2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

Integrated Laser-Camera Sensor for the Detection and Localization of Landmarks for Robotic Applications

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

***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 fail. 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 of an application in SLAM is presented to verify the results.**

I. INTRODUCTION

Among various sensors used in detection and localizing landmarks in robotics, laser range scanners received much of attention, mainly due to its response behaviour and ability to accurately scan a wider field of view. Laser range finders can precisely locate landmarks in environments having directional variant features, such as protruding edges in walls, edges of objects located in the field of view such as chairs, tables, and also moving objects such as humans [1]. However, in environments such as, corridors of having flat walls, long empty rooms and halls, the laser data will contain minimum number of features that can be detected as landmarks.

Recently, computer vision received much of attention for landmark localization, specially in simultaneous localization and mapping (**SLAM**) [2], [3], [4] as visually salient features can be easily extracted from camera images. 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, particularly for calibration and 3D estimates. Thus, it is possible to use the features of each sensor to overcome drawbacks of the other one. Hence this work demonstrates a novel application of a single laser-vision model. Early work of laser-vision model in SLAM uses two sensor readings separately and fuse the data two different maps. The maps are then fused at a post processing stage. In contrast 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 effective use of an integrated laser-camera model to solve the SLAM problem is demonstrated.

A. Related Work

The research in computer vision–based SLAM can be broadly categorized into two areas. They are: appearance based methods and feature or landmark based methods. In appearance based localization and mapping image features are collectively used to describe a scene. These feature based descriptions are used to compare and contrast the images that robot acquires along the way. Hence when a robot revisits an environment, the localization algorithm will be able to measure the similarity between the images of the current scene and the images that are registered in a database. In most cases this type of qualitative localization and mapping can only generate topological representations of the environment. Although it provides a viable and a more natural mapping and localization procedure, the qualitative algorithms does not provide detailed information about the environment. Details in such a map may be inadequate, specially when robots require accurate information about the structure of the environment for tasks such as path planning. Although appearance based methods has been used in SLAM [5], [6], [7], they are mostly used in the re-localization of the robots [8], [9], [10].

In contrast to the appearance based methods, landmark based methods uniquely identifies visually salient landmarks in the environment and calculate their position with respect to robot. Such measurements can be used in estimators to build the map of the visual landmarks while localizing the robot. The primary advantage of the landmark based methods over the appearance based methods is the higher fidelity of the map. In landmark based methods the range and bearing to the features can be calculated using different methods. The most common method is the use of stereo cameras [11], [12], [13], [14], [15]. Other methods include: single camera based feature position estimation [16], [17] and optical flow based calculation [18]. Although computer vision based SLAM methods shows significant advances, they exhibit one or more of the following drawbacks with respect to general SLAM applications.

This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC), Memorial University of Newfoundland and Canadian Foundation for Innovation (CFI)

Dilan Amarasinghe, George K. I. Mann, and Raymond G. Gosine are with Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL, A1B 3X5, Canada. {amarasin,gmann,rgosine}@engr.mun.ca

1) The methods were only demonstrated to work in small scale environments [11], [16], [17].
2) Employs a large number of landmarks in the environment [13], [18].

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 in the SLAM process, specially if the robot is moving at relatively higher speeds.

On the contrary laser range finder provides excellent range measuring capabilities and has been widely used in SLAM implementations. Landmarks that are generally invariant to the direction of scanning (such as chair and table legs, corners, tree trunks, poles, etc.) can be identified in laser range data. However, typical indoor environments with corridors, walls and other structured shapes, either does not have any corner features or have only very few features. During the estimation process when landmarks are absent in the environment uncertainty of the estimator rapidly grows. The landmarks that will be encountered with a higher robot uncertainty will have a higher uncertainty bound (Theorem 3 in [19]). This will lead to possible inconsistent data associations when the robot revisits the same area. Hence frequent featurelessness in the environment will lead to a highly unstable SLAM process. On the other hand computer vision can be used to detect visually salient features on walls and other places where laser range finder fails to detect landmarks and the laser range finder can be used to measure the range to those visually salient landmarks. On multi sensor SLAM, Castellanos et. al. [20] have presented a laser-camera based method that fuses landmark information from laser range finder data as well as image data. The method presented in [20] 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. In contrast, the proposed method uses vision as the primary sensor to obtain vertical edge features and then use data from the laser range finder to measure the range to those landmarks. Therefore, there is no dependency between the geometrical structure of the landmarks between the laser and vision data.

