



Journal of Artificial Intelligence

ISSN 1994-5450

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Novel RBPF for Mobile Robot SLAM Using Stereo Vision*

^{1,2}Zesu Cai, ²Jian Cao, ²Lining Sun and ¹Maohai Li

¹School of Computer Sciences and Technology,

²School of Mechatronics Engineering, Harbin Institute of Technology,
Harbin 150001, China

Abstract: This study presents the novel RBPF for mobile robot SLAM using stereovision to extract landmark information. The particle filter is combined with Gaussian Mixture Unscented Particle Filters (GMUPF) to extending the path posterior by sampling new poses that integrate the current observation that drastically reduces the uncertainty about the robot pose. The landmark position estimation and update is also implemented through GMUPF which a single update step from moving and sensing can be done and the change to the map certainty can be done in constant time. Furthermore, the number of resampling steps is determined adaptively, which seriously reduces the particle depletion problem. The 3D natural point landmarks are structured with matching Scale Invariant Feature Transform (SIFT) feature pairs. The matching for multi-dimension SIFT features is implemented with a KD-tree which introduce the Mahalanobis distance instead of the Euclidean distance for matching features in the time cost of $O(\log 2^N)$. Experiment results on real robot in our indoor environment show the advantages of our methods over previous approaches.

Key words: Mobile robot, SLAM, RBPF, SIFT, stereovision

INTRODUCTION

A key prerequisite for a truly autonomous robot is that it can simultaneously localize itself and accurately map its surroundings. The problem of achieving this is one of the most active areas in mobile robotics research, which is known as Simultaneous Localization and Mapping (SLAM). Most of the present research on SLAM is based on Extended Kalman filter (EKF) (Leonard *et al.*, 2002), whose primary advantage is that it estimates the fully posterior over landmark maps and robot poses in an online fashion. However, the use of the EKF in practice has two well-known drawbacks. The first one is that linearization can produce highly unstable filters if the assumptions of local linearity is violated. The second one is that the derivation of the Jacobian matrices is nontrivial in most applications and often leads to significant implementation difficulties.

Recently, particle filter have been at the core of solutions to higher dimensional robot problems such as SLAM, which, when phrased as a state estimation problem. RBPF (Murphy and Russell, 2001) is proposed as an effective way of representing alternative hypotheses on robot paths and associated maps. More recently, RBPF is used widely to building map (Stachniss *et al.*, 2005). Daily and Parnichkun (2005) describe the application of Fast SLAM using a trinocular stereo camera. Se *et al.* (2002) demonstrate the use of SIFT point features as landmarks for the SLAM problem using a trinocular Stereo vision. Gil *et al.* (2006) describes an approach that builds three dimensional maps using visual landmarks extracted from stereo images of an unmodified environment and solves the SLAM problem.

Corresponding Author: Zesu Cai, School of Computer Sciences and Technology, Harbin Institute of Technology, Harbin 150001, China

*Originally Published in *Journal of Artificial Intelligence*, 2008

In this study, we present an investigation into the use of stereo vision (Elinas *et al.*, 2006) for SLAM in indoor environment with 3D feature landmarks, which are structured from the SIFT feature matching pairs. These 2D SIFT features are used to structure 3D landmarks because they are invariant to image scale, rotation and translation as well as partially invariant to illumination changes and affine or 3D projection and their description is implemented with multi-dimensional vector (Lowe, 2004). This combination can result in many highly distinctive landmarks from environment, which simplifies the data association problem to only distinguishing unique landmarks. We present a fast and efficient algorithm which introduce the Mahalanobis distance instead of the Euclidean distance for matching features in a KD-Tree in the time cost of $O(\log 2^N)$. Following, our approach applies RBPF to estimate a posterior of the path of the robot, where each particle has associated with an entire map, in which each landmark is estimated and updated by a novel algorithm to recursively update the posterior density of SLAM called the Gaussian Mixture Unscented Particle Filters (GMUPF) (Wendel *et al.*, 2005). The filter has equal or better estimation performance when compared to standard particle filters and the Unscented Particle Filter (UPF) (Julier *et al.*, 2000), at a largely reduced computational cost. The GMUPF combines an Importance Sampling (IS) based measurement update step with an Unscented Kalman Filter (UKF) based Gaussian sum filter for time update and proposal density generation.

