A Solution for SLAM through Augmenting Vision and Range Information

Ali A. Aghamohammadi, *Student Member*, *IEEE*, Amir H. Tamjidi, Hamid D. Taghirad, *Member*, *IEEE*

Advanced Robotics and Automated Systems (ARAS),
Department of Electrical and Computer Engineering,
K.N. Toosi University of Technology

***Abstract*— This paper proposes a method for augmenting the information of a monocular camera and a range finder. This method is a valuable step towards solving the SLAM problem in unstructured environments free from problems of using encoders' data. Proposed algorithm causes the robot to benefit from a feature-based map for filtering purposes, while it exploits an accurate motion model, based on point-wise raw range scan matching rather than unreliable feature-based range scan matching, in unstructured environments. Moreover, robust loop closure detection procedure is the other consequence of this method. Experiments with a low-cost IEEE 1394 webcam and a range finder illustrate the effectiveness of the proposed method in drift-free SLAM at loop closing motions in unstructured environments.**

I. Introduction

The ability to construct a map of an environment and simultaneously using this map to estimate the location of the robot within this map (in the absolute coordinate) is an important prerequisite for fulfilling various functions of an autonomous mobile robot. Solving this problem, which is commonly known as Simultaneous Localization and Mapping (SLAM), has received immense attention from mobile robotics research community ([1],[2],[3]), in the last 20 years. One of the main objectives of solving this problem is to make it possible for a mobile robot to operate autonomously for a long period of time in an unknown unstructured environment.

Encoder data can be used to establish the motion model of a mobile robot [4], but traditional encoder-based dynamic modeling has its own well-known drawbacks. Therefore, Laser Range Finder (LRF) proposed to be used instead of the encoder to establish the motion model. There are two major approaches in the range scan matching: *feature-based* [5] and *point-wise scan matching* [6]. Extracting robust and reliable features in an unstructured environment is a knotty, if possible, task and due to that it is inevitable to adopt a point-wise range scan matching to provide acceptable motion estimation.

We believe that Kalman Filtering-based (KF-based) methods are efficient frameworks to tackle the SLAM problem and it is mainly because of their power in computing, storing, and representing the accurate correlations between the elements of the system's state vector. For exploiting this framework, there is a need to a feature-based method for environment representation as it enables us to augment robot's state vectors with features' state vectors and use the augmented state vector later in the KF-based framework.

As mentioned before, extracting reliable features from a range scan, in unstructured environments, is an intricate task. Various sensors can be mounted on a mobile robot to gather environmental data. However, especially in an unstructured environment and from feature extraction and data association point of view camera provides the most informative data. Many of data association algorithms use the traditional gating criterion to distinguish one feature from another, solely based on the innovation gate distance metric. Exclusive use of this criterion does not exploit other landmark attributes that may be useful for disambiguation. Particularly, in visual features, there exist many attributes for each feature, which makes them more distinctive and makes the association process very robust. For example, SIFT descriptors [7] or image patches [8] around the extracted features are of such attributes for each feature, which can be used as the additional data association metrics for making the association process more robust.

In the proposed method a sparse and persistent map of visual features is constructed and it is fused with the information, obtained from the point-wise range scan matching, for attaining the drift-free solution for SLAM problem in loop-closing motions of a mobile robot in unstructured environments. This fusion allows the robot to benefit from both an accurate motion model and an informative feature-based map, to efficiently solve the SLAM problem and correct the drifts, when confronts with the existing features in the map.

Since monocular camera is a bearing-only sensor, features' depths are initially unknown. There exist another fusion process between the LRF and camera's data in feature initialization level, in which the depth initialization process of some of features is carried out through LRF data. The schematic, shown in figure 1, depicts the data flow in the proposed method, whose components will be described in the following sections.

