

INTEGRATION OF KINECT AND LOW-COST GNSS FOR OUTDOOR NAVIGATION

D. Pagliari^{a,*}, L. Pinto^a, M. Reguzzoni^a, L. Rossi^a

^aDept. of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy -
(diana.pagliari, livio.pinto, mirko.reguzzoni, lorenzo.l.rossi)@polimi.it

Commission V, WG V/3

KEY WORDS: Kinect, GNSS, Low-cost, Navigation, Outdoor

ABSTRACT:

Since its launch on the market, Microsoft Kinect sensor has represented a great revolution in the field of low cost navigation, especially for indoor robotic applications. In fact, this system is endowed with a depth camera, as well as a visual RGB camera, at a cost of about 200\$. The characteristics and the potentiality of the Kinect sensor have been widely studied for indoor applications. The second generation of this sensor has been announced to be capable of acquiring data even outdoors, under direct sunlight. The task of navigating passing from an indoor to an outdoor environment (and vice versa) is very demanding because the sensors that work properly in one environment are typically unsuitable in the other one. In this sense the Kinect could represent an interesting device allowing bridging the navigation solution between outdoor and indoor. In this work the accuracy and the field of application of the new generation of Kinect sensor have been tested outdoor, considering different lighting conditions and the reflective properties of the emitted ray on different materials. Moreover, an integrated system with a low cost GNSS receiver has been studied, with the aim of taking advantage of the GNSS positioning when the satellite visibility conditions are good enough. A kinematic test has been performed outdoor by using a Kinect sensor and a GNSS receiver and it is here presented.

1. INTRODUCTION

1.1 The Microsoft Kinect device

Since its launch on the market, Microsoft Kinect has appeared as a great innovation, not only for gaming and entertainment but also in a number of different research fields, mainly because it combines the technology of RGB-D cameras with a low price (currently around 200\$). It has attracted researchers from the most varied research fields, from Robotics (El-laithy, 2012, Oliver et al., 2012, Samoil et al., 2014) to Biomedical Engineering (Alnowami et al., 2012) and Computer Vision (Schindhelm, 2012, Han et al., 2013). Kinect has been sold in 2010 as an accessory for the Xbox 360 console, allowing users to interact and control the console only with the use of voice and gestures, without any hand-device. It was developed by Microsoft and the Israeli company PrimeSense and it entered the World Guinness Record as the faster selling consumer device, with 8 million of sold unit in the first 60 days.

An RGB camera, an IR camera and an IR projector composed the first generation of the Kinect device, allowing to acquire coloured and depth maps and to perform skeleton tracking at a high frame rate (up to 30 fps). In the same case, a three-axes accelerometer and a microphone array are located too. The depth data are measured using structured light techniques, based on a triangulation measurement principle (Khoshelham, 2011). The sensor can be remotely controlled from a Personal Computer, thanks to a number of libraries and Software Development Kits (SDKs) realized from both Microsoft and third parts.

The complementary nature of the data deliverable by Kinect imaging has encountered a great success in the scientific and developers communities in order to solve navigation and mapping problems. It becomes the centre of a number of studies among which an important role is represented by the solution of the Simultaneous Location and Mapping (SLAM) problem (Oliver et al., 2012, Endres et al., 2012). However, the large

majority of these studies have been conducted in indoor environment, often regarding the use of the first generation of the sensor. Because the new generation of Kinect is capable of acquiring data even under sunlight radiation, we decided to extend the use of such a sensor to outdoor navigation, developing a solution for integrating the data deliverable by the Kinect to the ones of GNSS receivers.

The rest of the paper is organized as follows: in chapter 2 an overview of the measurement principle of Kinect for Xbox One is given, together with a geometric calibration of the optical sensors and a study of depth measurements in indoor and outdoor environment. In chapter 3 a study of the depth measurements considering materials with different reflective properties is given. The proposed navigation solution based on the integration of Kinect and GNSS receivers is presented in chapter 4, discussing a preliminary kinematic test too. At the end, some conclusions and remarks are drawn.

2. THE KINECT FOR Xbox ONE

2.1 The new device

In 2014, a second generation of Kinect sensor has been advertised as a controller for the new Microsoft console Xbox One. On the summer a Windows version of the device has been sold by the Microsoft store and a new dedicated SDK has been realized, allowing controlling and interacting with the sensor hardware and creating new applications thanks to the improved sensor capabilities. The new generation of Kinect sensor (from now on Kinect v2) represents a huge improvement over the previous one, mainly because of two factors. Firstly, the imaging sensor delivers higher resolution images (1280x1090 pixels for the RGB images and 512x424 pixels for the IR and depth images). Moreover, the depth measurements are performed using a time-of-flight technology providing a more accurate and complete acquisition of the 3D scene, but also a better skeleton tracking

* Corresponding author

and gesture recognition with the addition of 6 joints to be followed. Kinect v2 has the same number of sensors of its predecessor; it can deliver visual and depth images up to 30 fps and it is capable of acquiring depth information from 0.5 to 4.5 m. In Figure 1 the device is shown, together with the localization of the imaging sensors.

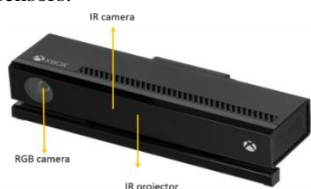


Figure 1. The Kinect v2 and the position of the imaging sensors and the IR projector

Details about the operating principles of Kinect v2 can be found in Sell and O'Connor (2014). However, it is important to underline how the IR camera is capable of delivering different output images (grey scale images dependent on the ambient lighting, active grey scale images independent from the ambient lighting and depth images, in which each pixel corresponds to the distance between the object and the sensor).

