2012 12th International Conference on Control, Automation, Robotics & Vision
Guangzhou, China, 5-7th December 2012 (ICARCV 2012)

Experimental Comparison of 3D Vision Sensors for Mobile Robot Localization for Industrial Application: Stereo–Camera and RGB–D Sensor

Lorenzo Sabattini, Alessio Levratti, Francesco Venturi, Enrica Amplo, Cesare Fantuzzi and Cristian Secchi
Department of Engineering Sciences and Methods
University of Modena and Reggio Emilia, Italy
Email: {lorenzo.sabattini, alessio.levratti, francesco.venturi, cesare.fantuzzi, cristian.secchi}@unimore.it

***Abstract*—While RGB–D sensors are becoming more and more popular in mobile robotics laboratories, they are usually not yet adopted for industrial applications. In fact, in this field, depth measurements are generally acquired by means of laser scanners and, when visual information is needed, by means of stereo–cameras. The aim of this paper is to perform an experimental validation, to compare the performance of a stereo–camera and an RGB–D sensor, in a specific application: mobile robot localization for industrial applications. Experiments are performed exploiting artificial landmarks (defined by a self–similar pattern), placed in known positions in the environment.**

I. Introduction

This paper aims at providing experimental comparison of 3D vision technologies in the field of mobile robot localization in industrial applications.

Mobile robots are nowadays widely adopted in manufacturing plants. A remarkable example is represented by AGVs (Automated Guided Vehicles) used for end–of–line activities [1]–[3]. As for most of mobile robots applications, localization is a crucial issue: robots need to be able to measure their current position, in order to perform their tasks.

Global localization is typically solved, in industrial environment, by means of laser trilateration: each robot is equipped with a laser scanner, which measures the distance with respect to opportunely placed reflectors, that are used as landmarks. The position of each landmark is stored in a map, so that trilateration can be used by the robots to compute their global position. The main drawback of this technology is that, in order to provide reliable localization in a cluttered environment (which is typically the case in a manufacturing plant) hundreds to thousands of landmarks are required [3]. This fact makes the installation of this kind of systems very expensive and intrusive.

To avoid this issue, natural landmarks could be used, instead of artificial ones. Natural landmarks [4] are features of the environment that are easily recognizable by means of sensors. If the position of these features is measured with respect to the map of the environment, they can be used for mobile robots localization. In [4] and references therein, laser scanners are used to identify natural landmarks: however, when moving in a manufacturing plant, robots need to cope with highly cluttered, wide environments, with the presence of several symmetries. Hence, vision sensors appear necessary to acquire sufficiently reach data.

More specifically, 3D vision sensors provide both visual and distance information: hence, they may be adopted to identify natural landmarks while measuring their relative position. On these lines, some pioneering works may be found in the literature [5], where 3D vision based localization algorithm are validated in simplified simulated frameworks.

Stereo–cameras has been adopted, since the 1950s, to acquire three–dimensional images. They are cameras with two (or more) lenses, with a separate image sensor for each lens. This allows the camera to simulate human binocular vision, and therefore gives it the ability to capture three–dimensional images. This is obtained comparing the two images: as the distance between the two lenses is known, the so called *disparity image* may be computed, from which depth may be determined [6]. Moreover, depth information can be obtained combining vision with laser, or structured light. On these line, in the last few years, RGB–D sensors are becoming more and more popular: these sensors exploit structured light to acquire both visual (RGB) and deph (D) information about the environment [7]. However, depth measurement performance is considerably lower, if compared to a traditional laser scanner [8].

Visual localization have been widely adopted, in the last few years, for mobile robotics applications, as described, for instance, in the survey paper [9]. In [10] stereo vision is used for localization and SLAM (Simultaneous Localization and Mapping), using a stereo–camera that acquires the position of known landmarks, in indoor environment. Conversely, the work in [11] presents an outdoor visual localization application, that exploits an omni–directional camera to acquire the relative positions of some features of the environment, namely edges. Stereo–cameras are exploited in [12] for rough localization in small agricultural fields.

