Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Robust Landmark Detection and Localization; A Multisensor Approach

Dilan Amarasinghe, George K. I. Mann and Raymond G. Gosine *

* *Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL, A1B 3X5, Canada. (e-mail: amarasin@engr.mun.ca, gmann@engr.mun.ca, rgosine@engr.mun.ca).*

Abstract: This paper describes a landmark position measurement system using an integrated laser-camera sensor. Laser range finder can be used to detect landmarks that are direction invariant in the laser data such as protruding edges in walls, edges of tables, chairs. When such features are unavailable the processes that depend on landmarks such as navigation and simultaneous localization and mapping (SLAM) algorithms will not be able to perform at the best accuracy. However, in many instances larger number of landmarks can be detected using computer vision. In the proposed method camera is used to detect landmarks while the location of the landmark is measured by the laser range finder using laser-camera calibration information. Thus, the proposed method exploits the beneficial aspects of each sensor to overcome the disadvantages of the other sensor. Experimental results with important statistics are provided and an application in SLAM is presented.

1. INTRODUCTION

Recently, computer vision received much of attention for landmark detection and localization, especially in SLAM applications (Jeong and Lee (2006); Mouragnon et al. (2006)). However there are many drawbacks in vision based sensors. Monocular SLAM implementations require the features to be present in the field of view for a longer duration to facilitate the proper convergence of the feature position estimate. However, stereo vision has the ability to overcome some of issues in single camera systems, but require a heavy computational overhead. Thus, it is possible to use the features of both sensors, laser and camera, to overcome drawbacks of each. Hence this work demonstrates a novel application of a single laser-vision model. This paper proposes feature extraction at the sensor level while using laser-vision model as a single sensor for detection and locating landmarks. Therefore this paper constitutes following key contributions. First, the work demonstrates effective integration of laser and camera as a single sensor. Secondly the work demonstrates how the integrated laser-camera model be used to effectively solve the SLAM problem. The sensor also has the ability to either work as a laser only sensor or vision only sensor.

1.1 Related Work

The range and bearing to unique visual landmarks can be measured using different methods. The most common method is the use of stereo cameras (Saeedi et al. (2006)). Other methods include: single camera based feature position estimation (Montiel et al. (2006)) and optical flow based calculation (Bouguet and Perona (1995)). Although computer vision based SLAM methods shows significant advances, they exhibit one or more of the following drawbacks with respect to general SLAM applications.

(1) The methods were only demonstrated to work in small scale environments (Montiel et al. (2006)).
(2) Employs a large number of landmarks in the environment (Saeedi et al. (2006)).

These issues can be primarily attributed to the the large uncertainties associated with the vision based feature position calculation. Further, in stereo and other vision based feature position calculation methods, uncertainty of the feature position increases as the distance to the feature increases. Additionally, regular camera lens provide only a limited field of view. This severely limits the amount of time that a feature is actively observed when the robot is moving at relatively higher speeds. On the contrary, laser range finder provides excellent range measuring capabilities to direction invariant landmarks (such as chair and table legs, corners, tree trunks, poles, etc.). On multi sensor SLAM, Castellanos et al. (2001) have presented a laser-camera based method that fuses landmark information from laser range finder data as well as image data. The method presented by Castellanos et al. (2001) detects landmarks using data from each sensor and calculates the individual and joint compatibility between them. From the laser range finder it locates the line segments, corners and semiplanes. Using camera data it obtain redundant information about the landmarks that were observed by the laser range finder. Thus this method only facilitates the laser based landmarks with additional redundant information about the corners and semiplanes from vision data.

⋆ The authors would like to thank Natural Sciences and Engineering Research Council of Canada (NSERC), Canadian Foundation for Innovation (CFI) and Memorial University of Newfoundland for providing financial assistance for the research described above.

1.2 Landmark Detection and Position Estimation Using a Single Sensor

This section explores the applicability of each sensor for landmark detection and position estimation. In robotics, the camera and the laser range finder are the most commonly used sensors for environment sensing. Computer vision based solutions have long been proposed for detection and in many cases for position estimation of visually salient landmarks. The most important advantage of using computer vision for landmark detection is that it can detect visually salient landmarks with a high degree of details that can later be used for tracking or association. Due to the inherent sensor model, computer vision can only capture the bearing to a feature. Therefore in computer vision, stereo vision is the most popular method for direct landmark position estimation. On the contrary, a laser range finder scans its field of view to measure the distances to closest object. Usually, the measurements are taken at very small angular resolution and a higher range accuracy than any of the other range sensors, providing a high resolution depth plan of the field of view of the scanner. Next, the issues relating to the landmark detection and position measurement using a single sensor are addressed.

