Experimental Validation of a Visual Odometry System for Indoor Unstructured Environments

João Rodrigues
IDMEC-IST
Technical University of Lisbon
1049-001 Lisboa, Portugal
Email:
joao.l.rodrigues@ist.utl.pt

***Abstract*—The main goal of this paper is to present a visual odometry system, for localization of mobile robot in indoor unstructured environments, using only ceiling depth images, captured by a Kinect sensor. The use of odometric sensors is a common practice for localization of mobile robots. The method proposed in this work exploits information from an independent source of depth data and thus allows to complement or substitute the use of classic odometric sensors, like wheels encoders, with well known limitations. The experimental validation of the proposed solution shows that the method is able to accurately compute the attitude and linear velocities that allow a more precise mobile robot localization, even in presence of corrupted data from the sensor. Furthermore, the method works in an extended range of lighting conditions, without the need to perform any specific feature extraction.**

I. INTRODUCTION

The localization of mobile robots to navigate in indoor environments has been a great challenge to the scientific community in the area of mobile robotics [10], [3], [8]. GPS (Global Position System) allows to obtain the global position and attitude of a mobile robot with great precision. Actually GPS became the standard solution for outdoor environments. However, for indoor environments, or any environment where the GPS signal is not available, to solve the localization problem the creation of alternative approaches is required to obtain the global position of mobile robots [1], [18], [12], [7].

The use of Computer Vision techniques is a common practice to obtain information about the environment in robot localization, due to the large amount of information that can be extracted from a RGB image [17], [16], [9]. However, cameras are very sensitive to the environment lighting conditions, which has a negative impact in the robustness of localization systems. Recently, new devices with a RGB and a depth camera developed initially for video games, e.g Kinect from PrimeSense and Microsoft, became very popular in the mobile robotics community[4], [13].

Whatever the sensors implemented in the location system, its main function is to look at the environment and rapidly help the robot to answer two questions: where am I? and what am I facing? The global localization of a mobile robot is obtained by the fusion of the information about the environment, captured by sensors, and the knowledge about the robot motion given by the robot model. A Bayes filter is an appropriated technique to fuse these information [19], [8]. Usually, the prediction of the robot motion is performed with the odometry of the wheels, obtained by the encoders installed on the robot. Nevertheless, due to different causes like uncertainty in robot dimensions, misalignment of wheels or slippage in uneven terrain or other adverse conditions, the wheel odometry rapidly degrades the measure of the prediction motion, causing a negative impact on the final results.

As an alternative to wheel odometry, some localization systems have implemented cameras to predict the robot motion based on computer vision techniques. This class of methods, denominated visual odometry, allows the robot position determination, velocity and acceleration based on examination of the changes that motion induces on consecutive images captured by the cameras [15]. In addition, visual odometry can complement other sensors systems like GPS or inertial measurement units (IMU) [2], [14]. The visual odometry has been successfully applied in places where the GPS signal is not available [1] and will be the central focus of this paper.

Unlike most common localization systems that uses visual odometry with RGB images captured by cameras pointed to the ground [5] or looking around [2], the method presented in this paper resorts to depth images captured by a Kinect sensor, installed onboard of a mobile robot and pointed upwards to the ceiling. The use of vision from the ceiling has the advantage that images can be considered without scaling, i.e. a 2D image problem results. The use of ceiling vision in mobile robot navigation is successfully implemented in [11], [20], [7].

The method described in this paper aims to present a visual odometry system that is experimentally validated in a mobile robot, namely to extract the attitude and linear displacement that is integrated in a localization system for indoor navigation. Furthermore, the proposed method allows the operation in unstructured environments, i.e. without the need of any knowledge about the environment or the extraction of specific features.

This paper is organized as follows: Section II presents the mobile robot platform and the motivation for the use of visual odometry, instead of wheel odometry; Section III presents visual odometry in more detail; Section IV shows in detail the experimental results that allow the assessment of the approach for a number of high level tasks. Finally, Section V presents some conclusions and unveils future work.

II. MOBILE ROBOT PLATFORM

The type of mobile robot used in this tests is a low cost mobile robotic platform [6], with a differential drive configuration. On top of the mobile robot there is a PC