Due to their relatively low cost, RDB–D sensors have been extensively adopted, in the last few years, in the robotics research community. These sensors are generally used in laboratory scale experiments, for visual odometry [13], detection of humans [14], object recognition [15], obstacle avoidance [16] and localization [17].

978-1-4673-1872-3/12/$31.00 ©2012 IEEE

To the best of the authors' knowledge, RGB–D sensors are not usually employed in industrial environment. A first comparison between stereo–cameras and RGB–D sensors may be found in the literature in [18], where the performance of the two sensors is compared in a domestic environment, for elderly people monitoring.

Then, the main objective of this paper is to provide an experimental evaluation of the performance of two different classes of 3D vision sensors, regarding the localization application: a stereo–camera, and an RGB–D sensor. Conversely, feature recognition is out of the scope of this paper, and will be addressed subsequently. Hence, we introduce visual artificial landmarks that, given a specific pattern, are easily recognizable in the environment. Localization is then obtained with respect to the landmarks themselves, whose positions are previously recorded with respect to the map of the environment [19]. This allows us to focus our attention on sensors' performance, without being influenced by specific issues, related to object recognition algorithm.

The contribution of the paper may then be summarized as follows:

1) We implement a localization algorithm that exploits Extended Kalman Filter for data fusion of odometry with 3D visual information. Specifically, 3D vision sensors are used to compute the relative positions of self–similar landmarks, that were previously exploited in the literature for navigation purposes.
2) We experimentally compare the accuracy of the relative position measurements provided by a stereo–camera and by an RGB–D sensor.
3) We experimentally evaluate the performance of the visual localization system, comparing the results obtained with the stereo–camera and with the RGB–D sensor.

The outline of the paper is as follows. Section II describes the adopted 3D vision sensors, and provides the main characteristics of the landmarks used for localization. In Section III we experimentally compute the covariance matrices associated to the measurements provided by the sensors. Section IV describes the the localization algorithm based on Extended Kalman Filter. Experiments are then described in Section V. Finally, Section VI gives some concluding remarks.

II. Description of the Adopted Technology

A. 3D vision sensors

As stated in the Introduction, we are interested in comparing the performances of two different technologies to acquire 3D vision data: stereo–camera and RGB–D sensor. In this section we briefly describe the sensors employed in our experimental setup.

1) Microsoft Kinect RGB–D sensor:

Microsoft Kinect is a low cost 3D vision sensor developed for video–game purpose. The Kinect sensor [20] is a horizontal bar connected to a small base with a motorized pivot (Fig. 1). The device features an RGB camera, coupled with a depth sensor, that is an infrared projector (infrared light emitter) combined with a monochrome CMOS sensor (infrared detector), which captures video data in three dimensions under varying ambient light conditions. The acquired image has a resolution of 640×480 pixel. Visible and depth information can be acquired at a maximum rate of $30Hz$.

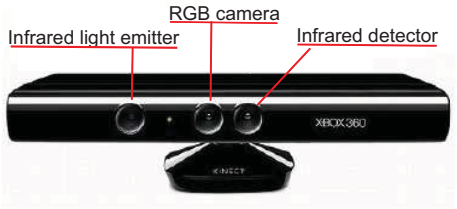


Fig. 1: Microsoft Kinect RGB–D sensor

As Kinect is intended for video–game purpose, it is a very low cost sensor: its price ranges 120.00 to 160.00 euro.

2) PointGrey Bumblebee2 Stereo–camera:

PointGrey Bumblebee2 [21] is a FireWire (IEEE–1394) stereo–camera, which is equipped with two Sony $1/3$" progressive scan color CCD sensors. The image may be acquired at a resolution of 1024×768 pixel with a rate of $20Hz$, or with a resolution of 640×480 pixel with a rate of $48Hz$. The

Fig. 2: PointGrey Bumblebee2 stereo–camera

camera is pre–calibrated against distortion and misalignment (RMS calibration error is guaranteed to be within 0.1 pixel). As Bumblebee2 stereo–camera is intended for professional use, it is considerably more expensive then Kinect: in fact, its price is around 2300.00 euro.

B. Self–similar landmarks

Exploiting information acquired by means of the 3D vision sensors, the robot is supposed to autonomously localize itself with respect to a map. More specifically, artificial landmarks are placed in the environment, and their positions are stored, with respect to the global map.