Monocular vision has been widely used in visual landmark detection in bearing only SLAM. Starting from the initial works of Andrew Davison (Davison and Murray (2002)) the research in vision based SLAM has moved to realtime monocular SLAM (Clemente et. al. (2007); Davison et. al. (2007)) implementations. In (Clemente et. al. (2007); Davison et. al. (2007)) the position (depth) of the visual landmarks is estimated using repeated observation of the landmark, and when the estimation converges it is initialized into the map. This type of feature initialization requires landmarks to be present in the field of view of the camera until the depth estimates converge to an acceptable level. Although these are pioneering methods in vision based SLAM, in typical application scenarios the landmarks cannot be guaranteed to remain in the field of view for a specific duration. In other methods the optical flow of a landmark, along with the robot velocities, can be used to calculate its position with respect to the robot frame.

In the detection of landmarks based on the laser range finder data, the corner and line (planes in the real world) features are the mostly used features (Martinez-Cantin et. al. (2006); Castellanos et al. (2001)). The landmarks that can be represented by a point in the map are often preferred over the line features, which can only be localized with a higher degree of freedom when the complete line segment is in the field of view of the scanner. The corner features that are invariant to the direction of the laser scan arise in the laser data due to objects such as corners in walls and other objects that have protrusions similar to legs of tables. However, in some cases these types of corner features may not be available in environments such as long corridors. Nevertheless, in most cases there are patterns on walls and other features that can be easily detected using computer vision. In addition, due to the differences in the appearance of surfaces under lighting, the regular corner features would usually appear as visually salient features. In the rest of this section two attempts in localizing features using computer vision and laser range data are discussed with their limitations. The next section introduces an integrated laser-vision sensor that exploits above mentioned properties of the visual features with the high accuracy of the laser based measurement.

Landmark localization using computer vision Landmark localization using only monocular vision has been achieved using two main methods: bearing only localization and optical flow based localization. Bearing only localization requires multiple wide baseline frames to infer the 2D position of a landmark. Therefore, the position estimation and the accuracy of the estimation of a landmark using bearing only readings are highly dependant on the movement of the camera and the number of sensor frames. In contrast, the optical flow based feature localization can be used to calculate the landmark position as soon as accurate optical flow data becomes available. Thus, in this paper, for monocular vision based landmark localization, only the optical flow based method was investigated.

From the six degree of freedom general model, the horizontal velocity of features (optical flow) ($\dot{p}$) on the image plane can be derived from the horizontal feature position (p), heading velocity (v), rotational velocity (ω), and focal length of the camera (λ) as follows:

$$\dot{p} = \frac{pv}{Z} - \frac{\omega}{\lambda}(\lambda^2 + p^2) \tag{1}$$

where Z is the distance to the feature in the direction of the heading velocity. Using the above equation and the camera model ($p/\lambda = X/Z$) where X is the perpendicular distance from the feature to the heading direction, the feature position with respect to the robot can be calculated by:

$$\begin{aligned} X &= \frac{p^2 v\lambda}{\dot{p}\lambda + (\lambda^2 + p^2)\omega} \\ Z &= \frac{pv}{\dot{p}\lambda + (\lambda^2 + p^2)\omega} \end{aligned} \tag{2}$$

The covariance of the calculated position can be found using the first order Taylor expansion of the feature position $[X, Y]^T$. The covariance matrix of the position calculation can be obtained from

$$\Sigma_{X,Z} = J \cdot \mathrm{diag}\left[\sigma_p, \sigma_{\dot{p}}, \sigma_v, \sigma_\omega\right] \cdot J^T \tag{3}$$

where

$$J = \begin{bmatrix} \frac{\partial X}{\partial p} & \frac{\partial X}{\partial \dot{p}} & \frac{\partial X}{\partial v} & \frac{\partial X}{\partial \omega} \\ \frac{\partial Z}{\partial p} & \frac{\partial Z}{\partial \dot{p}} & \frac{\partial Z}{\partial v} & \frac{\partial Z}{\partial \omega} \end{bmatrix}$$

and σ_p, $\sigma_{\dot{p}}$, σ_v, and σ_ω are the standard deviations of the horizonal feature position on the image, horizontal optical flow, heading velocity and rotational velocity of the robot, respectively. The uncertainty of the calculated locations can be evaluated by comparing the area of the ellipsoid defined by the 95% confidence interval. The uncertainty comparison for varying optical flows and feature positions is shown in Figure 1. From Figure 1 it is evident that at low optical flows the uncertainty increases regardless of the