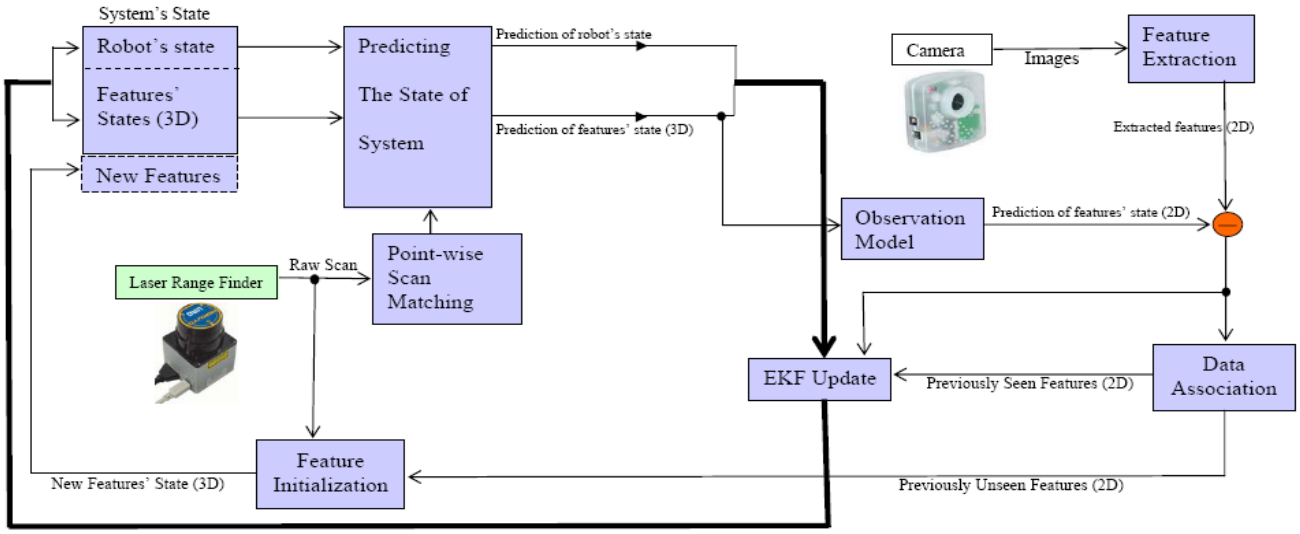

Figure 1: Data flow schematic of the proposed method.

II. PROBABILISTIC FRAMEWORK

The main map, used here for SLAM, is a feature-based map, updated dynamically by the Extended Kalman Filter. The probabilistic state estimates of the robot and features are updated during robot motion and feature observation. When new features are observed the map is enlarged with new states.

Mathematically, the map at time step k is represented by a system's state vector $\boldsymbol{x}_k$ and its covariance matrix $cov(\boldsymbol{x}_k)$.Here, *system* is the indication of collection of the robot and the environment. Thus, system's state vector is composed of the stacked state estimates of the robot and the environment's features.

$$\hat{\mathbf{x}}_{k+1} = \begin{pmatrix} \hat{\mathbf{r}}_{k+1} \\ \hat{\mathbf{F}}_k \end{pmatrix} \quad , \quad \hat{\mathbf{F}}_k = \begin{pmatrix} \hat{\mathbf{f}}_1 & \hat{\mathbf{f}}_2 & \cdots & \hat{\mathbf{f}}_n \end{pmatrix}^t \tag{1}$$

$\boldsymbol{F}$ is a vector, composed of the state vectors of existing features in the map, and $\boldsymbol{r}$ consists of robot's position, $\boldsymbol{p}$, and quaternion-based attitude, q. Quaternion-based attitude representation leads to a singularity-free, non-redundant, and efficient attitude computation.

$$\hat{\mathbf{r}}_k = \begin{pmatrix} \hat{\mathbf{p}}_k \\ \hat{q}_k \end{pmatrix} \tag{2}$$

$\boldsymbol{r}$ describes the robot-fixed coordinate frame's pose (position and orientation) with respect to global coordinate frame $\{G\}$. Indeed, robot-fixed coordinate frame, $\{R\}$, is the same as camera-fixed coordinate frame $\{C\}$. Also, we have

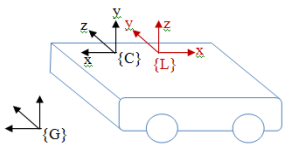

Figure 2: camera-fixed coordinate, $\{C\}$, is shown in black and LRF-fixed coordinate frame, $\{L\}$, is shown in red.

an LRF-fixed coordinate frame $\{L\}$, whose relationship to the camera frame $\{C\}$ is known and constant. Thus, global coordinate frame is defined as the camera-fixed frame at the initial location of the robot.