Hence, the robot is required to identify the landmarks, to measure their relative position, and subsequently to localize itself.

As stated in the Introduction, the main purpose of replacing laser scanners with vision sensors is the possibility to exploit the so called natural landmark: avoiding the need for artificial landmarks would significantly reduce the installation costs for mobile robotic systems in industrial applications. However, unique identification of natural landmarks in dynamic environments is a hard problem. Conversely, for localization purposes,

exact identification and measurement of the relative position of the landmarks is critical. Hence, as the purpose of this paper is to compare the performance of two different kind of 3D vision sensors, we consider a simplified localization framework: specifically, we introduce artificial landmarks that are reliably identifiable.

In particular, we exploit self–similar landmarks, introduced in [22] for navigation purposes. An example of self–similar landmark is represented in Fig. 3. More specifically, these

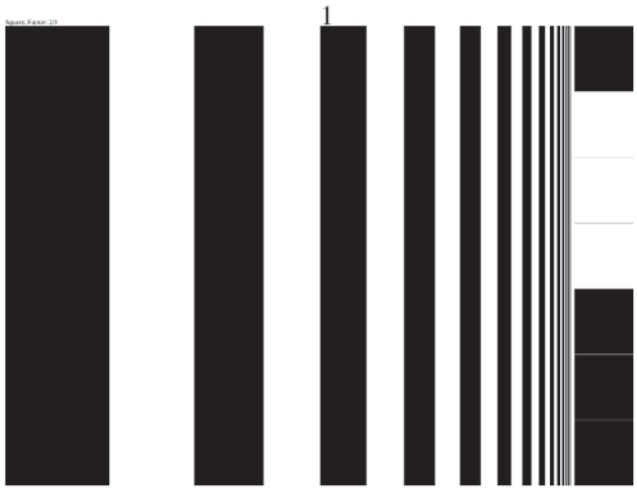


Fig. 3: Landmarks used for localization: self–similar pattern on the left–hand side, binary unique identifier on the right–hand side

landmarks are defined by means of self–similar intensity patterns. A function $f : \mathbb{R}^+ \mapsto \mathbb{R}$ is *p–similar* for the scale factor p, where $0 < p < 1$, if $f(x) = f(px)$, $\forall x > 0$. Hence, a p–similar function is identical to itself scaled by p in the x direction. Consequently, a p–similar pattern can be detected in the environment, by comparing the observed intensity pattern to a version of itself scaled by p. This characteristic makes the identification of this kind of landmarks very reliable [22].

As shown in Fig. 3, the left–hand side of the landmark represents a self–similar pattern, while the right–hand side of the landmark encodes the unique identifier of each landmark as a binary code. The position of each landmark (with respect to the map) is then stored, as well as its unique identifier. As will be explained in the next sections, these data will be exploited for localization purpose, acquiring the relative positions of the landmarks with the 3D vision sensors.

III. Experimental characterization of the 3D vision sensors

In order to evaluate the performance of a localization algorithm based on 3D vision, it is necessary to analyze the quality of the measurements provided by the sensors.

The quality of a measurement can be quantified by means of the associated covariance matrix [23]. Generally speaking, in fact, a sensor provides inaccurate data: the covariance matrix is a positive semidefinite and symmetric matrix, that measures the squared expected deviation from the mean value of the acquired data.

The covariance matrix may be estimated performing repeated experiments, and analyzing the acquired data set. In this particular application, we are interested in measuring the relative position of an object, with respect to the current position of the sensor. As we are considering wheeled mobile robots moving in indoor environment, we are interested in computing two–dimensional relative positions, on a plane identified by the ground floor.

More specifically, let (x_s, y_s) be the sensor's reference frame, where the y_s axis is parallel to the sensor's optical axis. Then, experiments have been performed to evaluate the accuracy of the sensors in measuring the relative position of an object, along the x_s and y_s axes.

In particular, we measured the relative position of a landmark (described in Section II-B) with respect to both the adopted sensors. We experimentally verified that, with the RGB–D sensor's resolution (640×480 pixel) a landmark is correctly identifiable if its distance is between $1m$ and $3m$. Conversely, exploiting the full resolution of the stereo–camera (1024×768 pixel) landmarks are identifiable up to $4m$.