2.2 Geometrical calibration of the optic sensors and depth accuracy

A standard calibration procedure has been realized to recover the Internal Orientation (IO) parameters of both imaging sensors, using a black and white checkboard and the camera calibration app embedded in the Matlab® 2015b, which implements the Heikkila calibration model (Heikkila and Silven, 1997). The estimated parameters and with corresponding precision are reported in Table 2. Moreover, the relative orientation parameters between the IR and the RGB camera have been estimated using the same black and white checkboard. The reference system has been considered with its origin corresponding to the IR camera projection centre, with the x-axis direct along the width of the IR imaging sensor, the y-axis direct along the height of the IR imaging sensor and the z-axis to complete the right handed Cartesian system. The estimated parameters with the corresponding precision are reported in Table 3.

	RGB camera		IR camera	
	Value	Std	Value	Std
Focal length [px]	1060.30	1.09	367.65	0.48
Principal point x [px]	961.05	0.82	253.72	0.32
Principal point y [px]	539.79	0.80	206.91	0.32
K_1	$5.13 \cdot 10^{-2}$	$1.90 \cdot 10^{-3}$	$1.12 \cdot 10^{-1}$	$2.51 \cdot 10^{-3}$
K_2	$-3.55 \cdot 10^{-2}$	$5.81 \cdot 10^{-3}$	$-2.27 \cdot 10^{-1}$	$7.95 \cdot 10^{-3}$
P_1	$-2.13 \cdot 10^{-3}$	$2.61 \cdot 10^{-4}$	$-1.90 \cdot 10^{-3}$	$2.95 \cdot 10^{-4}$
P_2	$3.28 \cdot 10^{-4}$	$2.82 \cdot 10^{-4}$	$-6.61 \cdot 10^{-4}$	$3.14 \cdot 10^{-4}$

Table 1. Estimated IO parameters of the imaging sensors

	Value	Std
X [mm]	50.98	2.72
Y [mm]	-0.36	2.92
Z [mm]	0.71	3.62
Omega [rad]	$3 \cdot 10^{-3}$	$9 \cdot 10^{-3}$
Phi [rad]	$6 \cdot 10^{-4}$	$7 \cdot 10^{-3}$
Kappa [rad]	$-1 \cdot 10^{-2}$	$3 \cdot 10^{-3}$

Table 2. Estimated relative orientation parameters between the two imaging sensors of Kinect v2

Kinect v2 is characterized by higher quality imaging sensors and the new device is a huge improvement over the first generation for indoor applications, see Pagliari and Pinto (2015), Lachat et al. (2015). Furthermore, the new generation of Kinect is capable of acquiring data even outdoor, so it is important to study the accuracy and the precision of the depth data acquired under the influence of sunlight radiation. Therefore, the depth calibration procedure presented in Pagliari and Pinto (2015) has been repeated outdoor. In Figure 2 the depth measurement error, as a function of the distance between a flat object and the sensor, is shown. For each step, the average distance of each pixel over a sample of 100 depth images has been computed; then these values have been in turn averaged obtaining a single value representative of the measurement performed by Kinect v2, to be compared with the distance measured using a laser meter. In Figure 3 the sensor noise is shown. It has been computed as the average value of the standard deviations computed over the 100 depth images, pixel by pixel.

From the calibration procedure it comes out that the Kinect v2 is capable of acquiring data even outdoor (in shaded conditions), with a level of accuracy comparable to the one that is obtained indoor. In both cases, the maximum error is below 0.024 m. The depth measurements performed outdoor are noisier than the ones realized indoor, however they are more precise of those obtainable using the old generation of Kinect. It is necessary to point out that a systematic error in the measurements is also included because the position of the optical centre of the depth camera is not known and the reference distances have been measured in correspondence with the external case of the sensor itself.

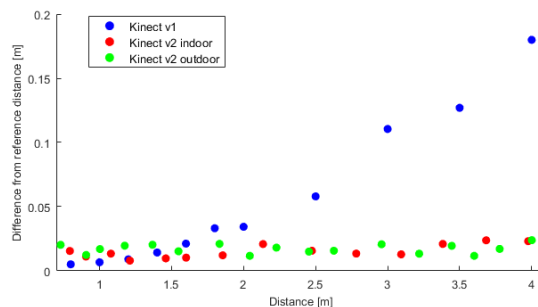


Figure 2. Estimation of the error committed by Kinect sensors as a function of the distance between the object and the device

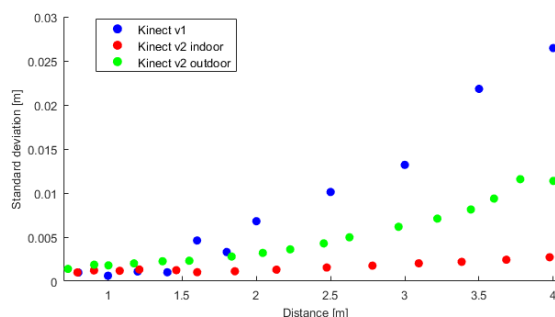


Figure 3. Standard deviations of the depth measurements acquired by Kinect sensors as a function of the distance between the object and the device

3. DEPTH MEASUREMENT RESPONSES OF DIFFERENT MATERIALS

3.1 The experimental set-up

The higher quality of the imaging sensors and the new measurement system allow acquiring scenes that are more complete; however, it is also important to evaluate some possible sources of error due to the properties of the inspected material, as well as the influence of the sunlight radiation. We realized an experimental setup similar to the ones presented in Lachat et al. (2015), but we also extended our analysis outdoor. In order to evaluate the effect of different materials and of the surface roughness, a panel with a number of square samples with a side of 0.2 m has been realized (see Figure 4). Four checkboard targets have been placed at the corners of the calibration panel.