Monocular camera is a bearing-only sensor, and for estimating the depth of a feature, camera has to move and observe it from different locations. The variations of the azimuth and elevation of the feature, in camera-fixed coordinate frame, allow the feature's depth to be estimated. On the ground of this fact, we have placed the camera on the robot so that it can observe the scene, at the left side of the robot, rather than the robot's forward scene. Proposed situation for the camera leads to a high amount of variations in the azimuth and elevation of the features during the robot's motion. Range finder is placed on the robot so that the origin of scans locates in the nearest possible point to the camera's optic center; this placement not only allows the LRF to accomplish the motion model, but also allows it to provide valuable information regarding the depth of visual features, located on the scanned slice of the scene by the LRF.

In the outdoor environments, most of the features are distant with respect to the robot, while also nearby features can be observed in the robot's field of view. Therefore, the initial uncertainty region for a feature has to cover the huge range depth. Huge uncertainties can considerably decrease the performance of the EKF through the nonlinearity effects of the measurement equation in this framework. Inverse depth parameterization [9] for coding features can handle this problem by providing high degree of linearity in the measurement equation.

Thus, all the extracted features are coded using the inverse depth parameterization, in which each individual feature is defined by the state vector with 6 elements:

$$\mathbf{f}_i = (\mathbf{r}_{c_i} \;\; \theta_i \;\; \phi_i \;\; \rho_i) \qquad \rho_i = 1/d_i \tag{3}$$

θ_i, φ_i, *and* d_i are the *i'th* feature's azimuth, elevation, and depth, coded in absolute frame, and $\boldsymbol{r_c}$ is the 3D location of

the camera, which is inferred at the time instant, at which the feature comes into the robot's field of view for the first time.

As a result, the directional vector along connecting line between camera optical center and 3D location of the feature, at feature's first observation step, can be calculated as follows:

$$\mathbf{m}_i = [\cos(\phi_i)\sin(\theta_i) \quad -\sin(\phi_i) \quad \cos(\phi_i)\cos(\theta_i)]^t \tag{4}$$

Thus, 3D position of *i'th* feature can be computed through:

$$\mathbf{f}_i^w = \mathbf{r}_{c_i} + \left(1/\rho_i\right)\mathbf{m}_i \tag{5}$$

III. MOTION PREDICTION

There exist many methods to provide an estimate of the robot's displacement. Traditional encoder-based dead reckoning methods have well-known disadvantages and we try to circumvent them by the information obtained from other sensors. [8] has proposed a constant-velocity motion model to address the SLAM problem using single monocular camera. This model certainly appears appropriate for a pure-vision SLAM with a hand-held camera with smooth motion; however, a significant limitation of this model is that the large uncertainties are needed to make the model able to cope with velocity variations, and actually it needs many accurate observations to bound the uncertainty and to provide precise estimates. Nevertheless, a range finder can solve this problem by providing accurate displacement estimation. As previously mentioned, in unstructured environments, it is inevitable to adopt the point-wise scan matching methods for motion prediction, which compare raw scan data. Here, popular ICP [6] method is adopted for addressing the motion prediction. The covariance of matching scans is calculated as follows [10]:

$$\mathbf{\Sigma}_D = s^2\left(\mathbf{A}^t\mathbf{A}\right)^{-1} \tag{6}$$

where

$$\mathbf{A} = \begin{pmatrix}\mathbf{A}_1\\ \vdots \\ \mathbf{A}_m\end{pmatrix}, \mathbf{A}_i = \begin{pmatrix}1 & 0 & -0.5e_{i2}\\ 0 & 1 & 0.5e_{i1}\end{pmatrix}, s^2 = \frac{\left(\mathbf{E}-\mathbf{AC}\right)^t\left(\mathbf{E}-\mathbf{AC}\right)}{2m-3} \tag{7}$$

in which $\boldsymbol{e_i}=(e_{i1} \; e_{i2})^t$ is the residual related to *i'th* pair, matched during current iteration of the ICP. If $\boldsymbol{g_i}$ and $\boldsymbol{g'_i}$ are two points from two consecutive scans, which are successfully matched with each other, the residual resulted from their matching is as follows:

$$\mathbf{e}_i = \left(\mathbf{g}_i - (\mathbf{R}\mathbf{g}'_i + \mathbf{\Delta t})\right), \mathbf{E} = \begin{pmatrix}\mathbf{e}_1\\ \vdots \\ \mathbf{e}_m\end{pmatrix}, \mathbf{C} = \left(\mathbf{A}^t\mathbf{A}\right)^{-1}\mathbf{A}^t\mathbf{E} \tag{8}$$

m is the number of points, have been matched with each other.