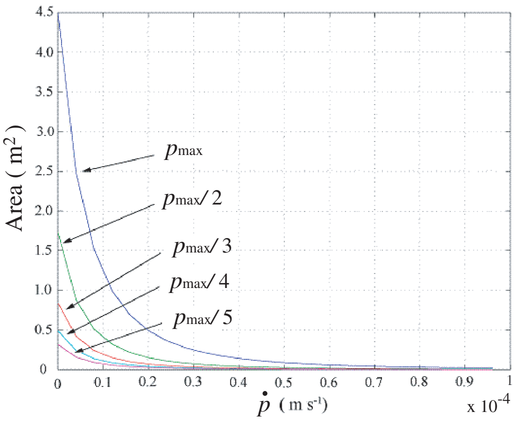

Fig. 1. Sensitivity of the uncertainty of the feature localization. The uncertainty is quantified by the area of the ellipse representing 95% confidence. $p_{\max} = 1.75mm$, $v = 0.092m/s$ and $\omega = 4 \times 10^{-3} rad/s$. ($\sigma_p = 10.9 \times 10^{-6} m$, $\sigma_{\dot{p}} = 0.3 \times 10^{-4} m/s$, $\sigma_v = 7.8mm/s$, $\sigma_\omega = 10^{-6} rad/s$)

feature position on the image. Moreover, as the feature moves closer to the edge of the image, the uncertainty increases even for the same optical flow value. Generally, a robot encounters many combinations of robot velocities and feature positions could give rise to high covariance values in the feature position calculations. The limitations in the usable range of optical flow and feature position make the optical flow based feature position calculation method unsuitable for SLAM applications.

Landmark localization using laser data The direction invariant features in the laser data can be identified as unique landmarks using the minimum points in the laser data plot (Schulz et. al. (2003)). These landmarks generally remain in the laser data regardless of the direction of scan. In addition to the convex features that appear as minimum points in the laser data, concave points such as sharp corners can be reliably detected in the laser data. However, as shown in Figure 2, in certain environments such as in long corridors, there might not be any directional invariant features. In such cases feature based laser only SLAM implementations will not be possible unless higher level features such as lines are used.

1.3 Objective

The main objective of this paper is to develop a reliable landmark detection and localization method that uses an integrated laser-camera sensor for SLAM applications. This papers presents a novel method for landmark detection and location calculation based on multisensor data in the context of SLAM. In contrast to the other notable works in multisensor SLAM Castellanos et al. (2001) the proposed method fuses the information in sensor domain rather than fusing map information that is being built using each sensor, as shown in Fig. 3. In the proposed work a camera is mounted on a laser range finder and the coordinate transformations have been obtained through a experimental calibration process Ortin et al. (2005). The vertical lines in environment are detected using the image data (bearing information) and the range to the vertical lines can be then interpolated using the laser readings

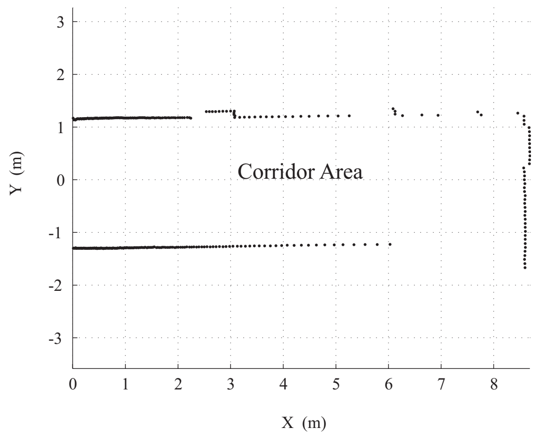

Fig. 2. A typical laser reading in an indoor environment where there are not sufficient direction invariant features.

and the coordinate transformation between the laser and the camera. These located features are then used in the extended Kalman filter based SLAM formulation.

