Robotica (2010) volume 28, pp. 663–673. © Cambridge University Press 2009
doi:10.1017/S0263574709990312

# Landmark detection and localization for mobile robot applications: 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, Canada, A1B 3X5.*

(Received in Final Form: July 3, 2009. First published online: August 11, 2009)

**SUMMARY**

This paper describes a landmark detection and localization 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, and chairs. When such features are unavailable, the dependant processes will fail to function. 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. While highlighting the drawbacks and limitations of single sensor based methods, an experimental results and important statistics are provided for the verification of the affectiveness sensor fusion method using Extended Kalman Filter (EKF) based simultaneous localization and mapping (SLAM) as an example application.

KEYWORDS: Sensor fusion; Landmark detection; Landmark localization; Simultaneous localization and mapping; Mobile robotics.

## 1. Introduction

Among various sensors used in solving the simultaneous localization and mapping (SLAM) problem in robotics, laser range scanners has received most attention, mainly due to its response behavior 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 or tables, and also moving objects such as humans.[1] However, when such features are unavailable, in such environments such as corridors having flat walls, long empty rooms and halls, the laser data will contain a minimum number of features that can be detected as landmarks.

Recently, computer vision received much attention in SLAM[2–4] and has the ability to extract visually salient features even in flat walls. 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 this issue in single camera systems, but requires a heavy computational overhead, particularly for calibration and 3D estimates. Thus, it is possible to use the features of both sensors, laser, and camera, to overcome the drawbacks of the other. Hence this work demonstrates a novel application of a single sensor based on a laser-vision model. Early work of the laser-vision model use two sensor readings separately and fuses the SLAM data in the postprocessing stage to estimate robot pose. In contrast, the method proposed in this paper performs feature extraction at the sensor level while using laser-vision model as a single sensor for detecting and locating landmarks. Therefore this paper presents the following key contributions. First, the work demonstrates effective integration of laser and camera as a single sensor for general purpose robot navigation. Second, the work demonstrates how the integrated laser-camera model can be used effectively to solve the SLAM problem. The integrated sensor also retains its ability to work as either a laser only sensor or a vision only sensor.

### *1.1. Related work*

The research in computer vision based SLAM can be broadly categorized into two areas. They are: appearance based methods and feature 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 the 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 more natural mapping and localization procedure, the qualitative algorithms do not provide detailed information about the environment. Details in such a map may be inadequate, especially when robots require accurate information about the structure of the environment for tasks such as path planning. Although appearance based methods have been used in SLAM,[5–7] they are mostly used in the relocalization of the robots.[8–10]

In contrast to the appearance based methods, feature based methods uniquely identify visually salient landmarks in the environment and calculate their position with respect to the robot. Such measurements can be used in estimators to build

* Corresponding author. Email: amarasin@engr.mun.ca

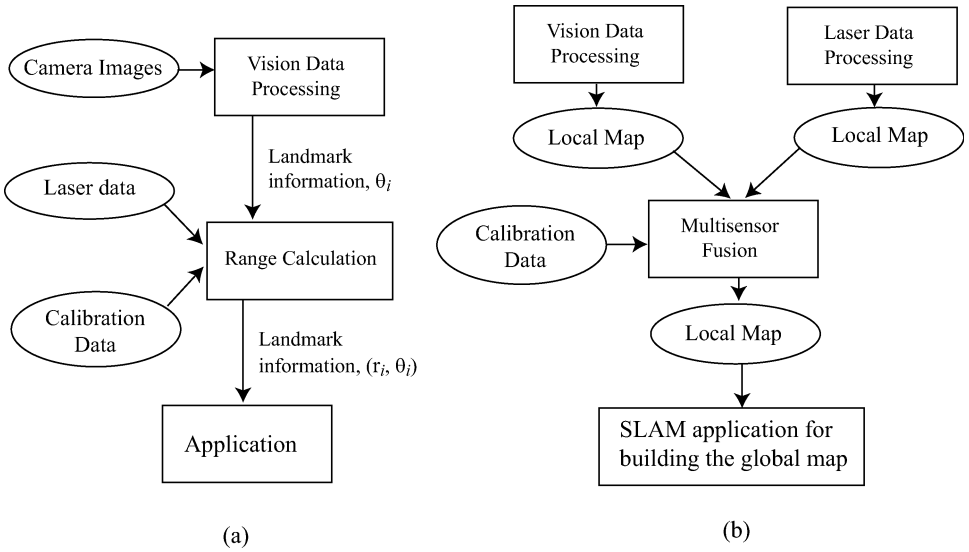


Fig. 1. Block diagram of (a) the proposed multisensor based SLAM process and (b) the multisensor based SLAM process presented in ref. [20]. The proposed method performs a sensor level fusion of information whereas the method in (b) fuses the information at map level.

the map of the visual landmarks while localizing the robot. The primary advantage of using feature based methods is the higher fidelity of the map. The feature based methods can be classified based on the method that they use to calculate the range and bearing to the features. The most common method is the use of stereo cameras.[11–15] Other methods used to calculate the feature position include: single camera based feature position estimation[16,17] and optical flow based calculation.[18] Although computer vision based SLAM methods have shown 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.[11,16,17]
(2) It is necessary to have a large number of features in the environment for the SLAM algorithms to properly converge.[13,18]

These issues can be primarily attributed to 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 with increased distance. Additionally, a regular camera lens provides only a limited field of view. This severely limits the amount of time that a feature is actively observed in the SLAM process, especially if the robot is moving at relatively higher speed.

On the contrary, the 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 do not have any 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 ref. [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. However, computer vision can be used to detect visually salient features on walls and other places where it is not possible to use a laser range finder to detect landmarks, and the laser range finder can be used to measure the range to the 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 with image data at map level. The method presented in ref. [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 obtains redundant information about the landmarks that were observed by the laser range finder. Thus this method only provides 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 uses data from the laser range finder to measure the range to those landmarks. Therefore, there is no dependency on the geometrical structure of the landmarks between the laser and the vision data. Figure 1 compares and contrasts the similarities and difference between the proposed method and ref. [20].