NOVEL RBPF FOR SLAM

Consider the case of a mobile robot moving through an unknown environment consisted of a set of landmarks. The landmark n is denoted θ_n . The robot moves according to a known probabilistic motion model $p(s_t | u_t, s_{t-1})$, where, s_t denotes the robot state at time t and the control input u_t carried out in the time interval $[t-1, t]$. As the robot moves around, it takes measurements z_t of its environment through observation model $p(z_t | s_t, \theta, n_t)$, where, θ is the set of all landmarks and n_t is the index of the particular landmark observed at time t .

The SLAM problem is to recover the posterior distribution $p(s^t, \theta_1, \dots, \theta_M | z^t, u^t, n^t)$, where, M is the number of landmarks observed so far and the notation s^t denotes s_1, \dots, s_t (and similarly for other variables). Murphy and Russel (2001) provide an implementation of RBPF for SLAM:

$$p(s^t, \theta_1, \dots, \theta_M | z^t, u^t, n^t) = p(s^t | z^t, u^t, n^t) \prod_{n=1}^M p(\theta_n | s^t, z^t, n^t) \quad (1)$$

This can be done efficiently, since the factorization decouples the SLAM problem into a path estimation problem and individual conditional landmark location problems and the $p(\theta_n | s^t, z^t, n^t)$ can be computed analytically once s^t and z^t are known and the amount of computation needed for each incremental update stays constant, regardless of the path length. Each map is constructed given z^t and the trajectory s^t represented by the corresponding particle. Each particle is of the form $s_i^{(i)} = \{s^{t(i)}, \mu_{1,t}^{(i)}, \Sigma_{1,t}^{(i)}, \dots, \mu_{M,t}^{(i)}, \Sigma_{M,t}^{(i)}\}$, Where (i) indicates the index of the particle; $s^{t(i)}$ is its path estimate, $\mu_{m,t}^{(i)}$ and $\Sigma_{m,t}^{(i)}$ are the mean and variance of the Gaussian representing the m -th landmark location.

In our methods, the i -th new pose $s_i^{(i)}$ is drawn from the posterior $p(s_i | s^{t-1(i)}, u^t, z^t, n^t)$, which takes the measurement z_t into consideration, along with the landmark n_t and $s^{t-1(i)}$ is the path up to time $t-1$ of the i -th particle. It can be shown that $p(s_i | s^{t-1(i)}, u^t, z^t, n^t)$ can be approximated as closely as desired by a Gaussian Mixture Model (GMM) of the following form:

$$p(s_t | s^{t-1, (i)}, u^t, z^t, n^t) \sim \sum_{g=1}^G \alpha^{(g)} N(s_t; \tilde{s}_t^{(g, (i))}, P_t^{(g, (i))}), \quad (2)$$

$$i = 1, 2, \dots, N$$

Where:

G = The number of mixing components

$\alpha^{(g)}$ = The mixing weights

$N(s_t; \tilde{s}_t^{(g, (i))}, P_t^{(g, (i))})$ is a normal distribution with mean vector $\tilde{s}_t^{(g, (i))}$ and positive definite covariance matrix $P_t^{(g, (i))}$.

GMUPF Algorithm

By using the UT transformation (Miaohai *et al.*, 2006), we follow GMUPF algorithm to extend the path $s^{(i)}$ by sampling the new poses $s^{(i)}$ from the posterior $p(s_t | s^{t-1, (i)}, u^t, z^t, n^t)$.