Let $\boldsymbol{p}_k$ denotes the camera's position and ${}^L_G q_k$ denotes the quaternion, represents the orientation of camera-fixed coordinate frame, *{C},* with respect to global frame, *{G},* at time step k. If $\boldsymbol{D}$ and $\Delta\theta$ represent the computed 2D displacement and rotation of the LRF, respectively, through the range scan matching, between time step *k* and *k+1*, robot's, i.e. camera's, next 3D position and orientation can be computed through the below equations:

$$\hat{\mathbf{p}}_{k+1} = \hat{\mathbf{p}}_k + \mathbf{C}({}^L_G\hat{q}_k)\times\mathbf{A}\times\mathbf{D} \tag{9}$$

$${}^L_G\hat{q}_{k+1} = {}^L_G\hat{q}_k \otimes \Delta q \;\; ; \;\; \Delta q = quat(\Delta\theta, {}^L z) \tag{10}$$

In which, $\boldsymbol{A}$ transforms the 2D displacement in LRF-fixed frame to the 3D displacement in camera-fixed coordinate. $\boldsymbol{C(.)}$ provides the equivalent rotation-matrix with its input quaternion vector, and $quat(\Delta\theta, {}^L z)$ returns the equivalent quaternion, in the camera-fixed coordinate, with $\Delta\theta$ rotation angle around ${}^L z$, z axis of LRF-fixed coordinate frame. $\otimes$ represents the quaternion product operator.

In order to propagate the uncertainty of the range scan matching on the uncertainty of the pose prediction, the derivatives of the predicted state to the stochastic parameters in the right-hand side of equations 9 and 10 are required

$$J_{p1} = \frac{\partial\mathbf{r}_{k+1}}{\partial\mathbf{r}_k} \;\; , \;\; J_{p2} = \begin{bmatrix}\frac{\partial\mathbf{r}_{k+1}}{\partial D} & \frac{\partial\mathbf{r}_{k+1}}{\partial\Delta\theta}\end{bmatrix} \tag{11}$$

Taking into account the statistical independency between robot pose at time step *k*, and, the predicted displacement between time steps *k* and *k+1*, the covariance matrix for the predicted pose can be calculated as follows:

$$\mathrm{cov}(\mathbf{r}_{k+1}) = J_{p1}cov(\mathbf{r}_k)J_{p1}^t + J_{p2}\,\mathrm{cov}(X)J_{p2}^t \tag{12}$$

where $X = (\boldsymbol{D}\;\Delta\theta)^t$ is the displacement vector.

In view of the fact that robot's displacement does not affect features' state, prediction of the system's state vector and its covariance takes the below form:

$$\hat{\mathbf{x}}_{k+1}^{(-)} = \begin{pmatrix}\hat{\mathbf{r}}_{k+1}\\ \hat{\mathbf{F}}_k\end{pmatrix},$$

$$\mathrm{cov}(\mathbf{x}_{k+1}^{(-)}) = \begin{pmatrix}\mathrm{cov}(\mathbf{r}_{k+1}) & \mathrm{cov}(\mathbf{r}_k,\mathbf{F}_k)J_{p1}^T\\ J_{p1}\,\mathrm{cov}(\mathbf{r}_k,\mathbf{F}_k) & \mathrm{cov}(\mathbf{F}_k)\end{pmatrix} \tag{14}$$

IV. MEASUREMENT MODEL

The main map, constructed in this method, is a sparse but persistent map of visual features, which can be simply referenced in the state-based representation of the environment and permits the method to correct long-term drifts in loop closures. Visual signal is rich in information

content, and comparing with LRF data, it is more likely to extract salient features from visual signal. Therefore, setting up the measurement model based on the visual features is beneficial and could considerably improve the robustness of the data association procedure and therefore enhance the accuracy of the EKF-SLAM.

The method, adopted for the feature extraction, is the Harris Corner Detector. We have made slight changes in the algorithm of Harris corner detector, so that the algorithm tends to choose the features as near as possible to horizontal center of the image. In other words, algorithm chooses the features based on their saliency and their distance to the horizontal center of the image. Figure 3 shows the features, extracted from a frame by this modified method.