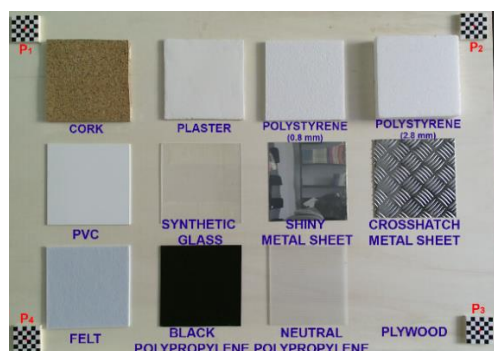


Figure 4. The panel with the samples of the chosen materials. The target points used to georeference the different point clouds are indicated by red dots

We acquired a reference dataset by means of a Leica MS60 multistation, which allows a mean spatial resolution of 0.001 m and millimetric accuracy. The coordinates of the four target points (located at the centre of the four checkboards) have been measured with the multistation too, thus allowing to georeference the different scans. The Kinect v2 has been placed in front of the panel and a total number of 100 depth images have been acquired, without moving the Kinect v2 sensor, both indoor and outdoor. At the end of the acquisition, only a single resulting image has been computed by averaging the values of each corresponding pixel of the image set. Then, it has been corrected from the lens distortion using the estimated IO parameters of the IR camera; this is possible because the IR and the depth camera are co-registered. The effect of the correction is quite remarkable: even if the plywood panel was not completely flat nor placed fully vertically, it is quite evident the presence of a barrel distortion (see Figure 5). Starting from the average depth image corrected from the lens distortions, the corresponding point cloud has been generated.

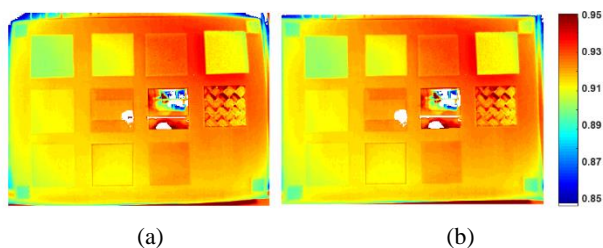


Figure 5. The resulting averaged depth images before (a) and after (b) the distortion correction. The colour scale represents the measured distances in meters

For the Kinect dataset, the position of the target has been automatically measured on the IR image (using the function detectCheckerboardPoints, embedded in Matlab® 2015b). Then, the corresponding object coordinates have been computed by using the data registered in the averaged depth map and the IO IR camera parameters. The reference system has been defined with its origin coinciding with the IR camera optical centre, the z-axis outgoing from the camera plane, the x-axis along the width of the imaging sensor and the y-axis to complete the right-handed Cartesian system. The seven parameters of the 3D Helmert transformation that describe the rototranslation and the scale factor between the Kinect scans and the multistation dataset have been estimated via Least Squares adjustment. The residuals between the coordinate measured with the multistation and the estimated ones are shown in Table 3.

	X [m]	Y [m]	Z [m]
P ₁	0.003	0.001	0.001
P ₂	-0.004	-0.001	-0.003
P ₃	0.005	0.001	0.002
P ₄	0.004	-0.001	0.001

Table 3. Residuals between the target points used to georeference the different scans

The georeferencing error can be ascribed to the GSD (Ground Sample Distance) of the IR image (equal to 0.003 m at a distance of 1 m, as during the survey). The differences between each point of the Kinect indoor point cloud and the nearest point in the reference scan, selected by considering the shortest 3D Euclidean distance, are shown in Figure 6.

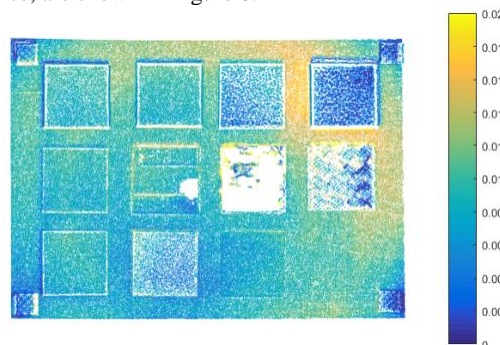


Figure 6. Differences between the reference scan and the point cloud derived from Kinect indoor acquisition (meters)

From the analysis conducted on the different materials some interesting considerations have emerged. Firstly, it is interesting to notice the presence of white borders around each material, due to the different perspective and resolution of the two instruments.

As one can expect, there is no correct response from the shining metal sheet, and the few distances that were measured correspond to objects reflected by the sheet itself. Because this material causes troubles to the MS60 too, it has been excluded from further analyses. Another interesting remark can be done by considering the reflective properties of different coloured materials. In particular, the Kinect depth measurement responses to the two polypropylene samples underline how the black version of the material was characterized by a lower response than the neutral one, in terms of number of reflected points. A more detailed analysis was conducted for the single material (see Table 4), reporting the average differences and the corresponding standard deviations for an indoor and outdoor test.

The differences between the reference scans and the point clouds generated from the depth images acquired by the Kinect v2 are in the order of 0.01 m, for all the considered materials. It is worth

to underline how the results delivered by Microsoft sensor are comparable for both indoor and outdoor shaded acquisitions. The only remarkable difference can be noticed for the crosshatch metal sheet, because this material partially disperses the sunlight radiation, so it is more difficult for the Kinect v2 device to receive a good reflected signal response when working outdoor.