B. 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 [20] 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. 1. 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 [21]. 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 and the coordinate transformation between the laser and the camera. These located features are then used in the extended Kalman filter based SLAM formulation.

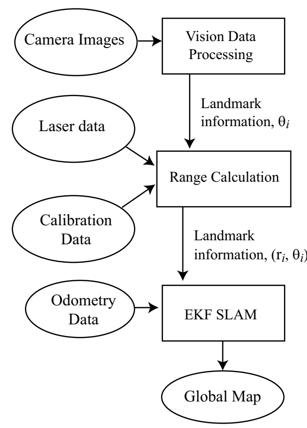

Fig. 1. Block diagram of the proposed SLAM process

C. Outline

The rest of the paper is organized as follows. The proposed method for landmark detection and localization with respect to the robot is presented in the Section II. Section III provides the experiments conducted to verify the algorithm and the results of an application of the extracted landmark data in SLAM. In section IV we provide a discussion on the proposed method and draw our conclusions.

II. Calibrated Laser-Vision Sensor

A camera is mounted on the laser range finder using a custom made bracket as shown in Fig. 2. 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. 2.

A. 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.

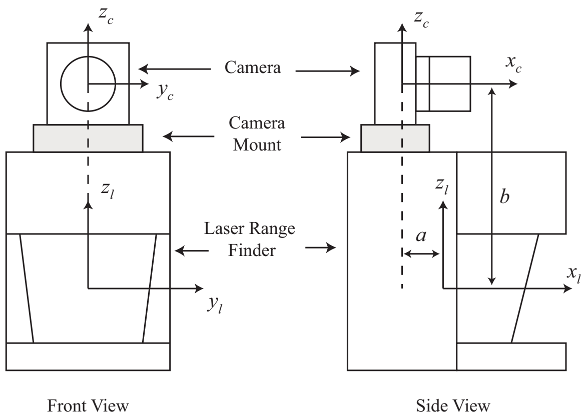

Fig. 2. Coordinate frames of calibrated laser-vision sensor

Fig. 3. The camera and the laser range sensor used in the experiments.

1) Hough transform based method.
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 4. Then, all the resulting corner features are detected using Harris corner detector [22] and are indicated by the white circles in Fig. 4.

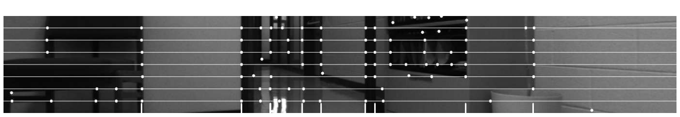

Fig. 4. Line feature detection using artificially generated corner features.

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. 4. 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. 5 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. 5 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.

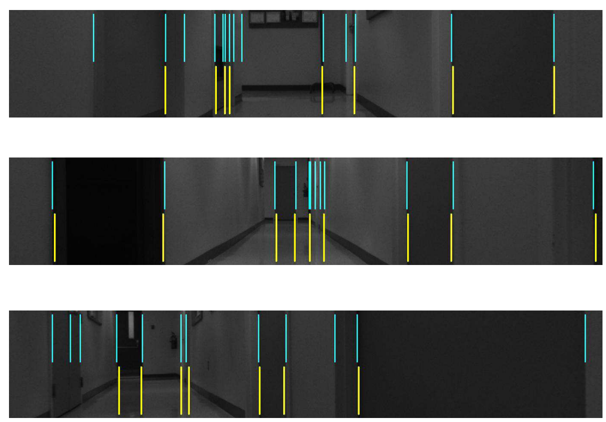

Fig. 5. Detected line features using Hough transformation and the corner based method.