### *1.2. Objective*

The main objective of this paper is to introduce a novel integrated laser-camera sensor that can be readily used in landmark based simultaneous localization and mapping algorithms. In contrast to the other notable works in multisensor SLAM[20] the proposed method fuses the information in the 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 are obtained through an experimental calibration process.[21] The vertical

https://doi.org/10.1017/S0263574709990312 Published online by Cambridge University Press

lines in the 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.

### *1.3. Outline*

The rest of the paper is organized as follows: the single sensor methods proposed to calculate the feature position with respect to the robot are provided in Section 2. Section 3 provides details of the integrated sensor. Finally, Section 4 provides the experiments conducted to verify the algorithm and the results. Section 5 provides a discussion on the proposed method along with the conclusions.

## 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. For example, scale invariant feature transform (SIFT) uses rich visual information to derive a multidimensional descriptor of visual features.[22] This type of rich description is useful for associating features in stereo vision[23] and for homography estimation.[24] 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,[12] the research in vision based SLAM has moved to realtime monocular SLAM[16,25,26] implementations. In refs. [16, 25, 26] 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 accuracy 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. However, due to the high sensitivity to noise in robot velocity measurements, the uncertainty of the final calculated values can be extremely large and the resulting position calculations will be of limited use. This uncertainty problem in the position calculation is magnified at low robot velocities. Further, the optical flow based method cannot directly calculate the object position when the robot is making pure translational motion.

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.[20,27] 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.

### *2.1. 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 work for monocular vision based landmark localization, only the optical flow based method was investigated.

From the six degree of freedom general motion 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 ($\omega$), of the robot and focal length of the camera ($\lambda$) as follows:[28]

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

where $X_c$ is the distance to the feature in the direction of the heading velocity as shown in Fig. 2. Using the above equation and the camera model shown in Fig. 2 ($p/\lambda = Y_c/X_c$) the feature position with respect to the robot can be calculated by

$$\begin{aligned} X_c &= \frac{pv}{\dot{p}\lambda + (\lambda^2 + p^2)\omega}, \\ Y_c &= \frac{p^2 v\lambda}{\dot{p}\lambda + (\lambda^2 + p^2)\omega}. \end{aligned} \tag{2}$$

The covariance of the calculated position can be found using the first-order Taylor expansion of the feature position

https://doi.org/10.1017/S0263574709990312 Published online by Cambridge University Press

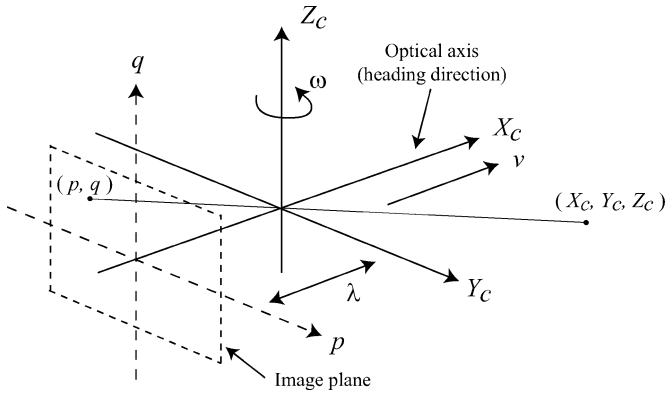


Fig. 2. The coordinate system and the imaging model of the camera.

$[X_c, Y_c]^T$. The covariance matrix of the position calculation can be obtained from

$$\Sigma_{X_c,Y_c} = J \cdot \mathrm{diag}[\sigma_p, \sigma_{\dot{p}}, \sigma_v, \sigma_\omega] \cdot J^T, \qquad (3)$$

where

$$J = \begin{bmatrix} \frac{\partial X_c}{\partial p} & \frac{\partial X_c}{\partial \dot{p}} & \frac{\partial X_c}{\partial v} & \frac{\partial X_c}{\partial \omega} \\ \frac{\partial Y_c}{\partial p} & \frac{\partial Y_c}{\partial \dot{p}} & \frac{\partial Y_c}{\partial v} & \frac{\partial Y_c}{\partial \omega} \end{bmatrix},$$

and $\sigma_p$, $\sigma_{\dot{p}}$, $\sigma_v$, and $\sigma_\omega$ 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 flow and feature positions is shown in Fig. 3. From Fig. 3 it is evident that at low optical flows the uncertainty increases regardless of the feature position

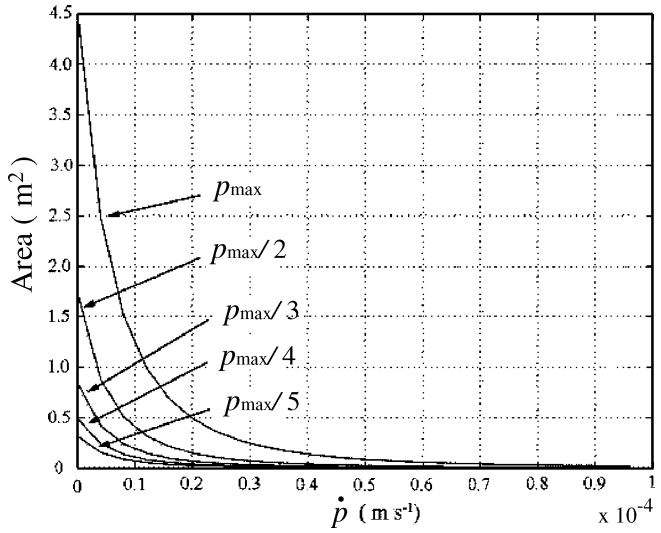


Fig. 3. Sensitivity of the uncertainty of the feature localization. The uncertainty is quantified by the area of the ellipse representing 95% confidence. $p_{\max} = 1.75$ mm, $v = 0.092$ m/s and $\omega = 4 \times 10^{-3}$ rad/s. ($\sigma_p = 10.9 \times 10^{-6}$ m, $\sigma_{\dot{p}} = 0.3 \times 10^{-4}$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