Material	Indoor		Outdoor	
	Mean Difference [m]	Std [m]	Mean Difference [m]	Std [m]
Cork	0.012	0.006	0.012	0.004
Plaster	0.009	0.002	0.012	0.003
Polystyrene [8 mm]	0.004	0.003	0.005	0.008
Polystyrene [28 mm]	0.006	0.002	0.007	0.002
PVC	0.009	0.001	0.011	0.001
Synthetic Glass	0.009	0.001	0.010	0.002
Crosshatch metal sheet	0.010	0.008	0.027	0.018
Felt	0.008	0.001	0.008	0.001
Black Propylene	0.006	0.001	0.007	0.002
White Propylene	0.008	0.001	0.008	0.003
Plywood	0.010	0.001	0.008	0.001

Table 4. Differences between the reference scan and the generated point clouds for each material

4. OUTDOOR NAVIGATION WITH THE KINECT

Since its launch, the Kinect has been proficiently used for a number of robotic applications, both for navigation and mapping tasks (see for instance Suarez and Murphy, 2012, Omara and Sahari, 2015, Fankhauser et al., 2015). Quite often, the proposed solutions are based on RGB-D SLAM; they can take advantage of pure depth information (Izadi et al., 2011) or depth-coloured data (Endres et al, 2012). However, they are restricted to indoor environments, because the poor performances of the first generation of the Kinect sensor in bright sunlight make it unsuitable for outdoor environment. To our knowledge, the use of Microsoft Kinect for outdoor navigation has been considered by few authors. Quayyum and Kim (2013) discussed an inertial-SLAM fused solution for outdoor navigation, while Fankhauser et al. (2015) evaluated the impact of the light incidence angle for depth measurements.

4.1 RGB-D and GNSS integration

We proposed to use the Kinect v2 for outdoor navigation, integrating the data deliverable by this sensor with a low-cost GNSS receiver, with the aim to take advantage of the satellite positioning when the received signal is good enough. The choice of a low-cost device was driven by the idea of realizing a low-cost integrated system together with the Kinect device. The visual and depth images acquired by the Kinect v2 are used to recover a first approximate trajectory that is then refined in a Kalman filter (in which visual, depth and GNSS data are integrated). All the data acquired during the survey are stored on a Personal Computer (PC), together with the PC acquisition time, that plays the role of the bridge between the Kinect v2 and GNSS receiver time scales. The visual and depth data acquired by the Kinect are pre-processed to correct them from lens distortions and they are then used to compute the photogrammetric-based solution.

Considering the huge amount of data delivered by the Kinect v2, the images are downsampled in time, preserving the correspondence between RGB and depths and ensuring a sufficient baseline between two subsequent acquisitions. Then the depth images are interpolated (with a sampling rate 4:1) in order to have a resolution comparable to that of the RGB camera. This step is required to create RGB-D images without sparse depth information. Each of the RGB-D image is a six dimensional array in which the first three channels contain the RGB information and the last three channels contain the point clouds derived from the depth image simultaneously acquired with the RGB image. The coordinates of the point clouds are referred to the RGB camera reference system. RGB-D images are very interesting for solving navigation problems because they allow integrating visual and 3-dimensional information within the same structure, allowing taking advantage of both kinds of data while determining the alignment between subsequent acquisitions. The core of the alignment procedure is based on the function presented in Xiao et al. (2013). It takes two RGB-D images as input, combining the reliability of SIFT (Lowe, 2004) keypoints, extracted from the RGB channels, with the 3-dimensionality of the point clouds to compute the roto-translation between them. The algorithm firstly detects the SIFT keypoints using the information stored in the RGB channels and estimates the alignment using a 3-point algorithm combined with RANSAC (Fisher and Bolles, 1981). This initial alignment is refined by using the ICP (Iterative Closest Point) on the point clouds. If the ICP solution drifts too much, it is discarded and only the SIFT alignment is considered. The final photogrammetric positions and orientations are computed incrementally, adding the roto-translation between the last couples to the previously computed solution. Because the trajectory is recovered starting from the previous results, the error accumulates over time, leading to a possible drift of the computed solution. For this reason, it is then integrated with the GNSS observations in a Kalman filter (Kalman, 1960).

4.2 Kalman filtering

In the proposed implementation of the Kalman filter, the state vectors $\underline{X}_j = \underline{X}(t_j)$ at the epoch $t_i = i = 1, 2, \dots, N$ is composed by the positions \underline{x}_i and the velocities \underline{v}_i , assuming that both the centre of projection of the Kinect RGB sensor and the phase centre of the GNSS antenna are located in the centre of mass of the vehicle. The basic consequence descending from it is that the motion between two subsequent epochs can be modelled ignoring the device rotations and considering the system as a point:

$$\underline{X}_j = \begin{bmatrix} \underline{x}_j \\ \underline{v}_j \end{bmatrix} \quad (1)$$

A uniform rectilinear motion is assumed between two epochs, using a velocity model error $\underline{\varepsilon}_i$ that allows changes of direction:

$$\underline{X}_{j+1} = T \cdot \underline{X}_j + \underline{\varepsilon}_j \quad (2)$$

The transition matrix T_i between two subsequent epoch t_i and t_{i+1} can be described as:

$$T_i = \begin{bmatrix} I_3 & \Delta t_{i,i+1} \cdot I_3 \\ 0 & I_3 \end{bmatrix} \quad (3)$$

where I_3 is a 3x3 identity matrix and $\Delta t_{i,i+1}$ is the time interval between two subsequent acquisition epochs, either of the Kinect or of the GNSS receiver.