Calculate the Sigma Points

$$\chi_{t-1}^{(i)} = \{\tilde{s}_{t-1}^{(i)}; \tilde{s}_{t-1}^{(i)} \pm \sqrt{(L + \lambda)P_{t-1}^{(i)}}\} \quad (3)$$

Using Motion Model to Predict

At time $t-1$, assume the posterior state density is approximated by the G -component GMM:

$$P_g(s_{t-1}^{(i)} | Z_{t-1}^{(i)}) = \sum_{g=1}^G \alpha_{t-1}^{(g)} N(s_{t-1}^{(i)}; \tilde{s}_{t-1}^{(g, (i))}, P_{t-1}^{(g, (i))}) \quad (4)$$

and the process and observation noise densities are approximated by the following I and J component GMMs, respectively

$$p_g(v_{t-1}) = \sum_{i=1}^I \beta_{t-1}^{(i)} N(v_{t-1}; \mu_{v_{t-1}}^{(i)}, Q_{t-1}^{(i)}) \quad (5)$$

$$p_g(o_t) = \sum_{i=1}^I \gamma_t^{(i)} N(o_t; \mu_{o_t}^{(i)}, R_t^{(i)}) \quad (6)$$

Where:

V_{t-1} = The process noise

O_{t-1} = The observation or measurement noise

For $j = 1, \dots, J$, set $\tilde{p}_j(o_t) = N(o_t; \mu_{o_t}^{(j)}, R_t^{(j)})$. For $j = 1, \dots, J$, set $\tilde{p}_j(v_{t-1}) = N(v_{t-1}; \mu_{v_{t-1}}^{(j)}, Q_{t-1}^{(j)})$ and for $g = 1, \dots, G$, set $\tilde{p}_g(s_{t-1}^{(i)} | Z_{t-1}^{(i)}) = N(s_{t-1}^{(i)}; \tilde{s}_{t-1}^{(g, (i))}, P_{t-1}^{(g, (i))})$.

For $\acute{g} = 1, \dots, \acute{G}$, use the motion model to calculate a $g'' = 1, \dots, G''$ Gaussian approximate and update the mixing weights. For $g'' = 1, \dots, G''$, complete the measurement update step to calculate a Gaussian approximate $\tilde{p}_{g''}(s_t^{(i)} | Z_t^{(i)}) = N(s_t^{(i)}; \tilde{s}_t^{(i)(g'')}, P_t^{(i)(g'')})$, update the mixing weights.

Generate the Proposal Distribution

The predictive state density is now approximated:

$$P_g(s_t^{(i)} | Z_{t-1}^{(i)}) = \sum_{g'=1}^{G'} \alpha_{t-1}^{(g')} N(s_t^{(i)}; \tilde{s}_t^{(i)(g')}, P_t^{(i)(g')}) \quad (7)$$

The proposal distribution is approximated by the GMM:

$$p_g(s_t^{(i)} | Z_t^{(i)}) = \sum_{g^*} \alpha_t^{(g^*)} N(s_t^{(i)}; \tilde{s}_t^{(i)(g^*)}, P_t^{(i)(g^*)}) \quad (8)$$

Measurement Update

Draw N samples $\{\chi_t^{(i)}; i=1, \dots, N\}$ from the proposal distribution $p_g(s_t^{(i)} | Z_t^{(i)})$ and calculate their corresponding importance weights:

$$\tilde{w}_t^{(i)} = \frac{p(Z_t | \chi_t^{(i)}) p_g(\chi_t^{(i)} | Z_{t-1})}{p_g(\chi_t^{(i)} | Z_t)} \quad (9)$$

Normalize the weights: $w_t^{(i)} = \tilde{w}_t^{(i)} / \sum_{i=1}^N \tilde{w}_t^{(i)}$

Use a weighted EM (WEM) algorithm to fit a G -component GMM to the set of weighted particles $\{w_t^{(i)}, \chi_t^{(i)}; i=1, \dots, N\}$:

$$p_g(s_t^{(i)} | Z_t^{(i)}) = \sum_{g^*} \alpha_t^{(g^*)} N(s_t^{(i)}; \tilde{s}_t^{(i)(g^*)}, P_t^{(i)(g^*)}) \quad (10)$$