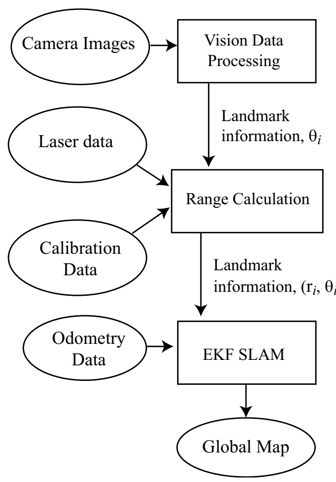

Fig. 3. Block diagram of the proposed SLAM process

2. CALIBRATED LASER-VISION SENSOR

A camera is mounted on the laser range finder using a custom made bracket as shown in Fig. 4. The camera is mounted at the center of the laser range finder to maintain the coordinate transformation between laser scanning plane and camera coordinate system as simple as possible. The coordinate frames are defined as shown in the Fig. 4.

2.1 Visual Landmark Detection

Landmarks in the camera images can take several forms. The most common landmarks are the visually distinct corner features. Other visually salient landmarks include, lines, arcs, and user defined objects. In this paper the visually salient vertical line features were detected in the captured images. Consistent lines features are the most robust in terms of detection accuracy and repeatability. In this work two algorithms has been evaluated for the detection of vertical lines in the images.

(1) Hough transform based method.

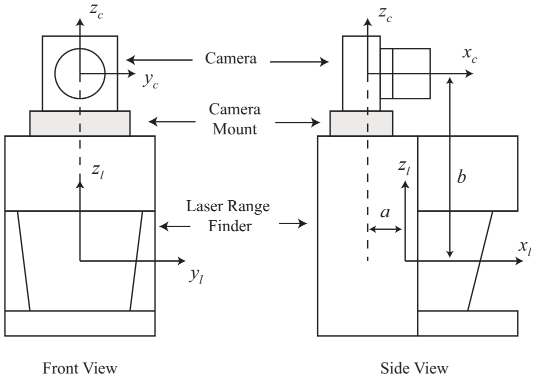

Fig. 4. Coordinate frames of calibrated laser-vision sensor

(2) Corner feature based method.

Line detection algorithms based on the Hough transformation is most popular in computer vision and pattern recognition. Hough transformation typically accumulates the votes for line configurations based on their support in the binary image. Since it is of interest to detect only the vertical (or close to vertical) lines, the search space can be restricted to compute the angle values in the vicinity of zero, thus reducing the computational cost. In addition to the hough transform based method, a simpler and computationally efficient corner based method was tested for vertical line detection. Initially, a set of horizontal lines were superimposed on the original image as shown in Fig 5. Then, all the resulting corner features are detected using Harris corner detector Shi and Tomasi (1994) and are indicated by the white circles in Fig. 5.

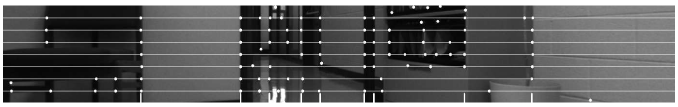

Fig. 5. Line feature detection using artificially generated corner features.

This list of corner features are then searched for sets of features that are vertically aligned. If the number of features in a set is greater than a threshold value then the average of the horizontal position is identified as a consistent vertical line. Identified lines are marked with white line stubs at the bottom of the image frame shown in the Fig. 5. The corner based method is approximately equivalent to the Hough transform based method. Instead of accumulating the pixel count at finer resolution for the full image, the corner based method samples the image at vertical line positions and accumulate the points where there is strong evidence for vertical lines.

A comparison of the two methods are shown in the Fig. 6 for three typical images that is taken during a robot run. The lines in the top part of the image are the ones detected using Hough transformation and the lines in the bottom part detected using corner based method. It is evident from the images that on average Hough transform returns more line images than the corner based method. This can be attributed to the fact that it accumulate the evidence for lines in the whole region than some sampled points in the image as in the case with corner based method. From the Fig. 6 it is evident that in addition to the ability to recover large number of landmarks the Hough transformation based method is more accurate as well. Therefore in the work described in this paper Hough transformation is selected.