The developed filter is based on the integration of two kinds of observations, computed respectively from the data acquired by

the GNSS receiver and the Kinect v2 sensor. The deterministic model of the observations \underline{Y}_i can be written as:

$$\underline{Y}_i = H_i \cdot \underline{X}_i \quad (4)$$

where H_i is the design matrix, also called transformation matrix. As for the Kinect, the observations consist in the displacement vectors as derived from the purely photogrammetric solution and computed considering two subsequent Kinect RGB-D acquisitions at the epochs t_k and t_{k+1} ; note that $\{t_k\} \subseteq \{t_i\}$. Because of the hypothesis of instantaneous uniform rectilinear motion, the rotation between two subsequent acquisitions is not modelled in the proposed vehicle dynamics. The design matrix H_k , that corresponds to a Kinect acquisition epoch t_k , is therefore described as:

$$H_k = [I_3 \quad \Delta t_{k,k-1} \cdot I_3] \quad (5)$$

where $\Delta t_{k,k-1}$ is the time interval between each Kinect acquisition and the previous one.

Concerning the GNSS observations, pseudoranges have not been directly considered; this would have required a proper linearization and the use of an Extended Kalman Filter (see Realini and Reguzzoni, 2013). Instead, the positions computed for each single epoch are used (see Brovelli et al., 2007). The design matrix H_j , that corresponds to a GNSS acquisition epoch t_j , with $\{t_j\} = \{t_i\} \setminus \{t_k\}$, is described as:

$$H_j = [I_3 \quad O_3] \quad (6)$$

where O_3 is a 3x3 null matrix. Note that actually H_j is the same for any GNSS acquisition time t_j .

As for the stochastic model of the observations, the GNSS solution is furnished with covariance matrices, which have to be taken into account. This kind of information is not available for the Kinect solution, which instead is equipped by an empirical value that is representative of the goodness of the rototranslation matrix estimated between two RGB-D frames. This empirical value, properly scaled, is taken as the error variance of the measured Kinect displacement, while the error correlation between the three components of this displacement is neglected.

4.3 The experimental test

An experimental test was realized to evaluate the precision that can be reached with the proposed method and the effect of combining visual and depth data information extracted from Kinect v2 for outdoor positioning. We installed the Kinect v2 on a cart; the device was slightly rotated looking downward in order to guarantee the presence of some surfaces in the range of 4 m for the entire survey. A u-blox AEK6T antenna was fixed on the top of the Kinect case, in correspondence with the RGB camera. The antenna was placed horizontally, due to its planar characteristic. On the same cart, the battery used to power up the Kinect and the PC used to control it were located too. On the PC, the software used to save the data acquired by the low cost GNSS receiver and the PC-time corresponding to each acquisition was installed too.

The kinematic test was realized by moving the cart at crawl velocity around Andrea Casella's fountain (Piazza Leonardo da Vinci, Milan). During the survey both RGB and depth images delivered by Kinect v2 were stored, together with the corresponding PC acquisition time. At the same time, the GNSS data (together with the corresponding PC time) were acquired by the u-blox antenna (at 1 Hz frequency) and saved on the same PC. In order to have a reference trajectory for evaluating the goodness of the recovered solution, we installed on the same

vehicle also a double frequency Leica GS14 GNSS receiver with its own antenna and a 360° reflective prism that is tracked by means of a Leica MS-60 multistation (see Figure 7).

The track was obtained by using both a GNSS receiver and a multistation, because the latter was not able to measure the full track due to the presence of some obstacles that limited the visibility of the eastern part of the track, while the GNSS solution guaranteed the continuity of the observations, even if with less accuracy. The GNSS observations were processed by using a RTK approach, relying on the SPIN GNSS network. The acquisition was done at 1 Hz frequency, measuring about 2300 points with a mean accuracy of 0.02-0.03 m. As for the multistation, the Leica MS-60 was set to record a measure when the displacement of the prism is greater than 0.01 m, observing in total about 4000 points.



Figure 7. The cart used for the experimental test with all the instruments installed on it

In order to be able to adjust the trajectory measured with the multistation in the same GNSS reference frame, two station points were materialized and reciprocally measured both with the multistation and with static GNSS observations. GNSS data are post-processed using the same permanent reference station used for the RTK solution, thus guaranteeing that the results are in the same reference frame. Those observations were used to fix the North direction of the multistation reference frame, adjusting all the observations together in the global GNSS reference frame using the open source GeoNet software (Rossi et al., 2011). This led to an estimation of the track with an accuracy of some millimetres. Finally transforming the results in a local East, North, Up reference system the GNSS RTK and the multistation track were synchronized on the GNSS local time, by minimizing the residuals between points for the East and North components. The Up direction was instead used to estimate the vertical bias between the prism and the antenna.

4.4 Data processing

The data delivered by Kinect v2 have been processed with a Matlab® in-house software that incorporates the RGB-D alignment function presented in Xiao et al., (2013). The Microsoft device can acquire a huge amount of data; it is capable of acquiring up to 30 fps, however we verified that with a battery supplied Netbook (i7 4700MQ @2.40GHz, 12 GB of RAM and SATA HD @5400 rpm class) the actual acquisition frame rate is about 5 fps. Because the cart was pulled very slowly, a high overlapping can be guaranteed without having to process the whole dataset. Therefore, the images have been downsampled in time and only one image out of ten has been processed, corresponding to one image every 2 seconds. The processed Kinect v2 dataset was composed by 218 couples of corresponding RGB and depth images, for a total length of about 50 m. The data acquired with the u-blox antenna have been processed by using the open source position software goGPS (Realini and Reguzzoni, 2013, Herrera et al., 2015). However, we experienced some problems due to a malfunctioning of the antenna, which may be caused by the Kinect interference. Further

investigations are required to understand the actual causes of the problem. For this reason, in the Kalman filter we decided to integrate the data acquired with the double frequency receiver in RTK mode. With the aim to simulate data acquired by a low-cost GNSS antenna, white noise with a standard deviation equal to 0.30 m has been added to the RTK solution and the error covariance matrices have been modified accordingly. The trajectory recovered by using the Kinect does not include any Ground Control Points, so it is a relative solution. In order to compare and integrate the data acquired with different sensors, the rototranslation that minimizes the discrepancies between the Kinect v2 solution and the GNSS one has been estimated, still assuming that the GNSS antenna phase centre and the Kinect are located in the same point.