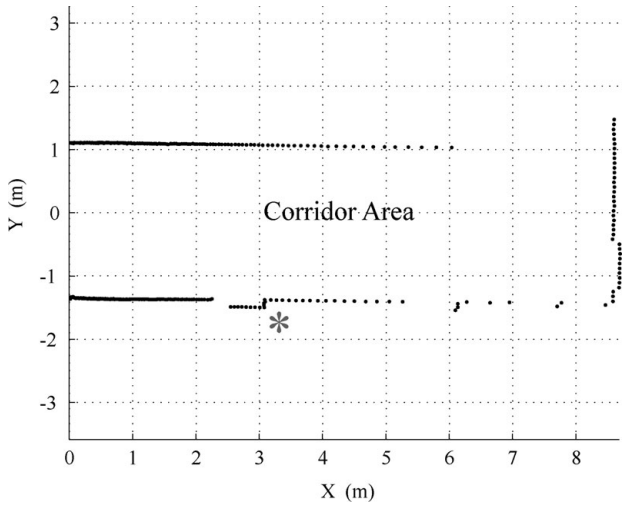


Fig. 4. A typical laser reading in an indoor environment where there is only one (*) direction invariant (corner) feature.

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 that 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.

### 2.2. *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.[29] 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 Fig. 4, certain environments such as long corridors might have only few or no directional invariant features. In such cases, reliable feature based, laser only SLAM implementations will not be possible unless higher level features such as lines are used.

## 3. Calibrated Laser-Vision Sensor

A camera is mounted on the laser range finder using a custom made bracket as shown in Fig. 5. The camera is mounted at the center of the laser range finder to maintain the coordinate transformation between the laser scanning plane and the camera coordinate system as simple as possible. The coordinate frames are defined as shown in Fig. 6. In our real setup the axes $z_l$ and $z_c$ coincide with each other (i.e., $a = 0$ and $b = 19$ cm). The laser range scanner is mounted on the center of the robot frame, thus the coordinate frame of the laser is also considered as that of the robot.

### 3.1. *Visual landmark detection*

Landmarks in the camera images can take several forms. The most common landmarks are the visually distinct corner

https://doi.org/10.1017/S0263574709990312 Published online by Cambridge University Press

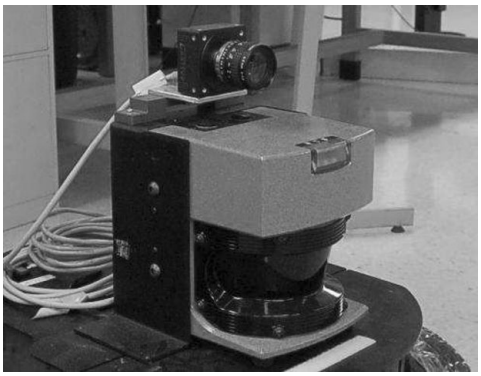

Fig. 5. The camera and the laser range sensor used in the experiments.

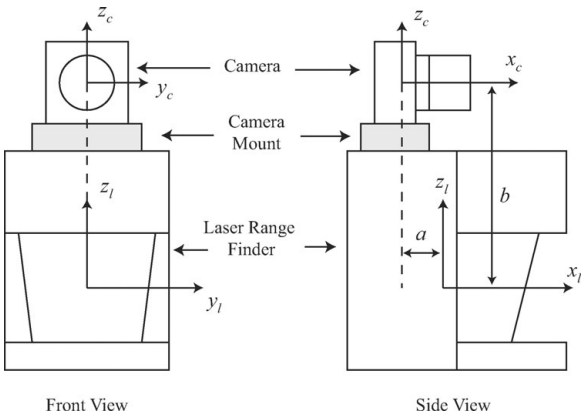


Fig. 6. Coordinate frames of calibrated laser-vision sensor.

features. Other visually salient landmarks include lines, arcs, and user-defined objects. In the proposed method visually salient vertical line features were detected in the captured images. Line features are the most robust in terms of detection accuracy and repeatability. In this work two algorithms have been evaluated for the detection of vertical lines in the images:

(1) Hough transform based method.
(2) Artificial corner feature based method.

Line detection algorithms based on the Hough transformation are 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,

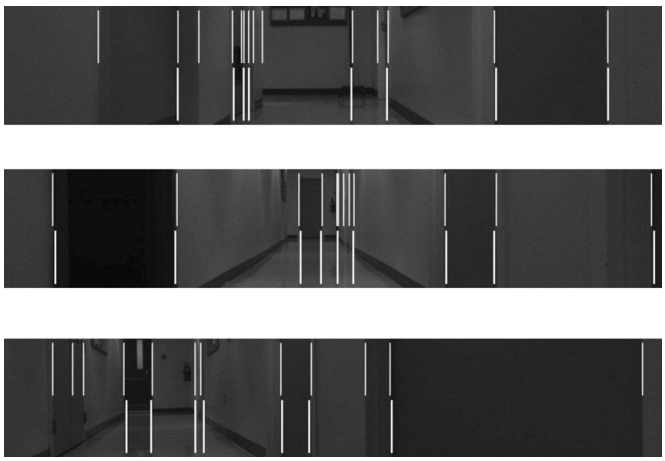

Fig. 8. Detected line features using Hough transformation and the corner based method.

a set of horizontal lines were superimposed on the original image as shown in Fig. 7. Then, all the resulting corner features were detected using a Harris corner detector[30] and are indicated by the white circles in Fig. 7.

This list of corner features is then searched for sets of features that are vertically aligned. If the number of features in a set is greater than the threshold value, then the average horizontal position of the features is identified as a consistent vertical line. Identified lines are marked with white line stubs at the bottom of the image frame shown in Fig. 7. A comparison of the two methods is shown in the Fig. 8 for three typical images that are 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 are detected using the 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 accumulates the evidence for lines in the whole region rather than for some sampled points in the image, as in the case with the corner based method.