As the aim of the experiments is to compare the performance of two different technologies, it is necessary to use the same conditions in both cases: specifically, we adopted the same resolution for both sensors, namely 640×480 pixel. Thus, the experiments were performed as follows:

- The measurement errors along the x_s and y_s axes have been measured independently:
  - the y_s coordinate of the landmark has been varied from $1m$ to $3m$, while keeping $x_s = 0m$.
  - the x_s coordinate of the landmark has been varied from $-1.25m$ to $+1.25m$, while keeping $y_s = 2.5m$.
- We acquired 25 images for each axis and for each sensor.

The covariance matrices $\mathcal{Q} \in \mathbb{R}^{2\times 2}$ have been approximated as diagonal matrices. Namely:

$$\mathcal{Q} = \begin{bmatrix} \sigma^2_{x_s} & 0 \\ 0 & \sigma^2_{y_s} \end{bmatrix} \tag{1}$$

where $\sigma^2_{x_s}$ and $\sigma^2_{y_s}$ represent the variance of the measurements along the x_s and y_s axis, respectively. In fact, as described in [23], the off–diagonal elements are negligible if compared to the diagonal elements.

We obtained the following experimental results (all results are in m^2):

- The covariance matrix of the measurements acquired with the Bumblebee2 stereo–camera is the following:

$$\mathcal{Q}_b = \begin{bmatrix} 1.8 \times 10^{-3} & 0 \\ 0 & 3.9 \times 10^{-4} \end{bmatrix} \tag{2}$$

- The covariance matrix of the measurements acquired with the Kinect RGB–D sensor is the following:

$$\mathcal{Q}_k = \begin{bmatrix} 2.7 \times 10^{-3} & 0 \\ 0 & 6.3 \times 10^{-3} \end{bmatrix} \tag{3}$$

It is worth noting that these results are obtained using the same resolution for both sensors, that is 640×480 pixel. Hence, it is possible to conclude that the stereo–camera provides more reliable relative position measurements: in fact, the values of the covariance matrix are significantly smaller, with respect to the ones computed for the RGB–D sensor.

IV. EKF–BASED LOCALIZATION USING SELF–SIMILAR LANDMARKS

We consider the problem of localizing a mobile robot operating in an indoor environment of known dimension, with the landmarks described in Section II-B scattered in it in known positions. The proposed localization method is based on the popular Extended Kalman Filter (EKF) Localization algorithm (see [23], [24]).

Algorithm 1 Extended Kalman Filter Localization

1: **Algorithm_EKF_Localization**$(\mu_{k-1}, \mathcal{S}_{k-1}, u_k, z_k, m)$
2: $\hat{\mu}_k = g\left(\mu_{k-1}, u_k\right)$
3: $\mathcal{G}_{k-1} = \left[\dfrac{\partial g_i}{\partial \mu_j}\right]$; $\mathcal{V}_k = \left[\dfrac{\partial g_i}{\partial u_j}\right]$
4: $\hat{\mathcal{S}}_k = \mathcal{G}_{k-1} \cdot \mathcal{S}_{k-1} \cdot \mathcal{G}_{k-1}^T + \mathcal{V}_k \cdot \mathcal{R}_k \cdot \mathcal{V}_k^T$
5: **for all** $z_{k,i}$ **do**
6: $\quad \mathcal{H}_k = \left[\dfrac{\partial h_i}{\partial \mu_j}\right]$; $\mathcal{P}_k = \left[\dfrac{\partial z_{k,i}}{\partial p_j}\right]$
7: $\quad \mathcal{K}_k = \hat{\mathcal{S}}_k \cdot \mathcal{H}_k^T \cdot \left(\mathcal{H}_k \cdot \hat{\mathcal{S}}_k \cdot \mathcal{H}_k^T + \mathcal{P}_k \cdot \mathcal{Q}_{x,y} \cdot \mathcal{P}_k^T\right)^{-1}$
8: $\quad \hat{\mu}_k = \hat{\mu}_k + \mathcal{K}_k \cdot \left[z_{k,i} - h\left(\hat{\mu}_k, m_i\right)\right]$
9: $\quad \hat{\mathcal{S}}_k = \left(\mathcal{I} - \mathcal{K}_k \cdot \mathcal{H}_k\right) \cdot \hat{\mathcal{S}}_k$
10: **end for**
11: $\mu_k = \hat{\mu}_k$; $\mathcal{S}_k = \hat{\mathcal{S}}_k$ **return** $\mu_k, \mathcal{S}_k$