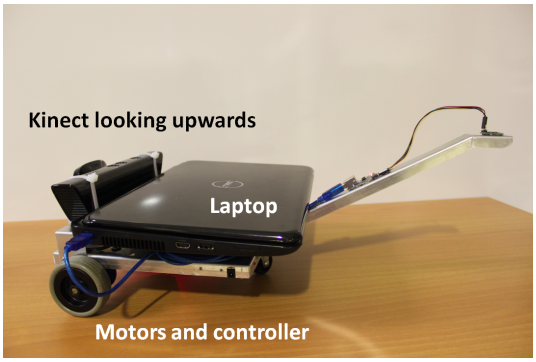

Fig. 1. Mobile platform equipped with Kinect sensor

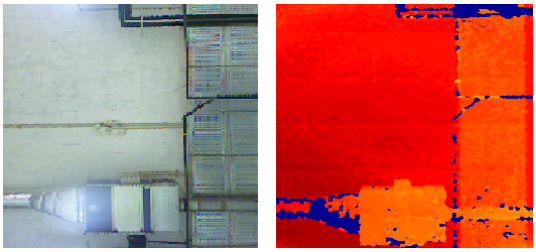

Fig. 2. RGB (left) and Depth (right) images captured by Kinect sensor

laptop that controls the motors and a Microsoft Kinect pointing upwards to the ceiling (see Fig. 1). On the right of the picture an extension with a magnetometer can be seen, to provide alternative attitude measurements, but not used in the scope of this paper.

The Microsoft Kinect is a bundle of sensors, which includes a RGB camera with a VGA resolution (640×480 pixels) using 8 bits and a 2D depth sensor (640×480 pixels) with 11 bits of resolution. In the work reported in this paper, the robot is moving in an indoor environment under a ceiling with some information (e.g. building-related systems such as HVAC, electrical and security systems, etc.), which is used to provide the ceiling vision navigation. Thus, the captured RGB and depth images that can be captured by the Kinect sensor are shown in Fig. 2.

For the development of a localization method to work in places with illumination changes and to reduce the computational efforts, the proposed visual odometry localization system only uses the depth signal of the Kinect sensor.

III. VISUAL ODOMETRY SYSTEM

In robotics, visual odometry [15] is the process that predicts the motion based on consecutive images captured by cameras installed onboard of the robot and has the advantage to be more immune to wheel slippage than wheel odometry and does not suffer from magnetic distortion effects observed in magnetometers. In this section a visual odometry approach is detailed, to be used in robot localization in unstructured environments. The proposed method just requires a depth camera pointing to the ceiling and uses the captured depth information to compute the localization of the mobile robot, without the need of a previous mapping and any feature extraction. Thus, analyzing the general architecture of the

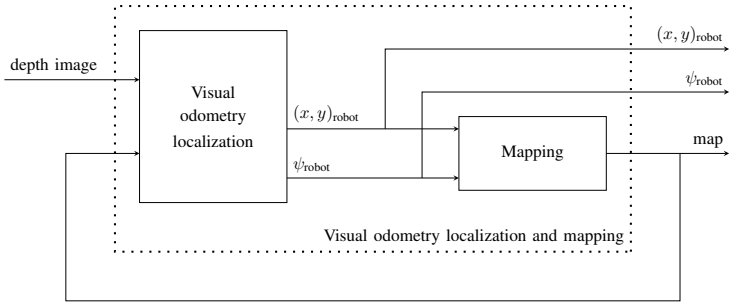

Fig. 3. General architecture of localization and mapping system

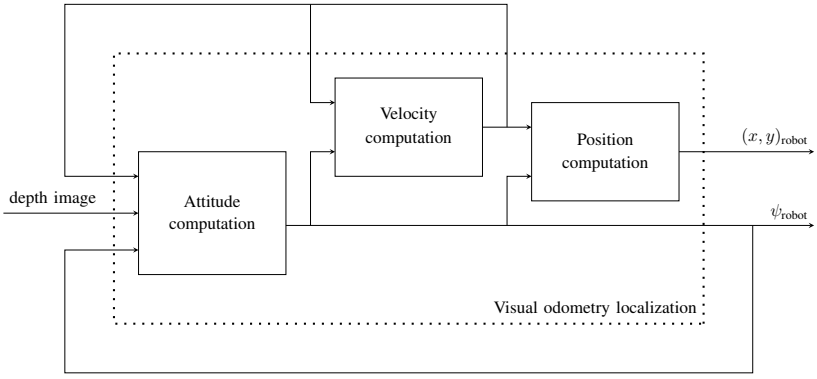

Fig. 4. Architecture of the visual odometry localization

proposed system (Fig. 3), the method consist in the mobile robot localization and the construction of the environment map based on depth images captured from the ceiling. The first step of the method is the definition of the new position in the map, which is performed with the depth image and the knowledge that the robot has about the environment and, in a second step, the new depth image is added to the map, increasing the database.

Looking to Fig. 4 that details the proposed visual odometry localization method, the localization of the robot is performed considering the computation of attitude and position of the mobile robot.