### *3.2. 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. 6) and the errors in camera calibration. Although the camera is calibrated using standard

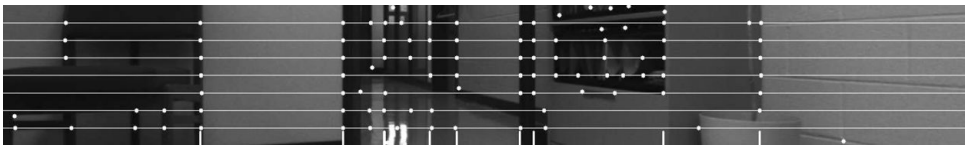

Fig. 7. Line feature detection using artificially generated corner features.

https://doi.org/10.1017/S0263574709990312 Published online by Cambridge University Press

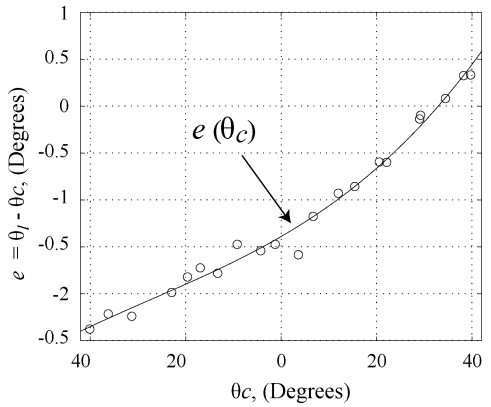


Fig. 9. Calibration curve for mapping between the field of view of the camera and the field of view of the laser range finder.

camera calibration techniques[1], the distortions especially at the edge of the images, contribute significantly to the errors.

The main objective of the sensor calibration method is to accurately map the field of view of the 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 ($\theta_c$) and in the laser range finder it is measured from the central laser scan ($\theta_l$). Thus, the error in the calibration can be calculated from $e = \theta_l - \theta_c$. As shown in Fig. 9, the error $e$ is approximated using a higher order polynomial $e(\theta_c)$ with respect to $\theta_c$. Thus for any new measurement in the image $\theta_c$, the corresponding mapping angle in the laser range finder can be calculated from $\theta_c + e(\theta_c)$. Similarly, the reverse mapping, the mapping of a reading in laser data onto the image, is also possible with the same data with a new calibration curve of $e(\theta_l)$ versus $\theta_l$.

### 3.3. Measurement model

The goal of defining a measurement model is to calculate the range to the landmark that has been detected by computer vision and then define its uncertainty. The bearing angle ($\theta_c$) of the detected landmarks (line features) can be calculated using a camera model with subpixel accuracy. A laser ranger provides a set of scanned readings 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. Therefore, using the coordinate transformation between the camera and the laser range finder along with the calibration information, the range to the line features can be calculated. Due to the resolution constraints in the laser data, the range value has to be interpolated from the data to increase its accuracy. This process of range interpolation is shown in Fig. 10. It should be noted that the coordinate

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

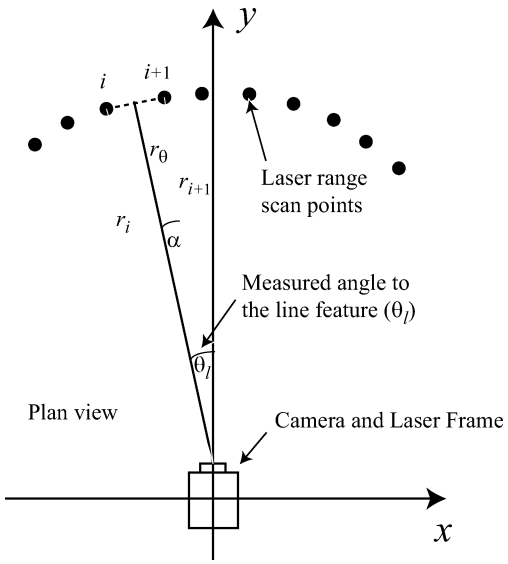


Fig. 10. Interpolation of the range to the line feature.

frame of the laser range data and the camera coincide with each other as the calibration is already applied to the bearing angle of the camera. Thus, in Fig. 10 bearing angle can be explicitly expressed as in the laser coordinate frame.

Assuming the resolution of the laser range scanner is at 0.5°, the range to the line feature can be calculated using the following interpolation:

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

Since the bearing to the feature is measured using the 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 increases 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 camera lens used 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

$$R = \mathrm{diag}\left[\ \sigma_r^2 \ \ \sigma_\theta^2\ \right], \quad (5)$$

where $\sigma_r$ and $\sigma_\theta$ are the standard deviations of the range and bearing measurement errors, respectively.

## 4. Experiments and Results

This section provides information about experiments that have been carried out to validate the suitability of the integrated sensor. Before the description of the experiments and their results, a key step in the selection of the detected visual landmarks has to be explained. In some cases, the visual landmark (the vertical line) would not cross the plane of the scanning laser. As an example, there could be visual features on protruding (or retracted) walls or objects on top

https://doi.org/10.1017/S0263574709990312 Published online by Cambridge University Press

Fig. 11. The curve constructed on the image plane by connecting the laser readings mapped from the laser coordinate frame to image frame.

of tables. In such cases the range to those landmarks cannot be guaranteed to be accurate. Thus the landmarks that do not intersect with the laser plane have to be removed from the list of detected landmarks before the calculation of the range to the landmarks. After the visual landmarks have been detected, the first step in the detection of such landmarks is the reverse mapping of the laser points onto the image using reverse sensor calibration (as described in the section 3.2) and coordinate transformation. Once the field of view of the two sensors is calibrated, the horizontal and vertical position ($p$ and $q$) of the laser point in the image can be calculated by using

$$p = \lambda r_i \tan(\theta_l),$$
$$q = \frac{\lambda b}{r_i \cos(\theta_l)},$$

where $\theta_l$ is the angle to the laser point from the vertical plane through the optical axis, $b$ is the vertical displacement of the camera and laser coordinates, and $r_i$ is the laser reading that is being mapped. Figure 11 shows the curve constructed from mapped laser readings.