The algorithm is recursive and the k–th step is discussed. Here we consider $\mu_k \in \mathbb{R}^3$ as the state vector representing the position and orientation of the robot, namely $\mu_k = [x, y, \vartheta]^T$.

Moreover, $u_k \in \mathbb{R}^2$ represents the input vector, namely $u_k = [\delta_R, \delta_L]$, with δ_R and δ_L the instantaneous displacement of the right and left wheels respectively.

The vector $z_{k,i} \in \mathbb{R}^2$ represents the measurement vector related to the i–th landmark, which has been detected by the on–board sensor, namely $z_{k,i} = [r_i, \alpha_i]^T$, where r_i and α_i are the range and bearing measurements, respectively.

Let $m_i = [x_m, y_m]^T$ be the position of the i–th landmark according to the reference frame of the map, we define m as the data structure containing the position information of the whole landmark set. The matrix $\mathcal{S}_k$ is the covariance matrix associated to the estimation uncertainty.

Line 2 of Algorithm 1 represents the computation of the time propagation of the system: we consider the motion model of the the differential drive wheeled mobile robot (refer to [25]):

$$g\left(\mu_{k-1}, u_k\right) = \begin{cases} x_k = x_{k-1} + \dfrac{\delta_R + \delta_L}{2} \cos\left(\vartheta_{k-1} + \dfrac{\delta_R - \delta_L}{2d}\right) \\ y_k = y_{k-1} + \dfrac{\delta_R + \delta_L}{2} \sin\left(\vartheta_{k-1} + \dfrac{\delta_R - \delta_L}{2d}\right) \\ \vartheta_k = \vartheta_{k-1} + \dfrac{\delta_R - \delta_L}{d} \end{cases} \tag{4}$$

where d is the distance between the two actuated wheels.

In Line 3 of Algorithm 1, the two jacobians of the time propagation function are computed, with respect to the state vector μ_k and the input vector u_k. Then, Line 4 computes the prediction of the covariance matrix, and $\mathcal{R}_k \in \mathbb{R}^{2\times 2}$ is the covariance associated to the input noise.

Lines 5 to 10 compute the update steps of the Extended Kalman Filter for each landmark which has been detected by the sensor. In particular, Line 6 computes the jacobian of the output function $h\left(\hat{\mu}_k, m_i\right)$, defined as follows:

$$h\left(\mu_k, m_{k,i}\right) = \begin{cases} \rho_{k,i} = \sqrt{\Delta x_{k,i}^2 + \Delta y_{k,i}^2} \\ \alpha_{k,i} = \begin{cases} \beta - \vartheta_k \text{ if } \beta > \vartheta_k \\ \vartheta_k - \beta \text{ if } \beta \leq \vartheta_k \end{cases} \end{cases} \tag{5}$$

where

$$\begin{aligned} \Delta x_{k,i} &= x_k - x_{m,i} \\ \Delta y_{k,i} &= y_k - y_{m,i} \end{aligned} \tag{6}$$

and

$$\beta = \operatorname{atan2}\left(\Delta y_{k,i}, \Delta x_{k,i}\right) \tag{7}$$

The vector μ_k represents the sate vector, and $p_k = [\Delta x, \Delta y]^T$.

Subsequently, Line 7 computes the so–called Kalman gain, $\mathcal{K}_k$, exploiting the covariance matrix associated to the measurement uncertainty, that is $\mathcal{Q}_{x,y} \in \mathbb{R}^{2\times 2}$. In particular, the term $\mathcal{P}_k \cdot \mathcal{Q}_{x,y} \cdot \mathcal{P}_k^T$ and the computation of the jacobian $\mathcal{P}_k$ are necessary because, as described in Section III, we characterized the sensors with respect to the (x_s, y_s) coordinates, and not according to the range and bearing considered in the output function. Then, in Lines 8 and 9, the estimation and covariance update are computed. Finally, in Line 11 the state and the covariance estimations are returned.