A. Attitude computation

Following the sequence of the algorithm, the value of the mobile robot attitude is obtained comparing the captured depth image with map testing possible turning angles of the robot. Thus, considering a depth image $\mathbf{dimage}(k)$ captured in instant k, a set of possible rotated images is created based on the robot attitude in the previous instant of time for $\psi_j \in [\psi(k-1) - \Delta\psi, \psi(k-1) + \Delta\psi]$, resulting in j rotated images

$$\mathbf{dimr}_j = imrotate(\mathbf{dimage}(k), \psi_j), j = 1, ..., M \qquad (1)$$

where M is the number of images to be analyzed, to be selected in the implementation phase.

Since the data is captured by a depth camera based on an infra-red grid, several waves are not well reflected, due to geometry and properties of some objects, resulting in missing

data in the depth image, represented by null value ($0\ mm$). Thus, to eliminate the possible disturbances caused by the missing data, the comparison between images is only calculated in pixels with non corrupted data, i.e. for values in the map and in the captured image with valid depth information.

The new robot attitude is computed, finding the angle ψ_j that minimizes the mean square error between the image stored in the map, in the previous position of the robot ($\mathbf{map}_{x,y}(k-1)$), and the rotation of the captured depth image (2)–(4),

$$\mathbf{m}_j = (\mathbf{map}_{x,y}(k-1) - \mathbf{dimr}_j)^2 \tag{2}$$

$$\bar{\mathbf{m}}_j = \frac{\sum_{i=1}^{N} \mathbf{m}_j}{nd} \tag{3}$$

where N is the number of pixels of map and nd, the number of pixels of $\mathbf{m}_j$ with depth information ($\mathbf{m}_j > 0\ mm$).

Finally, the robot attitude is obtained, finding the ψ_j that minimizes (3):

$$\psi(k) = \min_j(\bar{\mathbf{m}}_j). \tag{4}$$

B. Velocity computation

In a similar way, the velocity is computed by testing different values and finding the one that results in the best fit. Thus, considering a depth image $dimage$ captured in instant k and rotated by the obtained attitude previously mentioned in Section III-A, results in $\mathbf{dimt}$. A set of possible displaced images along the direction of $\psi(k)$ is created based on the velocity of the robot in the previous instant inside the range $u_j \in [u(k-1) - \Delta u, u(k-1) + \Delta u]$. Following the same process that lead to the attitude computation, the robot velocity value is obtained by the mean square error of the possible tested images (5)–(7).

$$\mathbf{mu}_j = (\mathbf{map}_{x,y}(k-1) - \mathbf{dimt}_j)^2, \tag{5}$$

where $\mathbf{dimt}_j$ is the image translated with the possible velocity u_j.

$$\bar{\mathbf{mu}}_j = \frac{\sum_{i=1}^{N} \mathbf{mu}_j}{nd} \tag{6}$$

Finally, the robot velocity is obtained, finding the u_j that minimizes (6):

$$u(k) = \min_j(\bar{\mathbf{mu}}_j). \tag{7}$$

C. Position computation

After the computation of the attitude and the velocity of the mobile robot based on the depth information, the robot kinematics is used to allow the computation of the new position, based on the well know Euler discretization of the differential drive robot:

$$x(k) = x(k-1) + u(k)T\cos(\psi(k)) \tag{8}$$

$$y(k) = y(k-1) + u(k)T\sin(\psi(k)) \tag{9}$$

where T is the sampling time.

D. Mapping

Mapping is crucial in mobile robot navigation because it improves the knowledge about the environment in future localization. Therefore, in the fourth part of the proposed method, the new captured depth image is added into the global map of the environment in the localization computed as described in Sections III-A, B and C. For experimental assessment purposes, a naive approach to map building was exploited in this phase of the work. Thus the addition of the new captured depth image in the map is performed replacing all null pixels existing in the global image by the pixels captured by the Kinect sensor. In this process only the non corrupted data of the captured depth image is considered. With this method, the created global map is immune to the corrupted data and only the rich information about the environment is stored.

IV. EXPERIMENTAL RESULTS

To test the proposed approach several tests have been performed with different trajectories, combining both straight lines and curves. In the experiences, the robot starts at $x_0 = 0$ m, $y_0 = 0$ m, $u_0 = \dot{y}_0 = 0.1\ \mathrm{m{\cdot}s^{-1}}$, $\dot{x}_0 = 0\ \mathrm{m{\cdot}s^{-1}}$. During the motion at constant speed along the predefined trajectory, the mobile robot captures depth images from the ceiling with 5 Hz of sampling rate and several points are marked on the ground to be measured afterwards the test end up . All experiences are performed under a ceiling height of $5.2\ m$ resulting in a depth images with resolution $7.8 \times 10^{-}3\ m/pixel$. Notice that mobile robot motion is aligned with the vertical axis of Kinect sensor, which has 480 captured pixels and a vision angle of $43°$. The attitude and velocity computation (green filled circle in Fig. 5) have been performed considering a range of $\Delta\psi = 7°$ and $\Delta u = 0.12\ m \cdot s^1$. In this process, the attitude step is $0.1°$ for $|\psi_j - \psi(k-1)| <= 0.5°$ and $2°$ for $|\psi_j - \psi(k-1)| > 0.5°$. The velocity step is $0.04\ m \cdot s^1$ in all range (see red circles in Fig. 5).