4.5 Preliminary results

The computed GNSS and Kinect v2 solution has been used as input for the Kalman filter, considering the stochastic observation model previously introduced. The RTK-GNSS observations were supplied with their covariance matrices, with a mean accuracy of the order of 0.02-0.03 m. Both the RTK solution and the covariance matrices have been degraded adding a white noise with a standard deviation of 0.30 m. The covariance matrices of the Kinect displacements have been empirically built by rescaling a quality parameter representative of the goodness of the RGB-D alignment so to have a standard deviation of the order of few centimetres, which is the expected accuracy of Kinect v2 outdoor depth measurements (see Figure 2). The dynamics error standard deviations have been empirically tuned, differentiating between straights and corners. They have been defined considering that the average variation of the mean velocity is of the order of 0.1 m/s; in particular, along straights it has been considered equal to 10% of this value, while along corners the dynamics constraint has been relaxed considering a standard deviation equal to 50% of the average velocity variation. In both cases a stronger constraint has been imposed along the height direction, because the surveyed area was practically flat. In Figure 8 the computed solutions, together with the reference trajectory acquired with the multistation, are shown. As underlined before, the Kinect v2 solution is a relative one, so some additional information is required to georeference it. After applying the rototranslation there is a good agreement between the horizontal components of the Kinect v2 solution and the reference one (see Figure 8a), meaning that the Microsoft device could represent an interesting solution for outdoor navigation. However, in the proposed method the trajectory is computed incrementally, by adding new displacements to the previously computed solution. For this reason, a drift is accumulated over time. In the case study, the path was quite short, but we experienced anyway a loop closure problem of about 1 m. One can easily imagine that in case of longer paths this problem would be more relevant. In this sense, it is quite important to integrate the Kinect solution with other kinds of information. From Figure 8 it is quite evident how the filtered solution allows correcting this problem. This effect is more evident in height, considering the part of the trajectory following the first corner (see Figure 8b), where the average distance from the reference dataset decreases from about 0.5 m to less than 0.1 m. The standard deviations between the computed solutions and the reference trajectory are shown in Table 5. These values can be considered valid under some approximations. Firstly, we assumed that all the instruments were located in the centre of mass of the vehicle. This was a simplification introduced to avoid considering rotations. Furthermore, we did not consider the lever arm between the reflective prism and the Kinect RGB camera projection centre, so there is a residual effect that was not modelled in the estimated rototranslation.

The Kalman filtered solution was able to better follow the reference trajectory; in fact, it is less disperse, if compared to both GNSS and the Kinect solutions. The pure photogrammetric solution presents standard deviations in the order of 0.50 m for horizontal coordinates and 0.20 for the height, while for the GNSS the standard deviations are in the order of 0.30 m in all the direction, in accordance to the imposed white noise. The Kalman filtered solution shows an improvement in all directions, reaching precisions of the order of 0.20 m for the horizontal directions and of less than 0.10 m for the vertical one.

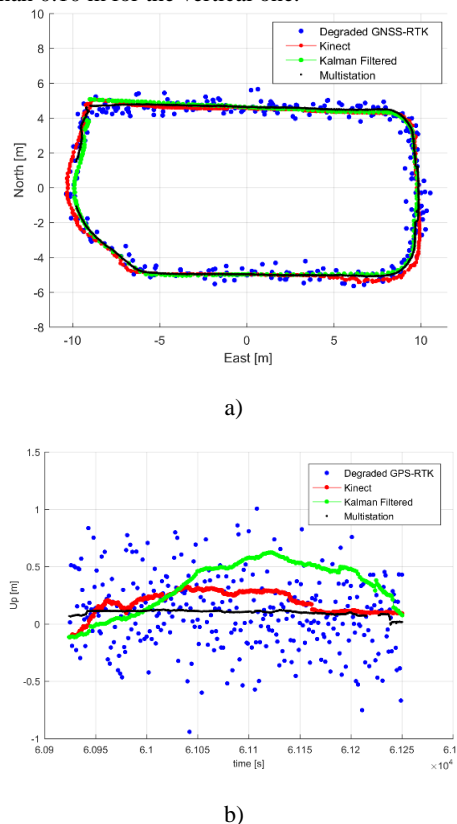


Figure 8. The computed solutions: a) planimetric view of the estimated trajectories – b) height behaviour of the estimated trajectory over time

Solution	North [m]	East [m]	Up [m]
GNSS	0.280	0.307	0.323
Kinect	0.524	0.545	0.206
Kalman filtered	0.183	0.195	0.067

Table 5. Standard deviations between the computed solutions and the reference trajectory measured with the multistation

The proposed solution is meant to be useful in situations in which it is not possible to rely on a good GNSS solution alone, for example in urban areas one can easily imagine to be unable to compute a solution using satellite positioning because of the presence of obstacles such as trees, buildings, colonnade, etc. For this reason we simulated the loss of the GNSS signal along the path, disregarding the GNSS estimated positions, for a portion of the second straight. In Figure 9 the corresponding filtered solution is shown. Thanks to the integration of the Kinect solution and the imposed vehicle dynamics, the trajectory has been well recovered even without the help of the GNSS observations. A third scenario has been studied by downsampling the GNSS observations, namely by considering a GNSS observation every 10 seconds. In Figure 10 the estimated

trajectory is shown, while in Table 6 the corresponding statistics are reported. From the results of this scenario, it is quite evident that the model used to describe the vehicle dynamics is too simple and without a proper number of GNSS observations the proposed Kalman filtered solution is not able to recover the correct trajectory after the corners.