V. EXPERIMENTAL RESULTS

The localization algorithm described in Section IV has been implemented on an Adept MobileRobots Pioneer P3–DX robot, controlled by means of the software framework ROS (Robot Operating System) [26], [27].

A variable number of landmarks (from 1 to 5) have been placed in the environment, in known positions. The experimental setup is represented in Fig. 4.

In order to precisely evaluate the performance of the localization system, experiments have been implemented in static conditions. Namely, the objective of the algorithm is to compute the position of the robot under the following conditions:

- The robot's initial position is

$$\begin{cases} x_0 = 0; \\ y_0 = 0; \\ \theta_0 = \dfrac{\pi}{2} \end{cases} \tag{8}$$

- The robot does not move, but an error is added to the odometry measurements. Specifically:
  - at the first iteration, the odometry measurement provides the following data: $\delta_R = \delta_L = 0.50m$.
  - at the subsequent iteration, the odometry measurement provides the following data: $\delta_R = \delta_L = 0.0m$.

Fig. 4: Experimental setup

Hence, the EKF–based localization algorithm is required to compensate the error provided by the odometry, acquiring the landmarks' relative position with the 3D vision sensors.

This experimental test have been performed with the Kinect RGB–D sensor, and with the Bumblebee2 stereo–camera, in three different experimental setup, characterized by the presence of 1, 3 and 5 landmarks, respectively. Statistical results obtained during the experiments are collected in Figs. 5, 6 and 7. These plots represent the result of the EKF–based localization algorithm (using data acquired with both sensors): blue solid line represent the real value, the red dotted line represents the measurement computed with the Kinect, black dashed line represents the measurement computed with the Bumblebee2. Standard deviation is shown as well. As expected, the uncertainty of the measurements decreases, as the number of iteration grows. Moreover, adding more landmarks, it is possible to obtain a higher accuracy and a faster convergence. The fact that, as shown in Section III, the Bumblebee2 camera provides more accurate measurements than the Kinect generates higher precision in the localization. More specifically, as shown in Figs. 5, 6 and 7, the Bumblebee2 camera leads to a more accurate localization, in term of position (Figs. 5 and 6) and orientation (Fig. 7).

VI. Analysis of the results and conclusion

In this paper we performed several experimental tests to compare the performance of 3D vision sensors for mobile robot localization. In particular, we adopted two different sensors: a PointGrey Bumblebee2 stereo–camera, and a Microsoft Kinect RBG–D sensor.

A localization algorithm based on Extended Kalman Filter has been implemented, to combine odometry with the measurement of the relative positions of artificial landmarks, defined by a self–similar pattern.

As expected, the higher quality of the stereo–camera produces better localization performance, in terms of precision, with respect to the RGB–D sensor. However, it is necessary to perform further experiments, in order to better evaluate the robustness of both sensors in realistic industrial applications, that is in the presence of worst environmental conditions.

Moreover, future work will also aim at better characterizing the quality of the measurements of the sensors. In particular, we will investigate the presence of possible systematic errors, that might be filtered to enhance the obtained results. Furthermore, we will provide better estimates of the covariance matrices, possibly characterizing them with respect to range and bearing measurements (instead of measuring the variance along (x_s, y_s) axes).

Acknowledgment

This work was funded by Regione Emilia Romagna in the scope of the research project DiRò (Distretto della Robotica Mobile).

References

[1] P. R. Wurman, R. D'Andrea, and M. Mountz, "Coordinating hundreds of cooperative, autonomous vehicles in warehouses," in *IAAI'07 Proceedings of the 19th national conference on Innovative applications of artificial intelligence*, vol. 2, 2007.

[2] R. Olmi, C. Secchi, and C. Fantuzzi, "Coordination of multiple agvs in an industrial application," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, may 2008, pp. 1916–1921.