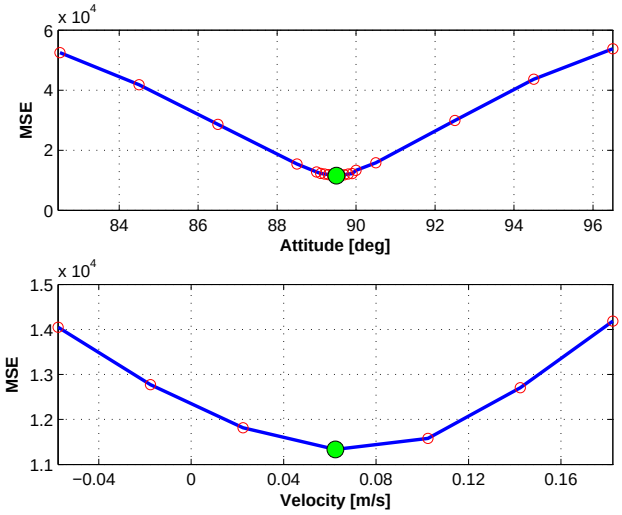

Fig. 5. Attitude and velocity computation in the instant $t = 2\ s$

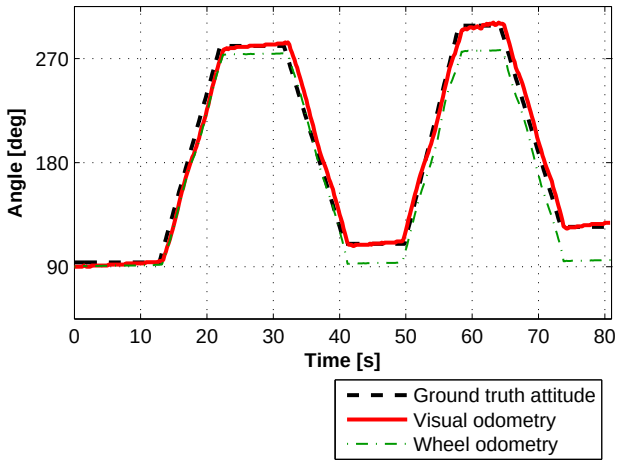

Fig. 6. Attitude of the robot along time

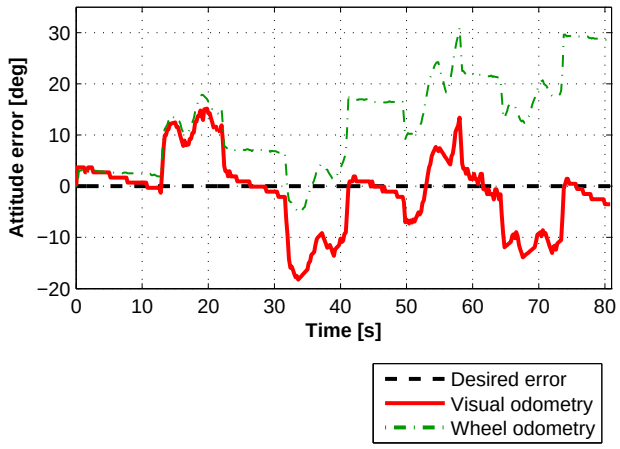

Fig. 7. Results for visual and wheel odometry errors along time

A. *Results for a lawnmower trajectory*

The first experience uses the classical lawnmower trajectory, which combines lines with curves, alternating the turning direction of the mobile robot and starting in with initial attitude of $\psi_0 = 90$ °. As it is possible to see in Fig. 6, the proposed method allows to estimate the attitude of the robot with accuracy, presenting results near of the measured ground truth attitude. Looking to Fig. 6, it is possible to observe that the results provided by the proposed method are indeed better than the usual wheel odometry.