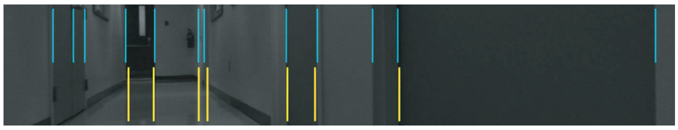

Fig. 6. Detected line features using Hough transformation and the corner based method.

2.2 Measurement Model

In this section it is assumed that the transformation between the sensor frames have been calibrated for lens distortion and sensor misalignment. Laser ranger provides a set of scanned reading that provides the range to the objects in the laser scan plane. The scanner is able to operate in a field of view of 180 degrees with a half a degree resolution. The bearing angle (θ_l) of the detected line features can be calculated using the camera model. Then using the coordinate transformation between the camera and the laser range finder and the calibration information the range to the line features can be interpolated using laser range scan. This process of range interpolation is shown in the Fig. 7.

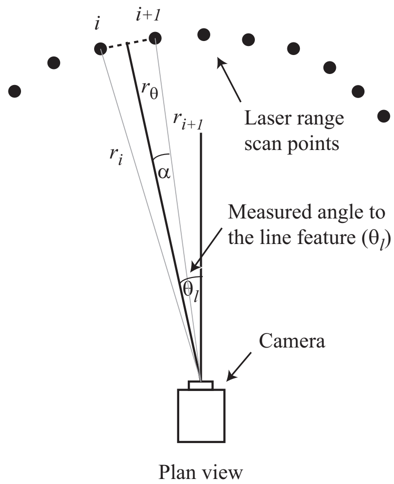

Fig. 7. Interpolation of the range to the line feature.

With a resolution of the laser range scanner at 0.5^0 the range to the line feature can be calculated using following interpolation.

$$r_\theta = \frac{r_{i+1}\cos(\theta - \alpha) + r_i \cos(0.5^0 - \alpha + \theta)}{2\cos(\theta)} \tag{4}$$

Since the bearing to the feature is measured using camera model and the range is measured using the interpolated range data, the uncertainty of the measurements also have to be calculated using the characteristics of each sensor. In the camera model, the incident angle for the same image area changes with the distance from the optical axis. Hence the bearing uncertainty increases when the distance to the line feature from the optical axis increases. But, since the used camera lens has only a narrow field of view, bearing uncertainty can be assumed to be a constant. For the range, usual constant uncertainty of the laser range finder is used. Thus, the covariance matrix of the measurements can be expressed as,

$$R = \text{diag}[\ \sigma_r^2 \ \ \sigma_\theta^2\ ]. \tag{5}$$

Where σ_r and σ_θ are the standard deviation of the range and bearing measurement errors, respectively.

3. EXPERIMENTS AND RESULTS

In this section two groups of experiments are carried out, the first for the verification of the method and the second is an application of the method to SLAM. In the verification experiments the vision data is superimposed on known laser data to test the accuracy of the method. In addition to that the vision based landmark detection is compared with a laser only method for the number of retrieved landmarks.

3.1 Verification of the Method

The laser data and the camera image is superimposed for the verification of the method. Fig. 8 shows the results of the feature detection and locating using integrated sensor for a typical set of image and laser scan data. Fig. 8 shows that vertical line features on the wall can be accurately localized using the proposed method.

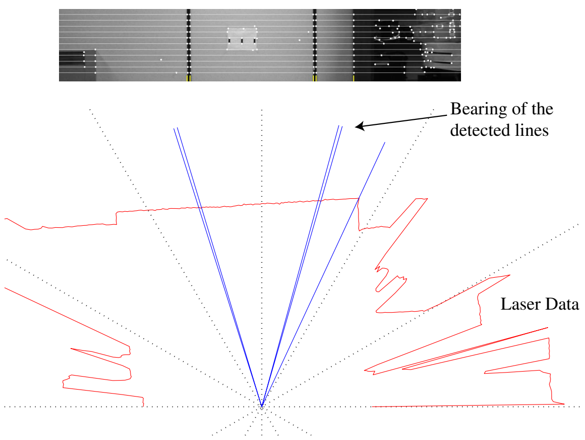

Fig. 8. The landmarks detected by the camera and their bearing angle superimposed on laser readings.