Inference

The conditional mean state estimate $\hat{s}_t^{(i)} = E[s_t^{(i)} | Z_t^{(i)}]$ and the corresponding error covariance $\hat{P}_t^{(i)} = E[(s_t^{(i)} - \hat{s}_t^{(i)})(s_t^{(i)} - \hat{s}_t^{(i)})^T]$ can be calculated in this ways:

$$\begin{aligned} \hat{s}_t^{(i)} &= \sum_{i=1}^N w_t^{(i)(0)} \chi_t^{(i)(0)} \\ \hat{P}_t^{(i)} &= \sum_{i=1}^N w_t^{(i)(0)} (\chi_t^{(i)(0)} - \hat{s}_t^{(i)})(\chi_t^{(i)(0)} - \hat{s}_t^{(i)})^T \end{aligned} \quad (11)$$

IMPLEMENTATION DETAILS OF USING STEREO VISION

SIFT Feature Extraction Based Stereo Vision

Let I_t^L and I_t^R denote the right and left gray scale images captured with a stereo camera at a time t . we compute image points of interest from both images by selecting maximal points in the scale space pyramid of a Difference of Gaussians. We extract natural landmarks that correspond to points in the 3-dimensional space. Each landmark is a vector $l = \{P^G, C^G, \alpha, s, f\}$, such that $P^G = \{X^G, Y^G, Z^G\}$ is a 3-dimensional position vector in the map's global coordinate frame, C^G is the 3×3 covariance matrix for P^G and α, s, f describe an invariant feature based on the Scale Invariant Feature Transform (SIFT). Parameter α is the orientation of the feature, s is its scale and f is the 128-dimensional key vector which represents the histogram of local edge orientations. Each point is accompanied by its SIFT descriptor and then matched across images. The matching procedure is constrained by the well-known epipolar geometry of the stereo rig and the Euclidean distance of their SIFT keys. In addition, a comparison between SIFT descriptors is used to avoid false correspondence. As a result, at a time t we obtain a set of B observations denoted: $O_t = \{o_1, \dots, o_n\}$.

$$\mathbf{o}_j = \{\mathbf{P}_{\mathbf{o}_j}^L = \{\mathbf{X}_{\mathbf{o}_j}^L, \mathbf{Y}_{\mathbf{o}_j}^L, \mathbf{Z}_{\mathbf{o}_j}^L\}, \mathbf{p}_{\mathbf{o}_j}^L = \{r_{\mathbf{o}_j}, c_{\mathbf{o}_j}, l\}, \mathbf{C}^L, \alpha, s, \mathbf{f}\} \quad (12)$$

Where, $\mathbf{p}_{\mathbf{o}_j}^L = \{r_{\mathbf{o}_j}, c_{\mathbf{o}_j}, l\}$ is the image coordinates of the point and $j \in [1, 2, \dots, n]$.

KD-Tree Based Feature Matching

Here we describes KD-tree algorithm for determining the matched SIFT feature pairs of left and right images captured along the robot's path by a stereo vision system. We use each associated SIFT descriptor to improve the data associated process. Each SIFT descriptor is a 128-long vector computed from the image gradient at a local neighbourhood of the interest point. In the approaches (Miaohai *et al.*, 2006), data association is based on the squared Euclidean distance between descriptors. However, when the same point is viewed from significantly different viewpoints, the difference in the descriptor is remarkable. In the presence of similar looking landmarks, this fact causes a significant number of false correspondences. In order to reduce the false correspondence problem, we propose a different method to present the distance function. We consider different views of the same visual landmark as different patterns belonging to C_i . Whenever a landmark is found, it is tracked along p frames and its descriptors d_1, d_2, \dots, d_p are stored. Then, for each landmark C_i , we compute a mean value \tilde{d}_i and estimate a covariance matrix P_i , assuming the elements in the SIFT descriptor independent. Based on this data we computer the Mahalanobis distance:

$$\begin{aligned} \forall \tilde{d}_k \in E, |\tilde{d}_i \leftrightarrow \tilde{d}_j| &\leq |\tilde{d}_i \leftrightarrow \tilde{d}_k|, \\ |\tilde{d}_i \leftrightarrow \tilde{d}_j| &= \sqrt{(\tilde{d}_i - \tilde{d}_j) P_i (\tilde{d}_i - \tilde{d}_j)^T} \end{aligned} \quad (13)$$