Analyzing the attitude error considering visual and wheel odometry shown in Fig. 7 it is also possible to conclude that visual odometry results are more accurate than the attitude obtained by wheels odometry. In Fig. 7, it can be seen that the attitude obtained by visual odometry normally has an error less than 5 °, but sometimes the error increases to values over 10 °. Comparing the time of these occurrences with the attitude presented in Fig. 6, it is possible to concluded that visual odometry provide the better results when the robot moves in a straight line, decreasing the accuracy along the curve

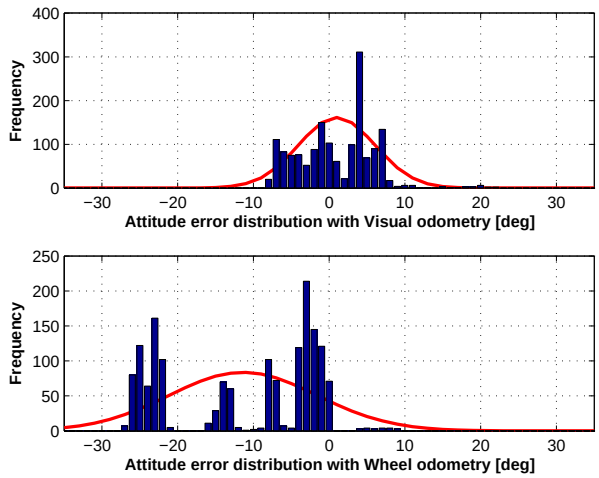

Fig. 8. Histogram of the attitude error

trajectories. The same effect happens with wheel odometry. However, the results of Fig. 6 show that visual odometry is able to recover the accuracy after finishing the curve, while wheel odometry is unable to do it, increasing the attitude error along time.

Moreover, observing Fig. 8 that shows the histogram of the attitude error of both odometry methods, it is also possible to conclude that visual odometry provides better results than wheel odometry. In this experience the visual odometry histogram is approximate to a zero mean Gaussian distribution, while wheel odometry presents a distribution with 13° of mean. The large value of the mean error, considering wheel odometry, denotes the existence of angular slippage in the motion. Notice that, in this experience, when the robot is navigating and mapping with a large image overlapping, visual odometry can compute an attitude value close of the real one, eliminating the offset caused by angular slippage.

Analyzing the results of the ground truth localization in Fig. 9, it is possible to observe that the trajectory obtained by visual odometry is able to follow the real one described by the mobile robot. Moreover, Fig. 9 shows that the position obtained by wheel odometry, due the wrong results of the attitude presented in Fig. 6, rapidly diverges from the ground truth trajectory, causing the robot to lose its own localization.

In addition to the estimation of the attitude, the proposed algorithm also computes the velocity of the robot. Although, due to the resolution of this estimator, imposed by the image translation in the grid, the results have lower accuracy than the odometer wheel (Fig. 10). The fact that the estimator provide results with approximately Gaussian error (Fig. 11), do not disturb the global robot position estimation. This can be verify by the results shown in Fig. 9, where the global position obtained by visual odometry are much more accurate than those obtained by the odometry wheel.

Finally, Fig. 12 shows the ceiling map computed along the trajectory. Comparing the depth image of Fig. 2, which shows the first data of this experience, it is possible to observe, that the knowledge about the environment is larger and well organized. Comparing the amount of missing data (blue color) present in map, after the first depth image acquisition (Fig. 2),

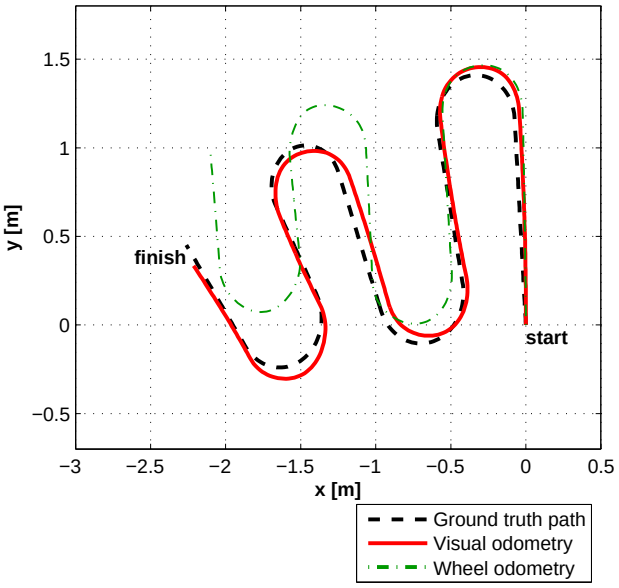

Fig. 9. Map with visual and wheel odometry position considering a ground truth path