As discussed previously, the protruding features in the laser data can be detected as landmarks in the laser data. These features can be detected using strong corner points in the plot of laser data. Fig. 9 shows a comparison between number of landmarks that can be detected in laser data and in image data during a robot run. It is clearly evident that there are significant periods where image features out number the laser based landmarks. Further, the number of image features remain much more steady compared to the large variations in the number of laser based landmarks. Additionally, it should be noted that where there is low number of visual features there is a significantly higher number of laser based landmarks. Therefore, landmark localization method that uses both methods of detection can benefit from the higher number of landmarks throughout the run of the robot. Although the results are purely specific to a given environment, the total number of landmarks can be improved using the proposed method in addition to the laser only methods.

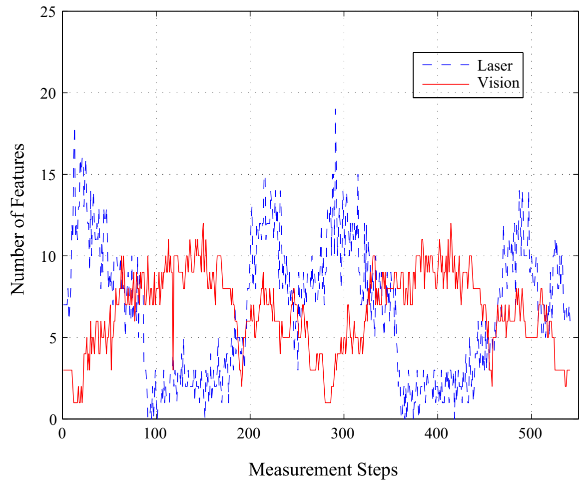

Fig. 9. Number of landmark features detected by vision and laser system.

3.2 Application in EKF SLAM

An experiment was conducted using the Pioneer 3AT robot in a typical indoor environment in order to illustrate the viability of the landmarks located using the laser-vision based in a typical SLAM scenario. The robot was driven approximately 67.5m forming two loops. During this experiment the laser range data, images from the camera and odometry data were logged at regular spatial intervals. After the landmarks are detected and located using laser data and images, the data is processed off-line using the EKF method Dissanayake et al. (2001). The Joint Compatibility Branch and Bound (JCBB)Neira and Tardos (2001) algorithm was used for the data association. A from the data gathered during the robot run map consisting of 71 landmarks that has been built (Fig. 10(b)). The Fig. 10(a) shows the robot path using pure odometry data, where there are significant errors.

4. CONCLUSION

In this paper it is shown that computer vision and laser range scanner can be used to accurately detect and measure the visually salient landmarks in the environment. Further, such measurements can be readily integrated into EKF based SLAM method to build maps of typical indoor environments. One possible pitfall of this method arises when the line features in the real 3D world does not intersect with the laser scan plane. However, this condition can be ensured by mapping the laser points to the image plane using the sensor calibration data and focusing on the vertical lines that intersect mapped laser data curve. In the current method this cannot be directly achieved as the Hough transform based method return generic vertical lines but not localized vertical lines. Although not directly comparable to the multisensor SLAM presented in Castellanos et al. (2001), it is possible to observe that the proposed method can be used localize strong (visually