In the next step the intersecting points between the vertical lines and the curve of the mapped laser readings are found. The vertical gradient in the neighborhood of each intersecting point can be calculated by a suitable gradient detector. Then the vertical landmarks corresponding to points with weak vertical total vertical gradients can be dropped. Although this method is able to remove most of the landmarks that do not intersect the laser plane, in rare cases two vertical aligned landmarks that belong to objects with different ranges could yield erroneous range information.

Two SLAM experiments were carried out to evaluate the fitness of the multisensor landmark detection and measurement method. In the first experiment the robot was driven through a regular office environment where it encountered narrow corridors, open office areas, and regular object clutter that are typical to an office environment. The robot travelled approximately 67 m making two loops through the office environment. In the second, longer experiment the robot was driven through the main corridors in a university building where the corridors were considerably wider compared to the first experiment. The robot traveled approximately 148 m while looping one and half times in the same environment. The experiments were carried out using a Pioneer 3AT robot equipped with a SICK® laser range finder and a camera with a regular off-the-shelf lens. During this experiment the laser range data, images from the camera and odometry data were logged at regular spatial intervals (20 cm or 2° apart, whichever occurs first). The noise levels that have been used in the map estimation and localization are listed in Table I.

Figure 12 shows process feature detection and localization using an integrated sensor for a typical set of image and laser scan data. As can be seen from Fig. 12 the laser range finder can only detect the landmark at location $C$ (using the intersection of two lines) while computer vision can be effectively used to detect other landmarks that can only be detected using a camera (at locations $A$ and $B$).

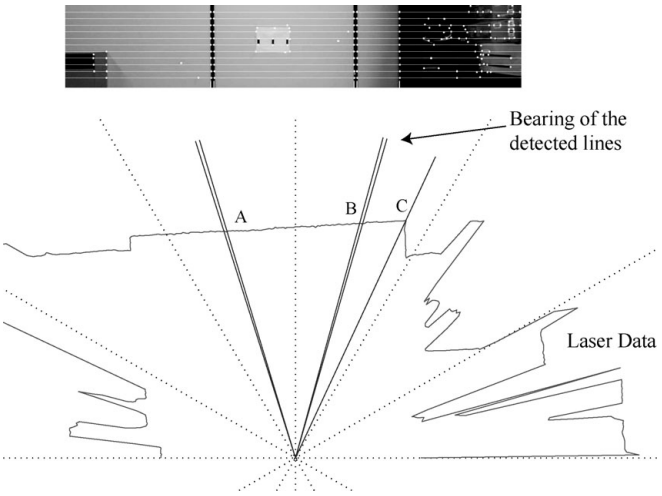


Fig. 12. The landmarks detected by the camera and their bearing angle superimposed on laser readings.

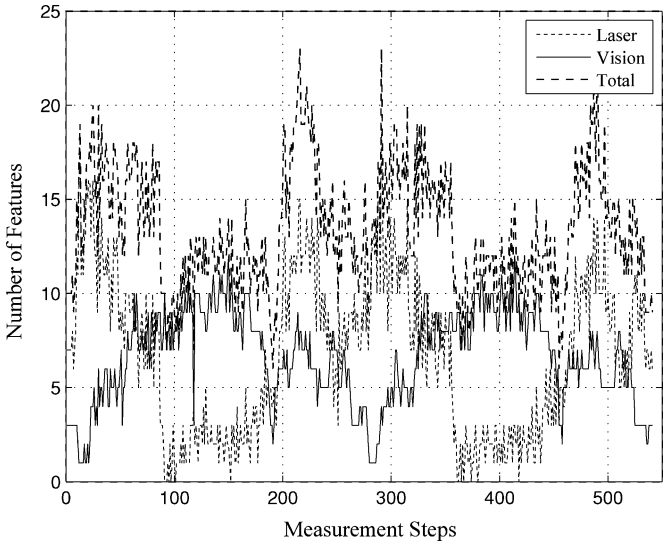


Fig. 13. Number of landmark features detected by vision and laser system.

https://doi.org/10.1017/S0263574709990312 Published online by Cambridge University Press

Table 1. The measurement of noise levels of the respective sensors that is used in the SLAM.

| Quantity | Measurement noise |
|---|---|
| Range to the landmark (cm), $\sigma_r$ | 5.0 |
| Bearing to the landmark (degrees), $\sigma_\theta$ | 1.0 |
| Robot heading velocity (cm/s), $\sigma_v$ | 0.5 |
| Robot rotational velocity (degrees/s), $\sigma_\omega$ | 0.025 |