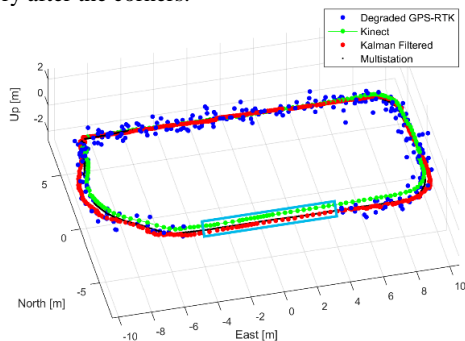


Figure 9. The computed solutions simulating the absence of GNSS signal along the second straight, indicated by the light blue rectangle

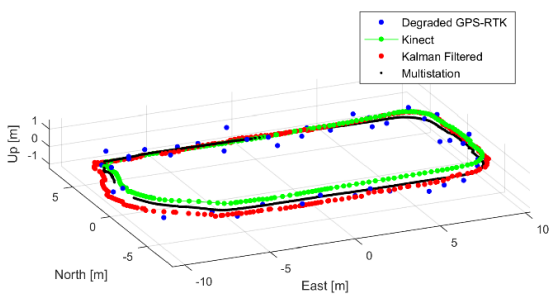


Figure 10. The computed solutions simulating a 0.1 Hz frequency for the GNSS observations

Solution	North [m]	East [m]	Up [m]
GNSS	0.314	0.284	0.349
Kinect	0.524	0.545	0.206
Kalman filtered	0.467	0.459	0.183

Table 6. Standard deviations between the computed solutions for the downsampled scenario and the reference trajectory measured with the multistation

The discrepancies between the purely photogrammetric trajectory and the filtered one are of the same order of magnitude, meaning that the proposed Kalman filter is not able to significantly correct the drift error accumulated in the Kinect solution.

5. CONCLUSIONS

In this paper, the use of the new generation of the Kinect device has been investigated for outdoor navigation, with the aim to create an integrated low-cost GNSS and photogrammetric navigation solution. The new Kinect has been advertised as capable of acquiring data even outdoor, under direct sunlight, so it is important to evaluate the expected accuracies and possible sources of error. The depth measurement accuracy of the Kinect v2 sensor has been evaluated outdoor, considering the depth response on a shaded planar surface. From the performed calibration procedure it comes out that Kinect v2 can acquire depth information with the same level of accuracy that characterizes indoor measurements. However, the data acquired outdoor are noisier. In order to evaluate the effect of the reflective

properties of the emitted signal on the different surfaces a panel with square samples of 0.2 m of different materials has been realized. The point cloud generated from the Kinect has been compared with the reference scan acquired with a Leica MS60 multistation. The procedure has been repeated indoor and outdoor (in shaded conditions) obtaining in both cases differences of the order of 0.01 m, for all the considered materials. From this analysis, it comes out how the Kinect v2 can be used indifferently indoor or outdoor in shaded conditions. Of course, reflective surfaces can cause troubles in both cases. Further investigations are required when the framed object is exposed to direct sunlight radiation.

Starting from these considerations, a solution for outdoor navigation has been proposed. In particular, a Kalman filter solution has been presented for integrating the Kinect trajectory (which is in turn recovered integrating depth and visual images) with the positions acquired with a low-cost GNSS receiver, discussing the different assumptions that have been done and the used mathematical model. This model has been realized under a number of simplified hypotheses (i.e. both the GNSS antenna and the Kinect device are assumed to be located in the centre of mass of the vehicle, the motion between subsequent time epochs is a pure translation, no lever arm has been considered between the different instruments), however the obtained results are promising. In fact, the Kalman filtered solution can correct some systematic errors present in the purely photogrammetric trajectory recovered from the Kinect data, especially in the Up direction. Moreover, the filtered solution allows correcting the loop-closure problem, which is a well-known issue for trajectory computed incrementally. We were able to obtain standard deviations of the order of 0.20 m for the horizontal coordinates and below 0.10 m in Up, comparing the filtered trajectory with the reference one acquired with the multistation. These results are comparable with the level of approximation that has been considered during the mathematical formalization of the problem; however, we expect to improve them by realizing a new Kalman filter, introducing a more complex dynamic model that for instance also takes into account a planar rotation of the Kinect itself. The RGB-D images obtained by the Microsoft device seems to be a suitable combined data for outdoor navigation. In fact, the Kinect v2 and the low-cost GNSS receiver are two sensors that can compensate their reciprocal weaknesses. The Kinect v2 can be useful to recover a very smooth trajectory in areas where the GNSS signal reception is difficult. For instance, a GNSS solution in urban areas is typically noisy and could be difficult to have a good positioning when getting closer to possible obstacles. This is exactly the field of application of Kinect v2, which can acquire depth information only at distances below 4 m. On the other hand, the Kinect needs external information, like the ones provided by GNSS, to be georeferenced. A remark is necessary for the different reference systems; the problem of the transformation between them has been empirically solved, namely that a best-fit solution has been used to estimate the rototranslation solution among the different trajectories. A more rigorous approach would require an a priori measurement of the lever arm between the Kinect and the GNSS antenna, e.g. using a calibration polygon to determine a set of double points.