Camera and range finder are calibrated offline [11], so that the slice of the environment, scanned by the LRF, coincides with the horizontal center of the image. Therefore, the depth of the features, lie on the horizontal center of the image, can be accurately inferred from the range scan data. Depth of the other features cannot be inferred from a single frame recorded by a monocular camera. Therefore, in initialization stage, depth of such features is initialized heuristically, and thus the initial uncertainty region of their depth, covers a huge range.

The problem, here, is that if such uncertainties are modeled by a standard Gaussian distributions in the Euclidean space, linearization error of the measurement equation in the EKF framework decreases the performance of the EKF estimation. The observation model based on the inverse depth parameterization can considerably decrease the resulted error from the nonlinearity of the observation equation in the EKF framework.

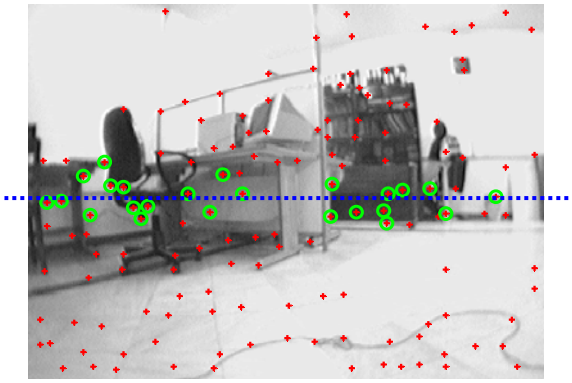

Figure 3: points, shown by the red +, have been extracted using Harris method. Extracted landmarks, enclosed by green circles, are considered as observed features. Blue dashed line is the horizontal center of image.

Actually, distorted pixel coordinates of features in recorded frame are the real observation of features, which have to be compared with predicted pixel coordinate of that feature. This prediction is accomplished through observation equation. Assuming that $\boldsymbol{h}^c=(h_x\ h_y\ h_z)$ is the connecting line between camera and an observed feature, in camera-fixed frame, undistorted pixel coordinates of the observed feature can be derived based on the pinhole camera model. Actually, the real camera lens distorts the image, and thus the radial distortion model, is adopted to remove this distortion.

$\boldsymbol{h}_c$ is the connecting vector between the camera and the observed feature, in the camera-fixed coordinate frame. So, for each of features, it can be easily predicted as follows:

$$\hat{\mathbf{h}}^c = {}^c_w\mathbf{C}\left(\hat{\mathbf{f}}_i^w - \hat{\mathbf{p}}\right) \tag{15}$$

$\hat{\mathbf{p}}$ is the prediction of the current position of the camera and ${}^c_w\mathbf{C}$ is the rotation matrix describes the relation of camera-fixed and global coordinate frames. As a result, *i'th* feature's prediction can be written as a function of system's state prediction.

$$\hat{\mathbf{z}}_{k+1}^i = h^i(\hat{\mathbf{x}}_{k+1}^{(-)}) \tag{16}$$

Total observation model, $\boldsymbol{h}$ and its covariance are obtained by considering all features in a single vector.

$$\mathbf{z} = \left(\mathbf{z}^1\ \mathbf{z}^2\ \cdots\ \mathbf{z}^m\right)^t\ ,\ \hat{\mathbf{z}} = \mathbf{h}\left(\hat{\mathbf{x}}_{k+1}^{(-)}\right)$$
$$\text{cov}(\mathbf{z}_{k+1}) = H_x\ \text{cov}(\mathbf{x}_{k+1}^{(-)})H_x^T\ ,\ H_x = \left.\frac{\partial h\ (x)}{\partial x}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k+1}^{(-)}} \tag{17}$$

The important attribute of this method is that, while the uncertainty of the depth of some features are huge, and while the algorithm cannot exploit them for estimating the translation of the camera, method can utilize them for estimating the camera's orientation. Moreover, due to the accurate motion model, presence of the features with known depths (through LRF initialization), and the manner of camera placement on the robot, the depth of other features are accurately estimated much faster than previous methods, which tried to estimate the depth of the features through the information of a monocular camera. Thus, most of such features can also effectively contribute in translation computing, usually few steps after their initialization.