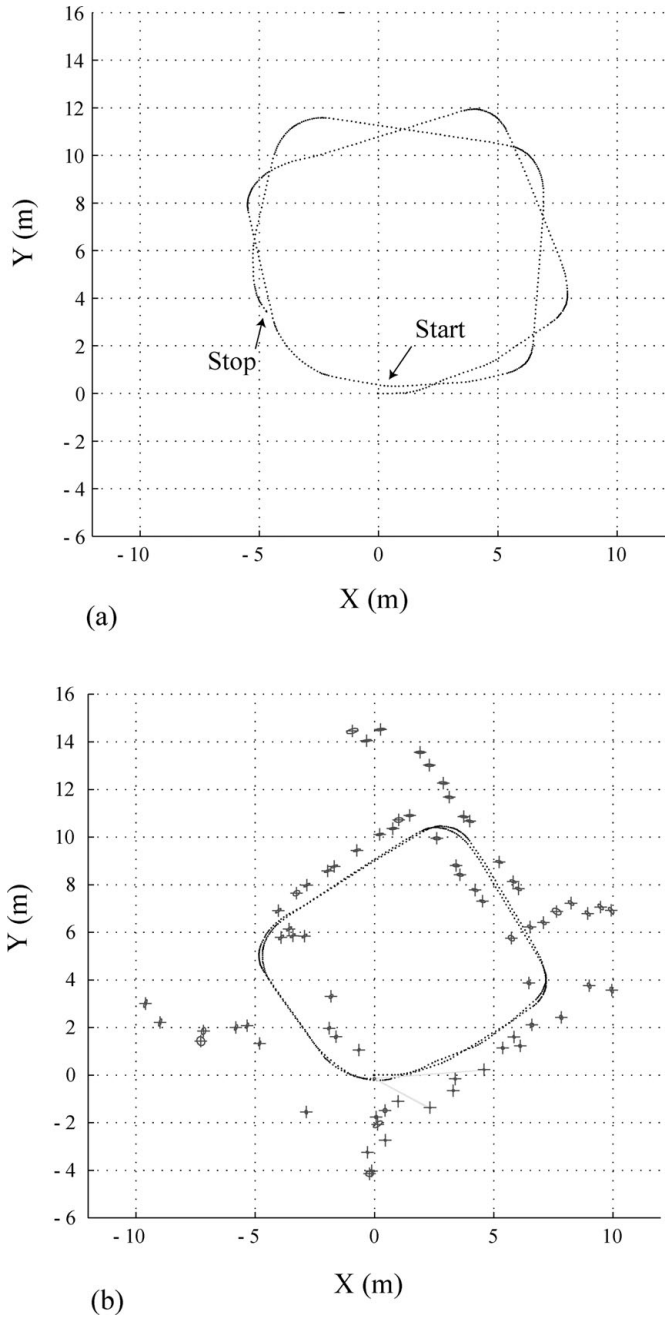


Fig. 14. Results of a localization and mapping of the first robot run: (a) with odometry, and (b) using EKF and vision-laser landmark localization (see the attached video file for incremental map building along with the current image frame).

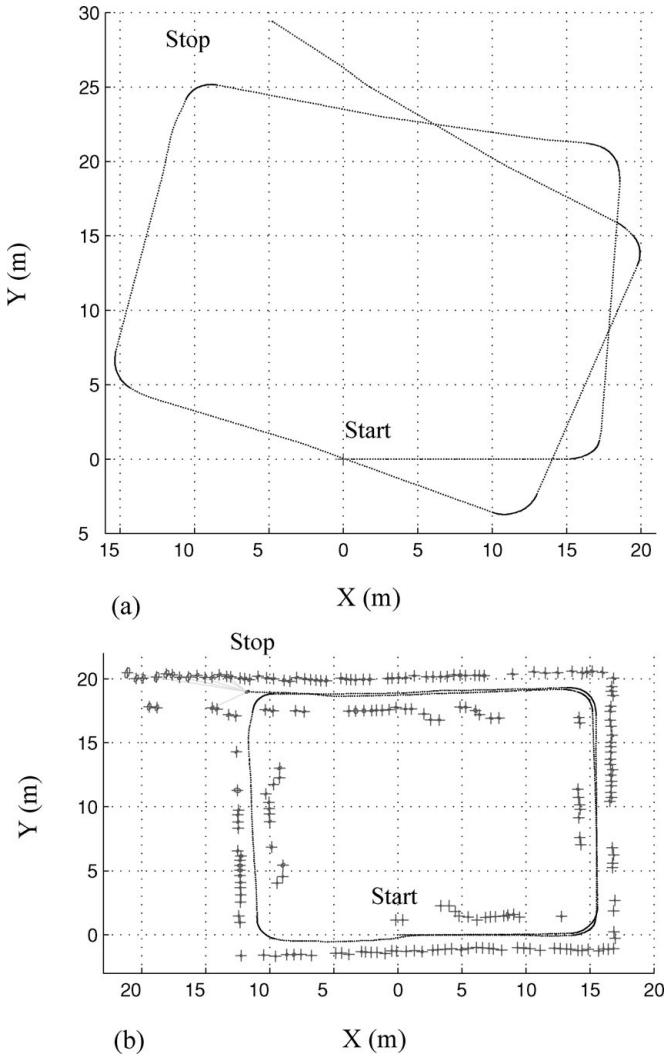


Fig. 15. Results of a localization and mapping of the second robot run: (a) with odometry, and (b) using EKF and vision-laser landmark localization (see the attached video file for incremental map building along with the current image frame).

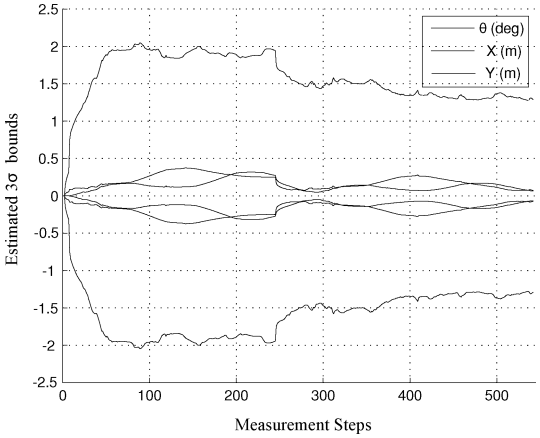


Fig. 16. $3\sigma$ bounds of the localization errors of the first experiment.

As discussed previously, the protruding features in the laser data can be detected as landmarks in the laser data. Figure 13 shows a comparison between the number of landmarks that can be detected in laser data and in image data during the first experiment. It is clearly evident that there are significant periods when image features outnumber the laser based landmarks. Further it should be noted that when there is a low number of visual features there is a significantly higher number of laser based landmarks. Although the results are purely specific to a given environment, the total number of

https://doi.org/10.1017/S0263574709990312 Published online by Cambridge University Press

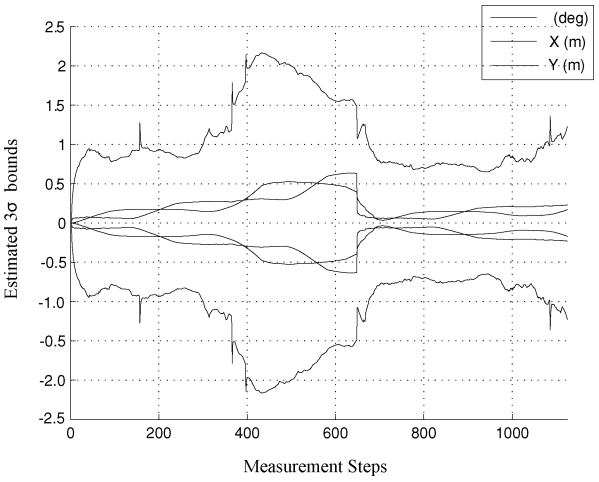


Fig. 17. $3\sigma$ bounds of the localization errors of the second experiment.