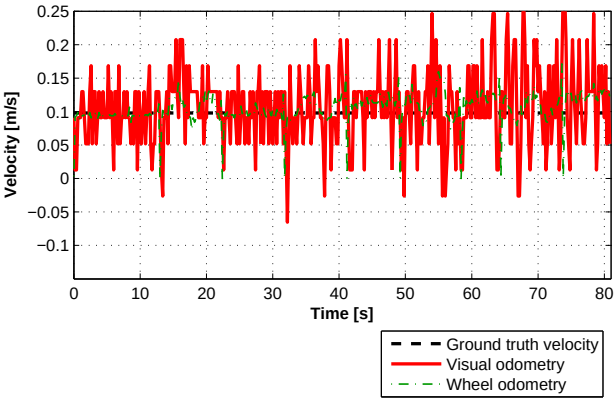

Fig. 10. Velocity of the robot along time

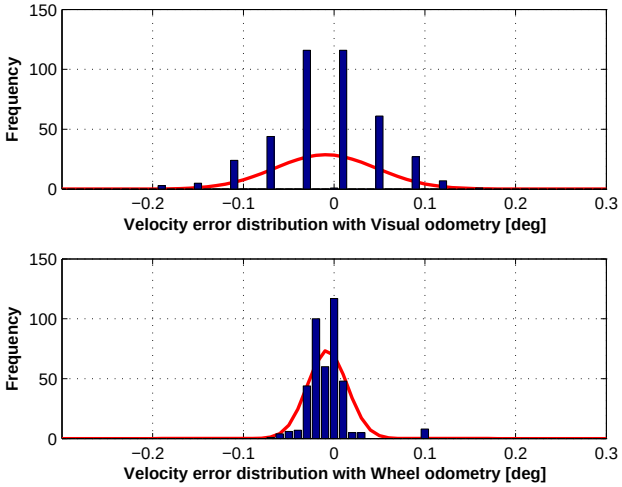

Fig. 11. Attitude of the robot along time

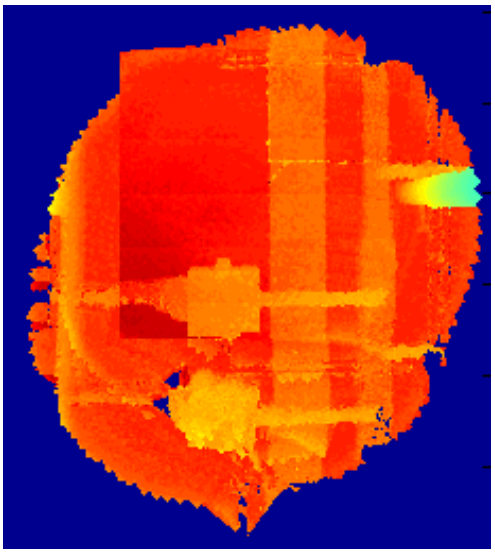

Fig. 12. Ceiling map built along the trajectory

is higher than the existed in the final map (Fig. 12). Notice that the blue area around the final map correspond to a non mapped area and not to missing data. This allows to concluded that the mapping method is able to build a ceiling map, reducing the amount of missing data existing in depth images.

B. Results for a longer trajectory

To test the robustness of visual odometry when the robot is moving in a longer straight line, a new experience has been performed. Thus, the robot has been moving along 32 m, capturing the depth ceiling images for its self-localization database and create the environment map (Fig. 13).

Analyzing the results presented in Fig. 14, it is possible to see that the attitude computed by the proposed visual odometry method provides results closer to the ground truth attitude than the computed by wheel odometry. This effect is more visible along time, what denotes that visual odometry presents a more robust method for attitude computation based on motion analyzing.

Examining the distribution of the attitude error in both odometry methods as depicted in Fig. 15, it is possible to observe that visual odometry is more accurate than wheel odometry, with an approximate zero mean Gaussian and low standard deviation ($\sigma^2 = 5^{\circ 2}$). On other hand, the distribution of wheel odometry attitude errors shows the few accuracy of this method, characterized by a large distributions with different modes and high values to attitude errors.

Comparing the results of the odometry localization obtained by wheel odometry with the visual odometry proposed in this paper (Fig. 16), it is possible to conclude that the proposed visual odometry provides a better localization than wheel odometry. However, results shows that, although this method can predict the mobile robot motion with more accuracy than wheel odometry, visual odometry does not have enough precision when the robot is navigating during a long

time in unmapped places.

Notice that, in this experience the robot is moving along 32 m, navigating 320 s (5 minutes and 20 seconds) with only one sensor in an unknown environment. This allows to conclude that, to develop localization systems able to navigate in an unknown environment, the proposed visual odometry method must be fused with other sensors. However, these results shows that the use of visual odometry can provide better motion prediction than wheel odometry. Moreover, this result is achieved without any previous knowledge about the environment.