B. Sensor calibration

In order to measure the distances to the visual landmarks using the laser range finder, the coordinate transformations of the two sensors have to be accurately calibrated. There are two possible sources for errors in the calibration information: the errors in the alignment of the frames of the sensors (parameters a and b in Fig. 2) and the errors in camera calibration. Although the camera is calibrated using standard camera calibration techniques[1], the distortions specially the edge of the images contribute significantly to the accurate alignment of the sensors.

The main objective of the sensor calibration method used in this paper is to accurately map the field of view of the

[1] MATLAB toolbox for camera calibration, http://www.vision.caltech.edu/bouguetj/calibdoc/

camera to that of the laser range finder. In order to achieve that objective, a 'v' shaped target with black and white faces is placed in front of the robot. In a series of image and laser data with the 'v' shaped object placed to span the field of view of the camera (since the field of view of the camera is less than that of the laser range finder), the angle to the tip of 'v' is measured from the center of each sensor. In the camera images it is measured in degrees from the optical axis (θ_c) and in the laser range finder it is measured from the central laser scan (θ_l). Thus, the error in the calibration can be calculated from $e = \theta_l - \theta_c$. As shown in Fig. 6 the error e is approximated using a higher order polynomial $f(\theta_c)$ with respect to θ_c. Thus, for any new measurement in the image, θ_c the corresponding mapping angle in the laser range finder can be calculated from $\theta_c + f(\theta_c)$.

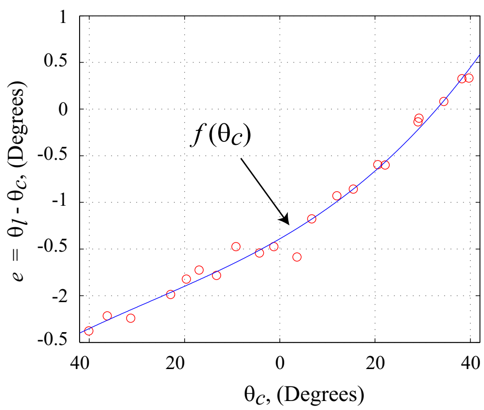

Fig. 6. Calibration curve for the mapping between field of view of the camera and the field of the view of the laser range finder.

C. Measurement Model

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 $180°$ 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.

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

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

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

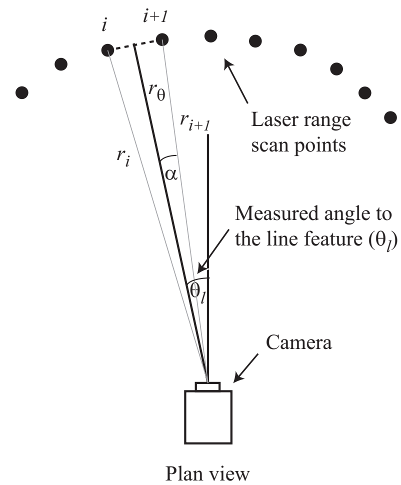

Fig. 7. Interpolation of the range to the line feature.

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,

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

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

III. 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.

A. 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.

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.

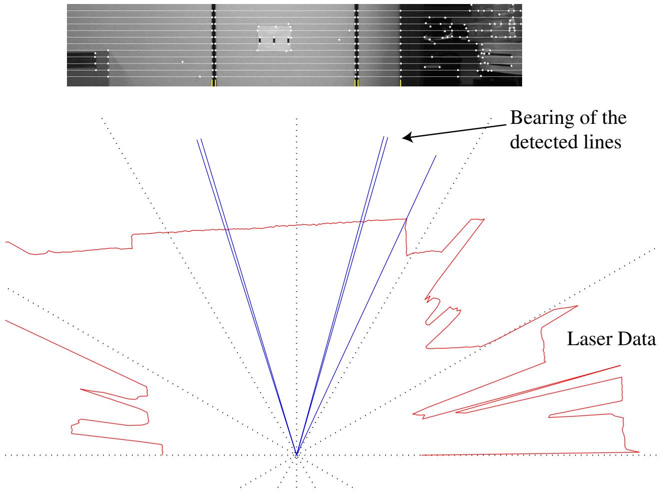

Fig. 8. The landmarks detected by the camera and their bearing angle superimposed on laser readings.