Where, p_i is a diagonal covariance matrix associated with the class C_i with mean value \tilde{d}_i . We computer the Mahalanobis distance for all the landmarks in the map of each particle and assign the correspondence to the landmark that minimizes Mahalanobis distance. Then we can use the following equation to judge the matching for two key-points:

$$|kp_1 \leftrightarrow kp_2| / |kp_2 \leftrightarrow kp_2| < \lambda \quad (14)$$

Where, λ is constant and $0 < \lambda < 1$, if this inequality is satisfied, then the matching is successful and simultaneously eliminates the false matching. Figure 1 shows an example of SIFT feature matching for a pair image from the labor with stereo vision and we obtain 76 matched pairs which the matching accurate rate is higher than 90%.

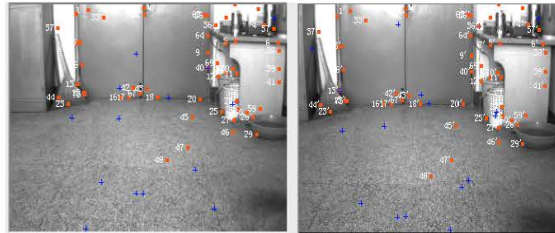


Fig. 1: The SIFT feature matches based on KD-tree and the matching pairs are represented by red rectangle. The mismatching pairs are represented by blue +. The matched key-points are 76. The corrected are 68

Observation Model Based Stereo Vision

In the stereovision case, we perform a linear search of the keys in the left image for the best match to each key in the right, subject to epipolar constraints and determine its 3D position and covariance according to the well-known stereo equations:

$$Z = fB/d, X = uZ/f, Y = vZ/f \tag{15}$$

Where, f is the focal length of the camera, B is the based-line of the stereo vision, d is the disparity between SIFT keys in the left and right images and $[u,v]$ is the pixel position of the key in the right camera.

An observation is defined as a set of k correspondences between landmarks in the map and the current view. Let $F_t = \{f_1, \dots, f_k\}$ be the k SIFT feature key-points observed at time t , in which there are n key-points matching with the 3D landmarks in the map database: $n_t^l = \cup_{i=1,2,\dots,n} (f_i \leftrightarrow L_{f_i})$ and there are m key-points matching the 2D SIFT feature key-points which observed at time $t-1$ and are not reconstructed and added to the map database: $n_t^v = \cup_{i=1,2,\dots,m} (f_{n+i} \leftrightarrow V_{f_{n+i}})$. Then the likelihood of the observation Z_t being obtained as:

$$p(z_t | s_t^{(i)}, \theta, n_t) = p(z_t^l | s_t^{(i)}, \theta, n_t^l) p(z_t^v | s_t^{(i)}, \theta, n_t^v)$$

Where, z_t^l represented the observation $F_t = \{f_1, \dots, f_n\}$ and z_t^v represents the observation $F_v = \{f_{n+1}, \dots, f_{n+m}\}$, $p(z_t^l | s_t^{(i)}, \theta, n_t^l)$ represents the likelihood of the observation z_t^l given the matching relation n_t^l and $p(z_t^v | s_t^{(i)}, \theta, n_t^v)$ represents the likelihood of the observation z_t^v given the matching relation n_t^v , these two likelihood can be calculated separately as follows:

$$\ln p(z_t^l | s_t^{(i)}, \theta) = \sum_{j=1}^n \ln p(f_j | s_t^{(i)}, L_{f_j})$$

$$\ln p(z_t^v | s_t^{(i)}, \theta) = \sum_{j=n+1}^{n+m} \ln p(f_j | s_t^{(i)}, V_{f_j})$$