V. CONCLUSION

In this paper a visual odometry localization system for mobile robots navigation in indoors unstructured environments is presented and experimentally validated. The localization system resorts only in depth images of the ceiling, captured by a Kinect sensor. Several experiments were carried out and the real trajectory of the robot was measured to performed a ground truth test of the localization system. Results show that visual odometry presents a far better result than wheel odometry, calculating the robot attitude with a good accuracy and very close to the ground truth. The attitude error obtained by wheel odometry increases significantly over time, causing the robot to degrade its own localization information, unlike visual odometry, which can get a better localization on the first trajectory and a reasonable localization in the second trajectory. Comparing the results considering the different levels of environment discovery, it is possible to conclude that, when the robot is moving in known environments, the proposed visual odometry is able to perform a global accurate localization, while in continuous environments discover, visual odometry needs to be fused with other sensor to eliminate the incremental errors. Hence, as visual odometry computes the attitude with more accuracy than wheels odometry, the incremental errors along time is less significant. In the future, the visual odometry localization system will be integrated in a self-localization system, complementing the existing odometry data (compass, magnetometers, encoders, ...), improving the accuracy of the overall system.

ACKNOWLEDGMENTS

This work was partially supported by Fundação para a Ciência e a Tecnologia, through IDMEC under LAETA.

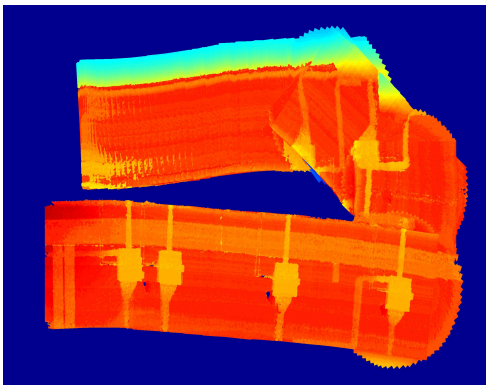

Fig. 13. Built map along a trajectory with few image overlapping

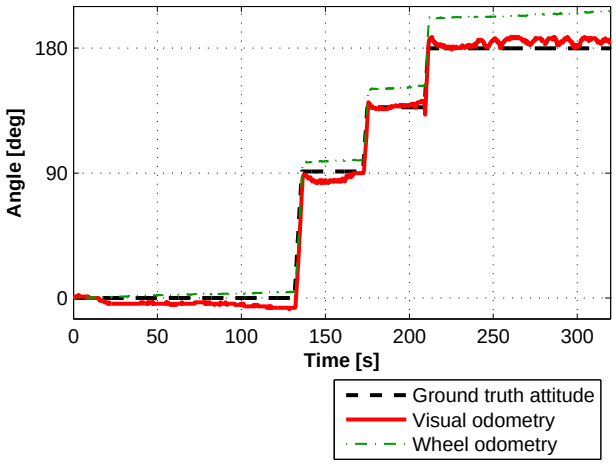

Fig. 14. Attitude of the robot along time

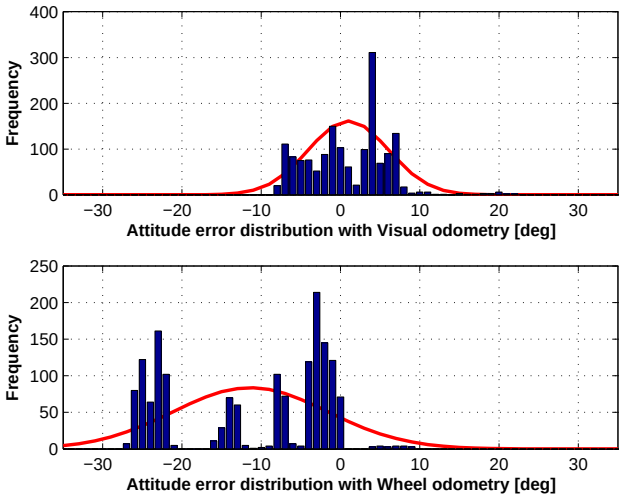

Fig. 15. Histogram of the attitude error

REFERENCES

[1] Markus Achtelik, Abraham Bachrach, Ruijie He, Samuel Prentice, and Nicholas Roy. Stereo vision and laser odometry for autonomous helicopters in gps-denied indoor environments. In *SPIE Defense, Security, and Sensing*, volume 7332, Orlando, Florida, USA, April 2009. International Society for Optics and Photonics.

[2] Motilal Agrawal and Kurt Konolige. Real-time localization in outdoor environments using stereo vision and inexpensive gps. In *2006 18th International Conference on Pattern Recognition, ICPR* ., volume 3, pages 1063–1068, Hong Kong, August 20-24 2006. IEEE.

[3] T. Bailey and H. Durrant-Whyte. Simultaneous localization and mapping (SLAM): part II. *Robotics Automation Magazine, IEEE*, 13(3):108 – 117, September 2006.

[4] J. Biswas and M. Veloso. Depth camera based indoor mobile robot localization and navigation. In *2012 IEEE International Conference on Robotics and Automation, ICRA.*, pages 1697–1702, St.Paul, USA, May 14-18 2012. IEEE.