V. Data Association and Filtering

After predicting the existing features in the map relative to the robot, data association procedure has to associate the extracted features from the current frame with the predicted ones. SIFT [7] is a robust and accurate method with respect to feature extracting and matching but its prohibitive computational cost in extracting features restricts its applicability in the on-line applications. Thus, we only use the SIFT descriptor for matching and Harris corner detector is adopted for extracting features as the measurements of the EKF framework.

At the initialization stage of each feature, its associated descriptor is stored. At the data association stage, the descriptor of each predicted feature, whose prediction remains within the image limits, is compared with the descriptors of other extracted features, from the observed image. Finally, the pairs resulted from the SIFT matching, which can pass the gating test, are considered as the matched pairs. The gating test for each pair is performed as follows:

Every pairs, resulted from the SIFT matching defines a residual. Residual for the *i'th* pair at time step *k+1* is:

$$\mathbf{res}_{k+1}^{i} = \left(\mathbf{z}_{k+1}^{i} - \hat{\mathbf{z}}_{k+1}^{i}\right) \tag{18}$$

Based on these residuals, the Mahalanobis distance is computed to perform a gating test and reject spurious associations.

$$\mathbf{res}_{k+1}^{i} \Sigma_{res} \mathbf{res}_{k+1}^{i\ t} < \gamma$$
$$\Sigma_{res} = \mathrm{cov}(\mathbf{z}_{k+1}^{i}) + \mathrm{cov}(\hat{\mathbf{z}}_{k+1}^{i}) \tag{19}$$

where γ is the 95-percentile of the χ_2^2 (Chi-square) distribution.

Exploiting obtained observation model, state estimate can be updated using the Extended Kalman Filter equations as follows [12]:

$$\hat{\mathbf{x}}_{k+1}^{(+)} = \hat{\mathbf{x}}_{k+1}^{(-)} + K_{k+1}[\mathbf{z}_{k+1} - \mathbf{h}(\hat{\mathbf{x}}_{k+1}^{(-)})]$$
$$\mathrm{cov}(\mathbf{x}_{k+1}^{(+)}) = \mathrm{cov}(\mathbf{x}_{k+1}^{(-)}) - K_{k+1} H_x \mathrm{cov}(\mathbf{x}_{k+1}^{(-)})$$
$$K_{k+1} = \mathrm{cov}(\mathbf{x}_{k+1}^{(-)}) H_x^T [H_x \mathrm{cov}(\mathbf{x}_{k+1}^{(-)}) H_x^T + \mathrm{cov}(\mathbf{z}_{k+1})]^{-1} \tag{20}$$

VI. Results

Proposed method has been implemented on the Melon, a mobile robot, which is equipped with a low-cost IEEE 1394 webcam with a wide angle lens, which can be seen in figure 1, and a Hokuyo URG_X002 range scanner

In the experiments, the robot was programmed to move along a 240 centimeter loop, along a square with 60 cm side length, in an unstructured environment. In this exploratory motion robot explores new areas before closing the loop, at the end of the motion. All of the features are distant and their depths, relative to robot, are more than 2 meters. The path, traveled by the robot, has been divided into 8 parts: Four 60 cm forward motions and four 90 degree rotations, alternately.

For the sake of studying the algorithm's behavior in loop-closing motions, robot is operated to exactly return and stop at its start location, at the end of the motion; due to that, 8th part has became longer than the other parts. First of all, the results of a localization using the ICP method, only based on the range scans' data, is shown in figure 4. Actually, subplots show the estimated location of the robot along x and y axes and its rotation about z axis of the LRF-fixed coordinate frame. Aforementioned parts are depicted by red double-head arrows, at the top of subplots, which are separated by red dashed vertical lines. Parts, specified by the odd numbers, relate to forward motions and those specified by the even numbers relate to 90 degree rotations. 5 and 10 centimeter drifts in robot pose estimation along x and y directions, respectively, and a 3 degree drift in its attitude, are observed at the end of the motion, using ICP method. Upper subplot in figure 7 shows the variances of one of the elements of robot's state (position along y axis) in this algorithm, which grows unbounded during robot's motion.