In the proposed photogrammetric solution, the recovered trajectory is computed incrementally so a drift is accumulated along the path. This effect was only partially visible in the presented preliminary test because the followed path was not so long (about 50 m); however, we can expect that it will increase in case of longest trajectories. In that case, the usefulness of the GNSS-Kinect integrated solution is expected to be more evident. We are planning to repeat the experiment covering more loops. Moreover, we want to investigate the behaviour of the Kinect v2

during fully sunny days, but also during transitions from different light conditions coupled with the completely loss of the GNSS signal (e.g. passing from outdoor to indoor or going under a colonnade).

REFERENCES

- Alnowami, M., Alnowami, B., Tahavori, F., Copland, F., Wells, K., 2012. A quantitative assessment of using the Kinect for Xbox360 for respiratory surface motion tracking. *Proc. SPIE 8316, Medical Imaging*.
- Brovelli, M.A., Realini, E., Reguzzoni, M., Visconti, M.G., 2008. Comparison of the performance of medium and low level GNSS apparatus, with and without reference networks. In: *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVI, part 5/C55, pp. 54-61.
- El-laithy, R.A., Jidong, H., Yeh, M., 2012. Study on the use of Microsoft Kinect for robotics applications. *2012 IEEE/ION Position Location and Navigation Symposium*, pp. 1280-1288.
- Endres, F., Hess, J., Engelhard, N., Sturn, J., Cremer, D., Burgard, W., 2012. An Evaluation of RGB-D SLAM System. *Robotics and Automation, 2012 IEEE International Conference on*, River Centre, Saint Paul, Minnesota, pp. 1691-1696.
- Fischer, M.A., Bolles, R. C., 1981. Random Samples Consensus: A paradigm for model fitting with application to image analysis and application cartography. *Communication of the ACM*, pp. 381-395.
- Fankhauser, P., Bloesch, M., Rodriguez, D., Kaestner, E., Hutter M., Siegwart, R., 2015. Kinect v2 for mobile robot navigation: Evaluation and modeling. *Advanced Robotics (ICAR), 2015 International Conference on*, Istanbul, 2015, pp. 388-394.
- Han, J., Shoo, L., Xu, D., Shotton, J., 2013. Enhanced Computer Vision with Microsoft Kinect Sensor: a Review. *IEEE Transaction on Cybernetics*, Vol. 43, n°5, pp. 1318-1334, October 2013.
- Heikkila, J., Silven, O., 1997. A four-step Camera Calibration Procedure with Implicit Image Correction. *IEEE International Conference on Computer Vision and Pattern Recognition*. pp. 1106-1112
- Herrera, A.M., Suhandri, H.F., Realini, E., Reguzzoni, M., de Lacy M.C., 2015. goGPS: open-source MATLAB software. *GPS Solutions*.
- Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A., 2011. Kinect Fusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. *Proc. ACM Symposium on User Interface Software and Technology (UIST'11)*.
- Kalman, R. E., 1960. A New Approach to Linear Filtering and Prediction Problems. *Transaction of the ASME—Journal of Basic Engineering*, pp. 35-45.
- Khoshelham, K., 2011. Accuracy analysis of Kinect depth data. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, Vol. 38(5/W12).
- Lachat, E., Macher, H., Mittel M.-A., Landes, T., Grussenmeyer, P., 2015. First experiences with Kinect v2 sensor for close range 3D modelling. *6th International Workshop 3D-ARCH 2015*.
- Lowe, D., 2004. *Distinctive Image Feature from Scale-Invariant*. *International Journal of Computer Vision*, 60(2), pp. 91-110.
- Oliver, A., Kong, S., Wünsche, B., MacDonald, B., 2012. Using the Kinect as a Navigation Sensor for Mobile Robotics. *Proceedings of the 27th Conference on Image and Vision Computing (IVCNZ'12)*, pp. 505-514.
- Omara, H.I.M.A., Sahari, K.S.M., 2015. Indoor mapping using kinect and ROS. *Agents, Multi-Agent Systems and Robotics (ISAMSR), 2015 International Symposium on*, Putrajaya, 2015, pp. 110-116.
- Pagliari, D., Pinto, L., 2015. Calibration of Kinect for Xbox One and Comparison between the Two Generations of Microsoft Sensors. *Sensors*, 15, 27569-27589.
- Qayyum, U., Kim, J., 2013. Inertial-Kinect Fusion for Outdoor 3D Navigation. *Australasian Conference on Robotics and Automation (ACRA 2013)*, Sydney, Australia, p. 8.
- Realini, E., Reguzzoni, M., 2013. goGPS: open source software for enhancing the accuracy of low-cost receivers by single-frequency relative kinematic positioning. *Measurement Science and Technology*, 24 (11), 115010.
- Rossi, L., Sampietro, D., Sansò, F., 2012. Geonet: un software per la compensazione di reti topografiche integrate. *Proc. of the 16th National Conference ASITA*, 6-9 novembre, Vicenza, p.1157-1164.
- Samoil, S., Lai, K., Yanushkevich, S.N., 2014. Multispectral Hand Biometrics. *Proceedings of the 2104 - 5th International Conference on Engineering Security Technologies*, pp. 24-29.
- Schindhelm, C.K., 2012. Evaluating SLAM approach for Microsoft Kinect. *Proceedings of ICWMC 2012- 8th International Conference on Wireless and Mobile Communications*.
- Sell, J., O'Connor, P., 2014. The Xbox One System in a Chip and Kinect Sensor. *Micro IEEE*, vol. 34, issue n°2, pp. 44-53, March-April, 2014.
- Suarez, J., Murphy, R. R., 2012. Using the Kinect for search and rescue robotics. *Safety, Security, and Rescue Robotics (SSRR), 2012 IEEE International Symposium on*, College Station, TX, 2012, pp. 1-2.
- Xiao, J., Owens, A., Torralba, A., 2013. SUN3D: A database of Big Spaces Reconstructed using SfM and Object Labels, ICCV.