[5] Jason Campbell, Rahul Sukthankar, Illah Nourbakhsh, and Aroon Pahwa. A robust visual odometry and precipice detection system using consumer-grade monocular vision. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, ICRA* ., pages 3421–3427, Barcelona, Spain, April 18-22 2005. IEEE.

[6] C. Cardeira and J. Sá da Costa. A low cost mobile robot for engineering

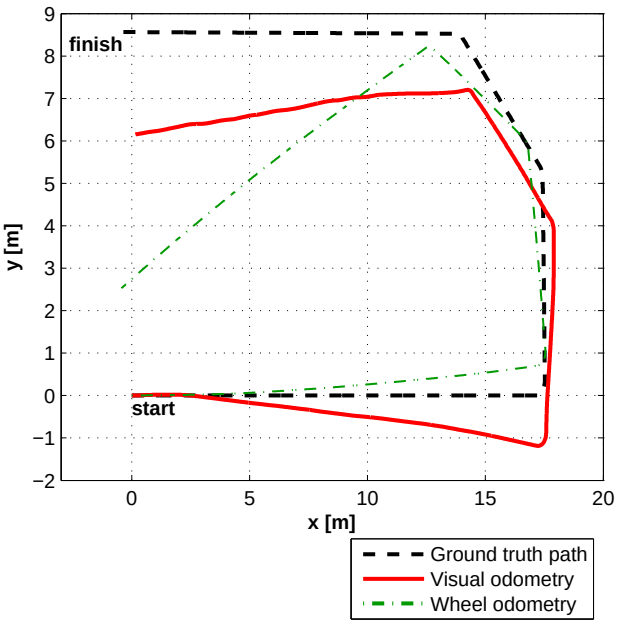

Fig. 16. Map with visual and wheel odometry position, considering a ground truth path

education. In *2005 31st Annual Conference of IEEE Industrial Electronics Society, IECON.*, pages 2162–2167, Raleigh, USA, November 6-10 2005. IEEE.

[7] F. Carreira, C. Christo, D. Valério, M. Ramalho, C. Cardeira, J. M. F. Calado, and P. Oliveira. 2d pca-based localization for mobile robots in unstructured environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)*, pages 3767–3868, Vilamoura, Portugal, October 2012.

[8] Howie Choset, Kevin M. Lynch, Seth Hutchinson, George A Kantor, Wolfram Burgard, Lydia E. Kavraki, and Sebastian Thrun. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. MIT Press, Cambridge, MA, June 2005.

[9] Peter Corke. *Robotics, Vision and Control: Fundamental Algorithms in MATLAB*, volume 73. Springer, 2011.

[10] H. Durrant-Whyte and T. Bailey. Simultaneous localization and mapping: part I. *Robotics Automation Magazine, IEEE*, 13(2):99 –110, June 2006.

[11] Yasuaki Fukutani, Tomoyuki Takahashi, Masahiro Iwahashi, Tetsuya Kimura, Samsudin Siti Salbiah, and Norrima Binti Mokhtar. Robot vision network based on ceiling map sharing. In *11th IEEE International Workshop on Advanced Motion Control (AMC 2010)*, pages 164–169, Nagaoka, Niigata, March 2010.

[12] Arturo Gil, Oscar Mozos, Monica Ballesta, and Oscar Reinoso. A comparative evaluation of interest point detectors and local descriptors for visual slam. *Machine Vision and Applications*, 21:905–920, 2010.

[13] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *The International Journal of Robotics Research*, 31(5):647–663, 2012.

[14] Kurt Konolige, Motilal Agrawal, and Joan Sola. Large-scale visual odometry for rough terrain. *Robotics Research*, pages 201–212, 2011.

[15] Davide Scaramuzza and Friedrich Fraundorfer. Visual odometry [tutorial]. *Robotics & Automation Magazine, IEEE*, 18(4):80–92, 2011.

[16] Davide Scaramuzza, Friedrich Fraundorfer, and Roland Siegwart. Real-time monocular visual odometry for on-road vehicles with 1-point ransac. In *IEEE International Conference on Robotics and Automation (ICRA2009)*, pages 4293–4299, Kobe, Japan, May 2009.

[17] Wolfram Burgard Sebastian Thrun and Dieter Fox. *Probabilistic Robotics*. Intelligent Robotics and Autonomous Agents. MIT Press, August 2005.

[18] C. Siagian and L. Itti. Biologically inspired mobile robot vision localization. *IEEE Transactions on Robotics*, 25(4):861–873, August 2009.

[19] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[20] De Xu, Liwei Han, Min Tan, and You Fu Li. Ceiling-based visual positioning for an indoor mobile robot with monocular vision. *IEEE Transactions on Industrial Electronics*, 56(5):1617–1628, May 2009.