[3] D. Ronzoni, R. Olmi, C. Secchi, and C. Fantuzzi, "AGV global localization using indistinguishable artificial landmarks," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, may 2011, pp. 287–292.

[4] S. F. Hernandez–Alamilla and E. F. Morales, "Global localization of mobile robots for indoor environments using natural landmarks," in *Robotics, Automation and Mechatronics, 2006 IEEE Conference on*, dec. 2006, pp. 1–6.

[5] G. Spampinato, J. Lidholm, L. Asplund, and F. Ekstrand, "Stereo vision based navigation for automated vehicles in industry," in *Emerging Technologies Factory Automation, 2009. ETFA 2009. IEEE Conference on*, sept. 2009.

[6] O. Faugeras, *Three–Dimensional Computer Vision*. The MIT Press, 1993.

[7] K. Lai, L. Bo, X. Ren, and D. Fox, "A large–scale hierarchical multi–view RGB–D object dataset," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, may 2011, pp. 1817–1824.

[8] C. S. P. S. W. Kim, D. Kim, and S. R. Oh, "Comparison of plane extraction performance using laser scanner and kinect," in *Ubiquitous Robots and Ambient Intelligence (URAI), 2011 8th International Conference on*, nov. 2011, pp. 153–155.

[9] G. N. Desouza and A. C. Kak, "Vision for mobile robot navigation: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 2, pp. 237–267, feb 2002.

[10] S. Takezawa, T. Ishimoto, and G. Dissanayake, "Optimal control for simultaneous localisation and mapping problems in indoor environments with stereo vision," in *IEEE Industrial Electronics, IECON 2006 - 32nd Annual Conference on*, nov. 2006, pp. 4749–4754.

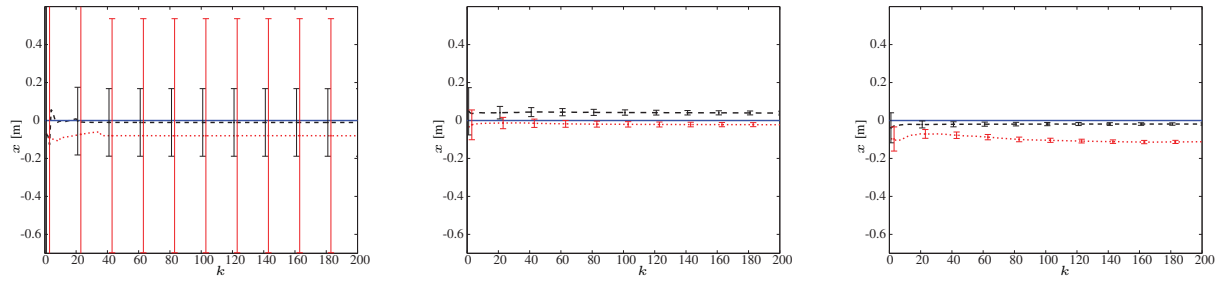
[11] S. Nuske, J. Roberts, and G. Wyeth, "Visual localisation in outdoor industrial building environments," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, may 2008, pp. 544–550.

[12] L. Piyathilaka and R. Munasinghe, "Vision–only outdoor localization of two–wheel tractor for autonomous operation in agricultural fields," in *Industrial and Information Systems (ICIIS), 2011 6th IEEE International Conference on*, aug. 2011, pp. 358–363.

[13] M. Fiala and A. Ufkes, "Visual odometry using 3–dimensional video input," in *Computer and Robot Vision (CRV), 2011 Canadian Conference on*, may 2011, pp. 86–93.

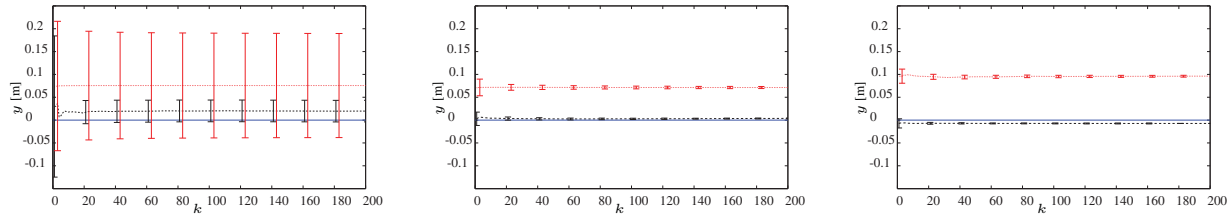
[14] L. Xia, C. C. Chen, and J. K. Aggarwal, "Human detection using depth information by kinect," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, june 2011, pp. 15–22.