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.

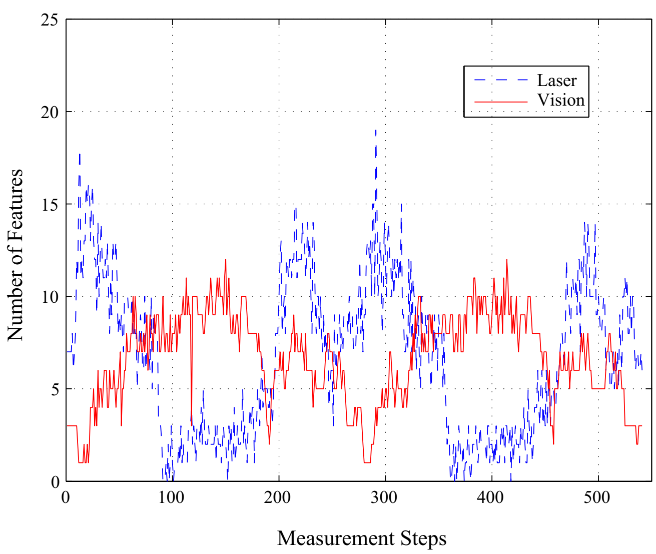

Fig. 9. Number of landmark features detected by vision and laser system.

B. Application in EKF based 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 [19]. The Joint Compatibility Branch and Bound (JCBB)[23] 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. The 95% confidence bounds of the errors in robot pose estimate are shown in Fig. 11. In Fig. 11 it is possible observe the effects loop closing in the robot position estimation around the midway point of the robot run.

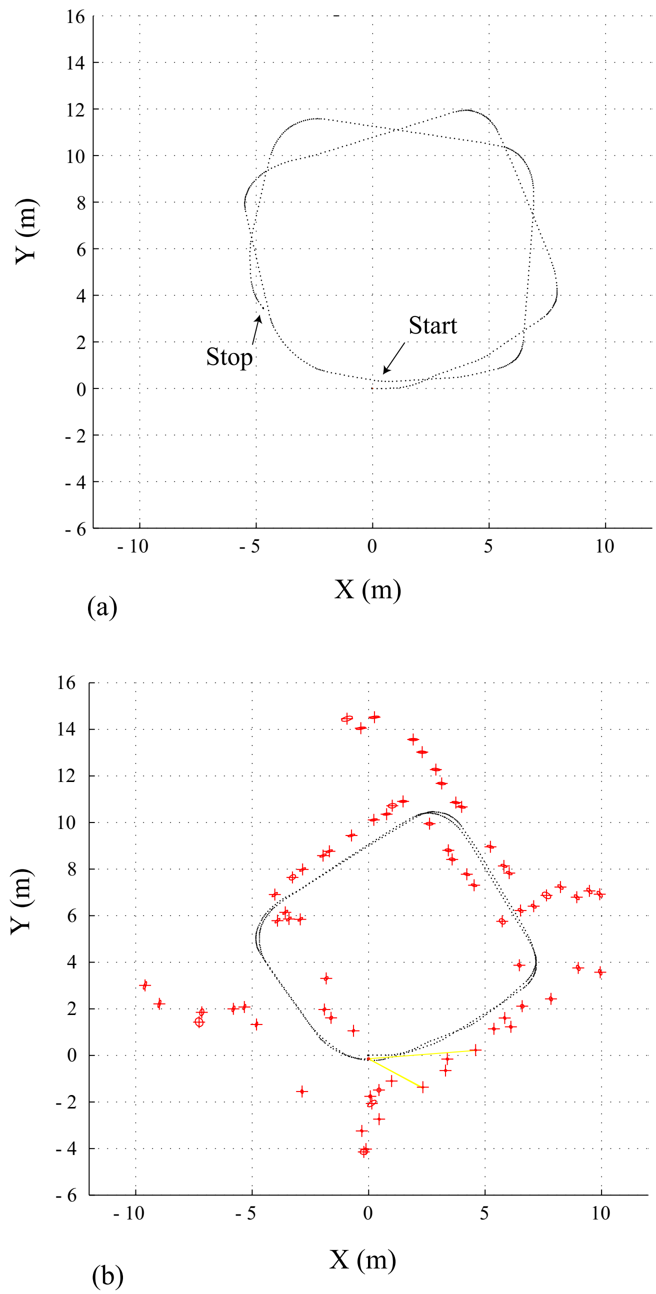

Fig. 10. Results of a localization and mapping of a robot run: (a) with odometry, and (b) using EKF and vision-laser landmark localization.