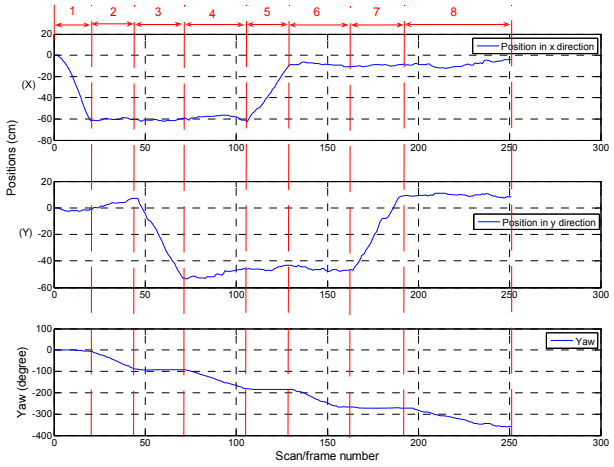

Fig. 4. Estimated location of the robot along x and y axes and its rotation about z axis of a laser-fixed coordinate frame, produced using ICP method.

Furthermore, we have implemented the MonoSLAM algorithm, based on the information of a monocular camera as the only data source, for localizing the robot using the recorded frames during robot's motion. Figure 5 shows the same elements of robot's state during its motion. Although this method has shown excellent results in localizing a hand-held camera with smooth motion, it is not a good choice for localizing a mobile robot with sudden breaks and with frequent variations in velocity. While the method generates a persistent feature-based map, which is the valuable information for correcting the accumulated drift in confronting with previously seen features, in this experiment it can not utilize this map for correcting drifts; because the method only relies on a single camera and thus constant-velocity model is used for motion prediction, which causes the estimation to deviate from the true path of the robot and thus the observation of previously seen features cannot pass the gating test of data association process. Therefore, the method could not correct the drift, produced during robot's motion, at the end of the motion.

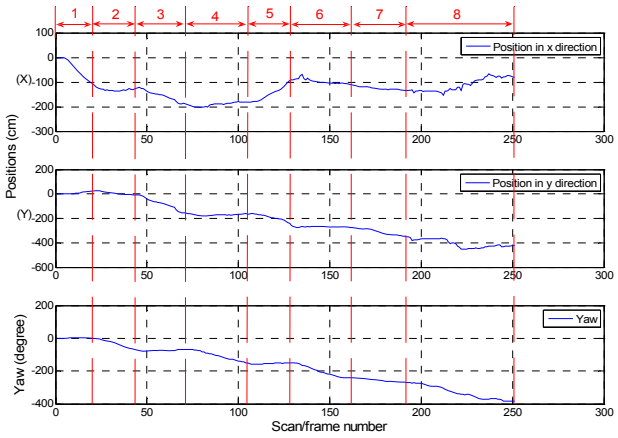

Fig. 5. Estimated location of the robot along x and y axes and its rotation about z axis of a laser-fixed coordinate frame, produced using MonoSLAM method.

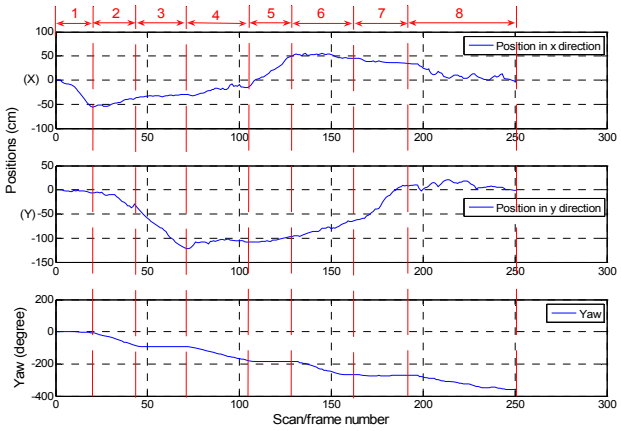

Fig. 6. Estimated location of the robot along x and y axes and its rotation about z axis of a laser-fixed coordinate frame, resulted from the fusion of the information of a monocular camera and an LRF.