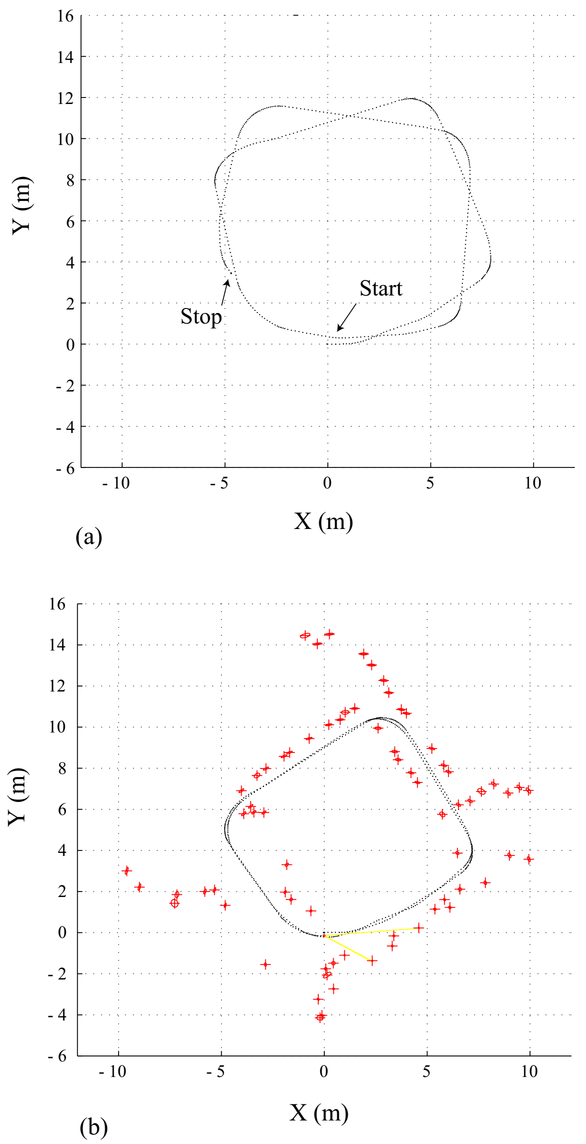

Fig. 10. Results of a localization and mapping of a robot run: (a) with odometry, and (b) using EKF and vision-laser landmark localization.

salient) landmarks using both camera and the laser than using data camera images as a redundant support role. Future extensions of this work include the use of more accurate sensor uncertainty modeling specially, in the case of bearing angle to the landmark and experimentation in large looping environments with possible sub-mapping.

5. ACKNOWLEDGMENT

Authors would like to acknowledge the work of the contributors to the implementation EKF SLAM algorithm[1] and the JCBB data association algorithm[2] .

REFERENCES

W. Y. Jeong and K. M. Lee, "Visual slam with line and corner features," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2006, pp. 2570–2575.

E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, , and P. Sayd, "Monocular vision based slam for mobile robots," in *Proceedings of International Conference on Pattern Recognition*, August 2006, pp. 1027–1031.

A. J. Davison and D. W. Murray, "Simultaneous localization and map-building using active vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 865–880, 2002.

P. Saeedi, P. D. Lawrence, and D. G. Lowe, "Vision–based 3–d trajectory tracking for unknown environments," *IEEE Transactions on Robotics*, vol. 22, no. 1, pp. 119–136, February 2006.

J. M. M. Montiel, J. Civera, and A. J. Davison, "Unified inverse depth parametrization for monocular slam," in *Proceedings of the Robotics, Science and Systems*, August 2006.

J.-Y. Bouguet and P. Perona, "Visual navigation using a single camera," in *Proceedings of the International Conference on Computer Vision*, 1995, pp. 645–652.

M. W. M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (slam) problem," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, June 2001.

J. A. Castellanos, J. Neira, and J. D. Tardos, "Multisensor fusion for simultaneous localization and map building," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 6, pp. 908–914, December 2001.

D. Ortin, J. Neira, and J. M. M. Montiel, "Relocation using laser and vision," in *Proceedings of the IEEE International Conference on Robotics and Automation*, April 2005, pp. 1505–1510.

J. Shi and C. Tomasi, "Good features to track," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 1994, pp. 593–600.

J. Neira and J. D. Tardos, "Data association in stochastic mapping using the joint compatibility test," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 6, pp. 890–897, December 2001.

Andrew J Davison and David W Murray. "Simultaneous localization and map-building using active vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 865–880, 2002.

Laura A Clemente, Andrew J Davison, Ian D Reid, Jose Neira, and Juan D Tardos. "Mapping large loops with a single hand-held camera,"In *Proceedings of the Robotics Science and Systems*, 2007.

Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. "Monoslam: Real-time single camera slam,"*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

Ruben Martinez-Cantin, Jose A Castellanos, Juan D Tardos, and J M M Montiel. "Adaptive scale robust segmentation for 2d laser scanner," In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 796–801, October 2006.

Dirk Schulz, Wolfram Burgard, Dieter Fox, and Armin B. Cremers. "People tracking with mobile robots using sample-based joint probabilistic data association filters," *The International Journal of Robotics Research*, vol. 22, no. 2, pp. 99–116, February 2003.

[1] http://www-personal.acfr.usyd.edu.au/tbailey/
[2] http://webdiis.unizar.es/ neira/slam.html