[15] K. Lai, L. Bo, X. Ren, and D. Fox, "Sparse distance learning for object recognition combining rgb and depth information," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, may 2011, pp. 4007–4013.



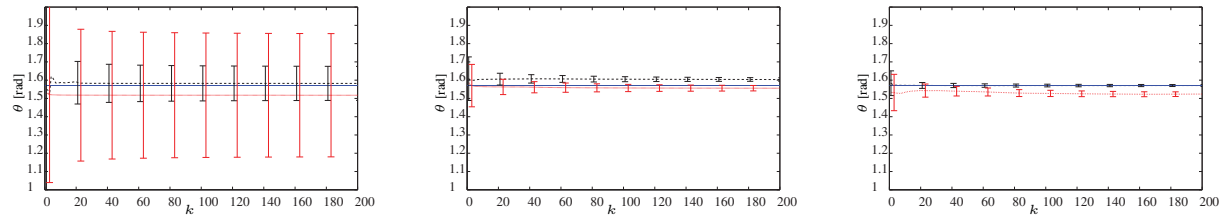
(a) 1 landmark (b) 3 landmarks (c) 5 landmarks

Fig. 5: Localization experiment in static condition: measurement of the x coordinate. Blue solid line represents the ground truth, red dotted line represents the measurement computed with the Kinect, black dashed line represents the measurement computed with the Bumblebee2



(a) 1 landmark (b) 3 landmarks (c) 5 landmarks

Fig. 6: Localization experiment in static condition: measurement of the y coordinate. Blue solid line represents the ground truth, red dotted line represents the measurement computed with the Kinect, black dashed line represents the measurement computed with the Bumblebee2



(a) 1 landmark (b) 3 landmarks (c) 5 landmarks

Fig. 7: Localization experiment in static condition: measurement of the orientation, θ. Blue solid line represents the ground truth, red dotted line represents the measurement computed with the Kinect, black dashed line represents the measurement computed with the Bumblebee2

[16] P. Benavidez and M. Jamshidi, "Mobile robot navigation and target tracking system," in *System of Systems Engineering (SoSE), 2011 6th International Conference on*, june 2011, pp. 299–304.

[17] N. Ganganath and H. Leung, "Mobile robot localization using odometry and kinect sensor," in *Emerging Signal Processing Applications (ESPA), 2012 IEEE International Conference on*, jan. 2012, pp. 91–94.

[18] E. E. Stone and M. Skubic, "Passive in–home measurement of stride–to–stride gait variability comparing vision and kinect sensing," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 6491–6494.

[19] X. Zhang and N. Navab, "Tracking and pose estimation for computer assisted localization in industrial environments," in *Applications of Computer Vision, 2000, Fifth IEEE Workshop on.*, 2000, pp. 214–221.

[20] "Microsoft kinect," http://www.xbox.com/en-US/Kinect.

[21] "Pointgrey bumblebee2," http://www.ptgrey.com/products/bumblebee2/.

[22] A. J. Briggs, D. Scharstein, D. Braziunas, C. Dima, and P. Wall, "Mobile robot navigation using self–similar landmarks," in *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, vol. 2, 2000, pp. 1428–1434 vol.2.

[23] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[24] G. Welch and G. Bishop, "An introduction to the kalman filter," University of North Carolina at Chaper Hill, 1995.

[25] G. Oriolo, A. De Luca, and M. Vendittelli, "WMR control via dynamic feedback linearization: design, implementation, and experimental validation," vol. 10, no. 6, nov 2002, pp. 835–852.

[26] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open–source robot operating system," in *ICRA Workshop on Open Source Software*, 2009.

[27] "ROS: Robot operating system," http://www.ros.org/.