Figure 6 shows the same elements of the robot's state, estimated as a result of fusing the information of a monocular camera and an LRF. The first parts of the estimated path have been slightly ruined, compared with the one in the pure ICP-based localization, as a result of the contribution of the features with unknown depths in estimation process. However, the method comes to its own when re-observes existing features in the map. Proposed method is able to recognize the previously seen areas after periods of neglect and correct the accumulated drifts in loop closures, exploiting the map, constructed by the visual features. If robot continues to re-observe other features and travel the path for the second time, depth of most of features will be estimated more accurately and their contribution in the SLAM process leads to a more accurate localization than previous two cases, and also such accurate observations prevent the unbounded growth of the uncertainty. Lower subplot in figures 7 shows the variance of the estimation of one of the elements (position along y axis) of robot's state. As it is seen in this figure loop-closing effect bounds this variance, as well as other elements' variances, in re-observing the existing features in the map, through the established correlations among all elements of the system's state vector in the covariance matrix of system's state. The robot confronts with the existing features in the map at the 218th step, for the first time and it returns to its initial location (origin of the global-fixed coordinate frame) at the 251st step.

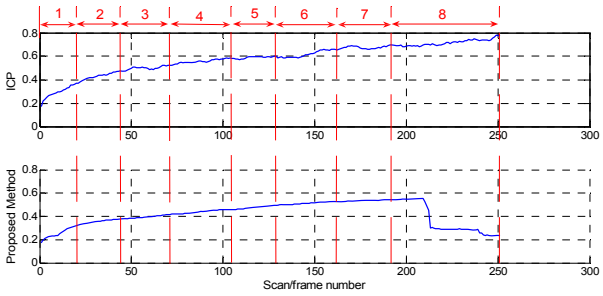

Fig. 7. Variance of the estimation of robot's position along y axis. Upper figure, has been resulted from the ICP method and lower one has been resulted from the proposed method.

Table 1 demonstrates a comparison of the drifts, resulted from these methods, at the end of robot's motion, when robot returns and stops at the origin of global coordinate.

VII. Conclusion

In this paper, a method has been proposed to augment the information of a monocular camera and a range finder. Point-wise scan matching methods are adopted as the best choice, in terms of accuracy, for constructing a motion model to localize a mobile robot in an unstructured environment, without using the encoder data. Augmenting the visual information, obtained from a monocular camera, with this motion model allows the method to construct a feature-based representation of the environment, which makes it possible to exploit the great benefits of the EKF, in terms of highly efficient representation of correlated uncertainty. Therefore, not only this method causes the robot to benefit from an accurate motion model but also makes it capable to recognize previously seen areas after periods of neglect and thus correct the accumulated drift in loop closing motions. This method could be a valuable step towards SLAM in unstructured environments, free from using encoders.

TABLE I
DRIFT COMPARISION AT THE END STEP OF THE MOTION

	ICP	Mono SLAM	Proposed Method
Drift along x axis (cm)	4.08	82.22	0.63
Drift along y axis (cm)	8.57	420.13	0.52
Drift about z axis (degree)	2.86	22.84	1.49

References

[1] M.W.M.G. Dissanayake, P. Newman, S. Clark, H.F. Durrant-Whyte, M. Csorba, A solution to the simultaneous localization and map building (SLAM) problem, IEEE Transactions on Robotics and Automation, 2001.

[2] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem, Proc. AAAI Nat'l Conf. Artificial Intelligence, 2002.

[3] S. Thrun, Y. Liu, D. Koller, A.Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, Simultaneous localization and mapping with sparse extended information filters International Journal of Robotics Research, 2004.

[4] R. Smith, M. Self, and P. Cheeseman, Estimating uncertain spatial relationships in robotics, In I.J. Cox and G.T. Wilfon, editors, Autonomous Robot Vehicles, pages 167-193. Springer-Verlag, 1990.

[5] A. A. Aghamohammadi, H. D. Taghirad, A. H. Tamjidi, and E. Mihankhah, Feature-Based Range Scan Matching For Accurate and High Speed Mobile Robot Localization, ECMR, 2007.

[6] F. Lu, E. Milios, Robot pose estimation in unknown environments by matching 2D range scans, Journal of Intelligent and Robotic Systems, 1997.

[7] S. Se, D. Lowe, J. Little, Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks, Int. J. Robot, 2002.

[8] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera, Proc. ICCV, 2003.

[9] J.M.M. Montiel, J. Civera, and A.J. Davison, "Unified Inverse Depth Parametrization for Monocular SLAM," Proc. RSS, 2006.

[10] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. Autonomous Robots, 1997.

[11] Q. Zhang, R. Pless, Extrinsic calibration of a camera and laser range finder (improves camera calibration), IROS, 2004.

[12] Gelb, A. 1984. Applied Optimal Estimation. M.I.T. Press