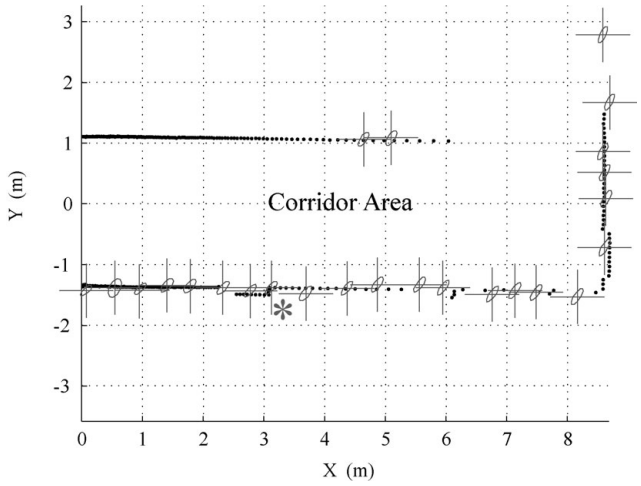


Fig. 19 Visual features detected in SLAM overlayed on the laser data from the same area of the environment. This further illustrates the fact that computer vision can be used to detect greater number of landmarks compared to laser data.

landmarks can be improved using the proposed method in addition to the laser based landmarks.

After the landmarks are detected and located using laser data and images, the data is processed off-line using the EKF method. The global nearest neighbor (GNN) algorithm was used for the data association. In the first experiment a map consisting of 71 landmarks has been built at the end of the run (Fig. 14(b)). Figure 14(a) shows the robot path using pure odometry data. The 95% confidence bounds of the errors in the robot pose estimate are shown in Fig. 16. In Fig. 16 it is possible to observe the effects in the robot position estimation around the midway point of the robot run. In the second experiment the robot constructed a map (as shown in Fig. 15) that contains 271 landmarks. Although the robot travels a considerably longer distance in a different environment compared to the first experiment, a similar pattern can be observed in robot position uncertainty when loop closing as shown in Fig. 17. The accuracy of the EKF based SLAM algorithm was enough to robustly close the loop in the long run, but during the initial steps of the loop closing there were erroneous data associations. The ability of the algorithm to recover from the initial errors data association can be mainly attributed to the large size of the map compared to the number of erroneous data associations. The data from these experiments were processed off-line using MATLAB software. The EKF SLAM algorithm is known to have a time complexity of $O(N^2)$ where N is the number of landmarks in the map.[31] Figure 18 shows the processing time for the second experiment. As shown, the processing time EKF algorithm increases with the increase in the number of landmarks in the map and levels of when the number of landmarks becomes constant when the robot closes a loop in the environment. The processing time for data association depend of the number of landmarks in the current observation as well as the overall number of landmarks in the map. Initially, data association time increases with the increase in the number of landmarks in the map but after loop closing the variances are mainly govern by the number of landmarks in each observation.

## 5. Conclusion

In this paper it is shown that computer vision and a 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 the EKF based SLAM method to build maps of typical indoor environments. Through the detection of greater number of landmarks the integrated sensor increases the reliability of the SLAM process. In typical indoor environments cameras can be used to detect significantly higher number of landmarks as shown in the Fig. 19 compared to laser data. Thus the proposed

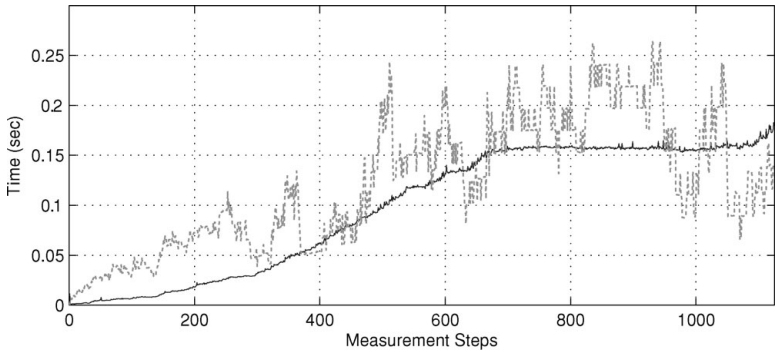


Fig. 18 The processing time for EKF (solid line) and the GNN data association (dashed line) of second experiment.

https://doi.org/10.1017/S0263574709990312 Published online by Cambridge University Press

method minimizes the risk of robot running without any observable landmarks and also improves the quality of the data of the landmarks by using the appropriate sensor for detection and position calculation. However, when robot is in a featureless environment, in adverse lighting conditions or in smoke or dust filled environment, the proposed method will have a very low reliability due to the limitations in the vision system. Future extensions of this work include the use of more accurate sensor uncertainty modeling especially in the case of bearing angle to the landmark and experimentation in large looping environments with possible submapping. Some visual landmarks are present in the form of wide vertical strips, in which two side edges are detected as vertical lines. Thus they are recognized as two landmarks and the SLAM algorithm will attempt to initialize them as such in the map. But since they are often physically close together only one of them will be initialized into the map. Further, when the robot is away from visual features as described above, the line detection algorithm will often detect a single line due to the limitation in the resolution of the camera. However, as the robot gets closer to the object, it will appear as two landmarks and the data association algorithm will have to decide the best edge to be assigned to the feature that is already in the map. Thus a better sensor model that can handle this type of composite objects is required.