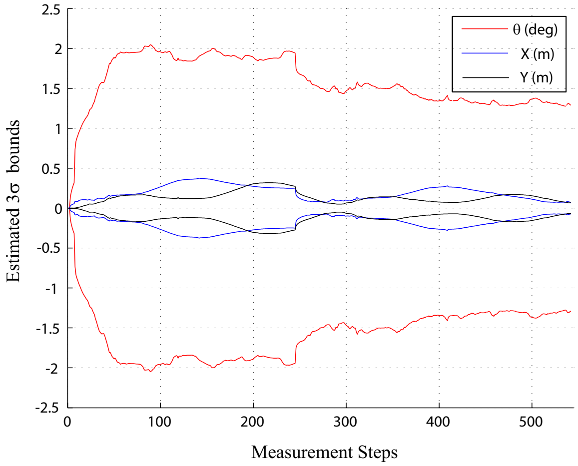

Fig. 11. 3σ bounds of the localization errors.

IV. 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 [20], it is possible to observe that the proposed method can be used localize strong (visually 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.

V. Acknowledgment

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. Also authors would like to acknowledge the work of the contributors to the implementation EKF SLAM algorithm[2] and the JCBB data association algorithm[3].

References

[1] J. E. Guivant, F. R. Masson, and E. M. Nebot, "Simultaneous localization and map building using natural features and absolute information," *Robotics and Autonomous Systems*, vol. 40, no. 2, pp. 79–90, 2002.

[2] 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.

[3] 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.

[4] J. V. Miro, G. Dissanayake, and W. Zhou, "Vision-based slam using natural features in indoor environments," in *Proceedings of International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, December 2005, pp. 151–156.

[5] P. E. Rybski, S. I. Roumeliotis, M. Gini, and N. Papanikolopoulos, "Appearance–based minimalistic metric slam," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2003, pp. 194–199.

[6] J. M. Porta and B. J. A. Krose, "Appearance–based concurrent map building and localization using a multi–hypotheses tracker," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2004, pp. 3424–3429.

[7] H. M. Gross, A. Koenig, and S. Mueller, "Omniview–based concurrent map building and localization using adaptive appearance maps," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, October 2005, pp. 3510–3515.

[8] J. M. Porta, J. J. Verbeek, and B. Krose, "Enhancing appearance–based robot localization using sparse disparity maps," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2003, pp. 980–985.

[9] B. J. A. Krose, N. Vlassis, and R. Bunschoten, "A probabilistic model for appearance–based robot localization," *Image and Vision Computing*, vol. 12, no. 6, pp. 381–391, April 2001.

[10] I. Ulrich and I. Nourbakhsh, "Appearance–based place recognition for topological localization," in *Proceedings of the IEEE International Conference on Robotics and Automation*, April 2000, pp. 1023–1029.

[11] A. J. Davison, "Mobile robot navigation using active vision," Ph.D. dissertation, University of Oxford, June 1998.

[12] 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.

[13] 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.

[14] S. Se and P. Jasiobedzki, "Photo–realistic 3d model reconstruction," in *Proceedings of the IEEE International Conference on Robotics and Automation*, May 2006, pp. 3076–3082.

[15] S. Se, D. Lowe, and J. J. Little, "Vision–based global localization and mapping for mobile robots," *IEEE Transactions in Robotics*, vol. 21, no. 3, pp. 364–375, June 2005.

[16] A. J. Davison, "Real-time simultaneous localization and mapping with a single camera," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, October 2003, pp. 1403–1410.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] 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.

[2] http://www-personal.acfr.usyd.edu.au/tbailey/

[3] http://webdiis.unizar.es/ neira/slam.html