiabrando, R. Chiabrando, D. Piatti, F. Rinaudo, Sensors for 3d imaging: metric evaluation and calibration of a ccd/cmos time-of-flight camera, Sensors 9 (12) (2009) 10080–10096.

[13] S. May, D. Droeschel, D. Holz, C. Wiesen, S. Fuchs, et al. 3d pose estimation and mapping with time-of-flight cameras, in: International Conference on Intelligent Robots and Systems, IROS, 3D Mapping Workshop, Nice, France, 2008.

[14] B. Langmann, K. Hartmann, O. Loffeld, Depth camera technology comparison and performance evaluation, ICPRAM (2) (2012) 438–444.

[15] T. Stoyanov, R. Mojtahedzadeh, H. Andreasson, A.J. Lilienthal, Comparative evaluation of range sensor accuracy for indoor mobile robotics and automated logistics applications, Robot. Auton. Syst. 61 (10) (2013) 1094–1105.

[16] T. Breuer, C. Bodensteiner, M. Arens, Low-cost commodity depth sensor comparison and accuracy analysis, in: SPIE Security + Defence, International Society for Optics and Photonics, 2014, p. 92500G.

[17] H. Gonzalez-Jorge, P. Rodríguez-González, J. Martínez-Sánchez, D. González-Aguilera, P. Arias, M. Gesto, L. Díaz-Vilariño, Metrological comparison between kinect i and kinect ii sensors, Measurement 70 (2015) 21–26.

[18] H. Fürntratt, H. Neuschmied, Evaluating pointing accuracy on kinect v2 sensor, in: International Conference on Multimedia and Human-Computer Interaction, MHCI, 2014.

[19] C. Amon, F. Fuhrmann, F. Graf, Evaluation of the spatial resolution accuracy of the face tracking system for kinect for windows v1 and v2, in: Proceedings of the 6th Congress of the Alps Adria Acoustics Association, 2014.

[20] T. Butkiewicz, Low-cost coastal mapping using kinect v2 time-of-flight cameras, in: Oceans-St. John's, IEEE, 2014, pp. 1–9.

[21] P. Fankhauser, M. Bloesch, D. Rodriguez, R. Kaestner, M. Hutter, R. Siegwart, Kinect v2 for mobile robot navigation: Evaluation and modeling.

[22] E. Lachat, H. Macher, M. Mittet, T. Landes, P. Grussenmeyer, First experiences with kinect v2 sensor for close range 3d modelling, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS.

[23] L. Li, Time-of-Flight Camera—An Introduction, Texas Instruments-Technical White Paper.

[24] A.A. Dorrington, J.P. Godbaz, M.J. Cree, A.D. Payne, L.V. Streeter, Separating true range measurements from multi-path and scattering interference in commercial range cameras, in: IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, 2011.

[25] Kinect for Windows Features, http://www.microsoft.com/ (accessed: 6.01.15).

[26] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, University Press, Cambridge, 2003.

[27] D. Piatti, F. Rinaudo, Sr-4000 and CamCube3.0 time of flight (tof) cameras: Tests and comparison, Remote Sens. 4 (4) (2012) 1069–1089.

[28] T. Instruments, Introduction to The Time-of-Flight (tof)—System Design, User's Guide.

[29] H. Rapp, Experimental and theoretical investigation of correlating tof-camera systems.

[30] T.T. Ratshidaho, J.R. Tapamo, J. Claassens, N. Govender, An investigation into trajectory estimation in underground mining environments using a time-of-flight camera and an inertial measurement unit, South African J. Ind. Eng. 25 (1) (2014) 145–161.

[31] A. Bhandari, A. Kadambi, R. Whyte, C. Barsi, M. Feigin, A. Dorrington, R. Raskar, Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization, Opt. Lett. 39 (6) (2014) 1705–1708.

Andrea Corti, born in 1989, Giussano (MI), Italy; he received his high school diploma in 2008 at LICEO SCIENTIFICO E.MAJORANA (DESIO - ITALY) and got the Bachelor Degree in Automation Engineering in 2012 at POLITECNICO DI MILANO. Now he is preparing for a master degree in Automation Engineering at the same university, working on a thesis focused on the "Metrological Characterization of the Kinect V2 Time-of-Flight Camera" in order to develop a computer application for autonomous post-traumatic upper limb rehabilitation. He is also interested in micro-controller based electronic devices development and programming.

Silvio Giancola, born in 1990 in Strasbourg (FRANCE), he graduated from a scientific high school at 17 years old. He proceeded with his study at the "Institut National des Sciences Appliquées" in his home city, with a 5- years Master Degree in Mechatronics Engineering. He spent the 5th year at the Politecnico di Milano for its master thesis, in the VBLab where he specialized in Computer Vision and 3D vision systems. He then worked at the laboratory spinoff ISS, as a software development engineer. In late 2014, he joined for Ph.D. program in the same Italian University, in 3D computer vision application.

Giacomo Mainetti, born in 1986 in Lecco (ITALY), received his high school diploma in 2005 at Liceo Scientifico G.B. Grassi (Lecco-ITALY). He got the Bachelor and Master Degree in Automation Engineering at Politecnico di Milano in 2008 and 2011. In 2010 he studied in Simon Fraser University in Vancouver (BC, CANADA). He is now a Ph.D. candidate at Mechanical Engineer Department of Politecnico di Milano. His main research activity is focused on no-contact measurement and computer vision. In particular, it is focused on the metrological analysis of 3D reconstruction techniques as stereoscopy and time of flight.

Remo Sala, born in 1961 in Casatenovo (LC, ITALY), he got the Master Degree in Mechanical Engineering at Politecnico di Milano in 1985. After some years of experiences at IBM Italia, he got his PhD in Mechanical Engineerig Department of Politecnico di Milano in 1993. He founded in 2007 the "VBLab", the computer vision laboratory of the Mechanical Department of Politecnico di Milano.

He is also the founder and current manager of two spin-offs: Ïnnovative Security Solutions" since 2006, specialized in 3D computer vision systems for industrial applications and "Steriline Robotics" since 2014, specialized in automation systems for hospital and pharmaceutical applications.