## Acknowledgments

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 in this paper.

## References

1. J. E. Guivant, F. R. Masson and E. M. Nebot, "Simultaneous localization and map building using natural features and absolute information," *Robot. Auton. Syst.* **40**(2), 79–90 (2002).
2. W. Y. Jeong and K. M. Lee, "Visual SLAM with Line and Corner Features," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China (Oct. 2006) pp. 2570–2575.
3. E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser and P. Sayd, "Monocular Vision Based SLAM for Mobile Robots," *Proceedings of International Conference on Pattern Recognition*, Hong Kong (Aug. 2006) pp. 1027–1031.
4. J. V. Miro, G. Dissanayake and W. Zhou, "Vision-Based Slam Using Natural Features in Indoor Environments," *Proceedings of International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (Dec. 2005) pp. 151–156.
5. P. E. Rybski, S. I. Roumeliotis, M. Gini and N. Papanikolopoulos, "Appearance-Based Minimalistic Metric Slam," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, USA (Oct. 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," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai, Japan (Oct. 2004) pp. 3424–3429.
7. H. M. Gross, A. Koenig and St. Mueller, "Omniview-Based Concurrent Map Building and Localization Using Adaptive Appearance Maps," *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Hawaii, USA (Oct. 2005) pp. 3510–3515.
8. J. M. Porta, J. J. Verbeek and B. Krose, "Enhancing Appearance-Based Robot Localization Using Sparse Disparity Maps," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, USA (Oct. 2003) pp. 980–985.
9. B. J. A. Krose, N. Vlassis and R. Bunschoten, "A probabilistic model for appearance-based robot localization," *Image Vis. Comput.* **12**(6), 381–391 (Apr. 2001).
10. I. Ulrich and I. Nourbakhsh, "Appearance-Based Place Recognition for Topological Localization," *Proceedings of the IEEE International Conference on Robotics and Automation*, San Francisco, USA (Apr. 2000) pp. 1023–1029.
11. A. J. Davison, Mobile Robot Navigation Using Active Vision. *PhD thesis* (Oxford, UK: University of Oxford, 1998).
12. A. J. Davison and D. W. Murray, "Simultaneous localization and map-building using active vision," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 865–880 (2002).
13. P. Saeedi, P. D. Lawrence and D. G. Lowe, "Vision-based 3d trajectory tracking for unknown environments," *IEEE Trans. Robot.* **22**(1), 119–136 (Feb. 2006).
14. S. Se and P. Jasiobedzki, "Photo-Realistic 3D Model Reconstruction," *Proceedings of the IEEE International Conference on Robotics and Automation*, Orlando, USA (May 2006) pp. 3076–3082.
15. S. Se, D. Lowe and J. J. Little, "Vision-based global localization and mapping for mobile robots," *IEEE Trans. Robot.* **21**(3), 364–375 (Jun. 2005).
16. A. J. Davison, "Real-Time Simultaneous Localization and Mapping with a Single Camera," *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Nice, France (Oct. 2003) pp. 1403–1410.
17. J. M. M. Montiel, J. Civera and A. J. Davison, "Unified Inverse Depth Parametrization for Monocular SLAM," *Proceedings of the Robotics, Science and Systems*, Pennsylvania, USA (Aug. 2006).
18. J.-Y. Bouguet and P. Perona, "Visual Navigation Using a Single Camera," *Proceedings of the International Conference on Computer Vision*, Boston, USA (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 Trans. Robot. Autom.* **17**(3), 229–241 (Jun. 2001).
20. J. A. Castellanos, J. Neira and J. D. Tardos, "Multisensor fusion for simultaneous localization and map building," *IEEE Trans. Robot. Autom.* **17**(6), 908–914 (Dec. 2001).
21. D. Ortin, J. Neira and J. M. M. Montiel, "Relocation Using Laser and Vision," *Proceedings of the IEEE International Conference on Robotics and Automation*, Barcelona, Spain (Apr. 2005) pp. 1505–1510.
22. D. G. Lowe, "Object Recognition from Local Scale Invariant Features," *Proceedings of the Seventh International Conference on Computer Vision (ICCV99)*, Kerkyra, Greece (Sep. 1999) pp. 1150–1157.
23. S. Se, D. Lowe and J. Little, "Vision-Based Mobile Robot Localization and Mapping Using Scale Invariant Features," *Proceedings of the IEEE International Conference on Robotics and Automation*, Seoul, Korea (May 2001) pp. 2051–2058.
24. M. Brown and D. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.* **74**(1), 59–73 (Aug. 2007).
25. L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira and J. D. Tardos, "Mapping Large Loops with a Single Hand-Held Camera," *Proceedings of the Robotics Science and Systems*, Atlanta, USA (2007).
26. A. J. Davison, I. D. Reid, N. D Molton and O. Stasse, "Monoslam: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(6), 1052–1067 (2007).
27. R. Martinez-Cantin, J. A. Castellanos, J. D. Tardos and J. M. M. Montiel, "Adaptive Scale Robust Segmentation for

2d Laser Scanner," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China (Oct. 2006) pp. 796–801.
28. S. Hutchinson, G. D. Hager and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Autom.* **12**(5), 651–670 (1996).
29. D. Schulz, W. Burgard, D. Fox and A. B. Cremers, "Peopletracking with mobile robots using sample-based joint probabilistic data association filters," *Int. J. Robot. Res.* **22**(2), 99–116 (Feb. 2003).
30. J. Shi and C. Tomasi, "Good features to track," *IEEE Conf. Comput. Vis. Pattern Recognit.* (Jun. 1994) pp. 593–600.
31. U. Frese, "Treemap: An $o(\log(n))$ algorithm for indoor simultaneous localization and mapping," *Auton. Robots* **21**(2), 103–122 (Sep. 2006).

https://doi.org/10.1017/S0263574709990312 Published online by Cambridge University Press