Where, $p(f_j | s_t^{(i)}, L_{f_j})$ represents the likelihood of the observation being f_j when robot at pose $s_t^{(i)}$ observing the landmark L_{f_j} , $p(f_j | s_t^{(i)}, V_{f_j})$ represents the likelihood of the observation being f_j when robot at pose $s_t^{(i)}$ observing the landmark V_{f_j} . A pose of the camera, s_t , defines a transformation $[R, T]_s$ from the camera to the global coordinate frame. Let the 3D coordinates of the landmark L_{f_j} be $(X^{(i)}, Y^{(i)}, Z^{(i)})$, using the Mahalanobis distance metric, we can obtain $\ln p(f_j | s_t^{(i)}, L_{f_j})$ as follows:

$$\ln p(f_j | s_t^{(i)}, L_{f_j}) = -0.5 \min (T_1, (\hat{I}_j - I_j)^T S^{-1} (\hat{I}_j - I_j))$$

$$S = J(R_t G_{f_j} R_t^T) J^T \tag{16}$$

Where:

- J = The Jacobian matrix of the observation equation
- G_{f_j} = The covariance of L_{f_j}

The maximum observation innovation T_1 is constant (in our case, 3.0), which is selected so as to prevent outlier observations from significantly affecting the observation likelihood.

While the feature $V_{\tilde{f}_i}$ has no 3D spatial information, $\ln p(f_j | s_t^{(i)}, V_{\tilde{f}_i})$ is only calculated according to epipolar constrain:

$$\ln p(f_j | s_t^{(i)}, V_{\tilde{f}_i}) = -0.5(\text{dist}(I_j, H_{\tilde{f}_i}) + \text{dist}(I_{\tilde{f}_i}, H_j))$$

Where, $I_{\tilde{f}_i}$ is the image coordinate of the feature $V_{\tilde{f}_i}$, $H_{\tilde{f}_i}$ is the epipolar line on the image plane corresponding to $V_{\tilde{f}_i}$ at time t and H_j is the epipolar line on the image plane corresponding to the feature f_j at time $t-1$, $\text{dist}(\bullet)$ is the function of the distance between point and line.

After calculating the observation model $p(z_t | s_t, \theta, n_t)$, which can be used to evaluate the i -th particle weight $w_t^{(i)}$ and $w_t^{(i)}$ is taken as follows:

$$w_t^{(i)} = \frac{p(z_t | s_t^{(i)}, \theta, n_t)}{\sum_{j=1}^N p(z_t | s_t^{(j)}, \theta, n_t)} \quad (17)$$

Motion Model

An essential component to the implementation of RBPF is the specification of the robot's motion model. Let I_t and T_{t-1} represent the pairs of stereo images taken with the robot's camera at two consecutive intervals with the robot moving between the two. For each pair of images we detect points of interest, compute SIFT descriptors for them and perform stereo matching, resulting in 2 sets of landmarks L_{t-1} and L_t . We compute the camera motion using the Levenberg-Marquardt (LM) algorithm minimizing the re-projection error of the 3D coordinates of the landmarks.

Let \tilde{s}_t be the 3-dimensional vector $\tilde{s}_t = \{\text{roll}, T_x, T_y\}$ corresponding to a given $[R, T]$. Our goal is to iteratively compute a correction term χ :

$$\tilde{s}_t^{i+1} = \tilde{s}_t^i - \chi$$

such as to minimize the vector of error measurement ϵ for a known stereo vision calibration matrix K , ϵ is defined as:

$$\epsilon = \begin{bmatrix} \epsilon_0^T \\ \epsilon_1^T \\ \vdots \\ \epsilon_k^T \end{bmatrix} = \begin{bmatrix} p_t^0 - K(Rp_{t-1}^0 + T) \\ p_t^1 - K(Rp_{t-1}^1 + T) \\ \vdots \\ p_t^k - K(Rp_{t-1}^k + T) \end{bmatrix} \quad (18)$$

Given an initial estimate for the parameters, we use LM algorithm to solve for χ that minimizes ϵ in three steps:

- According to the vector of residuals $\epsilon(\chi)$, compute its matrix of derivatives J with respect to the components of χ .

$$J = \begin{bmatrix} \frac{\partial \epsilon_0}{\partial \chi} & \dots & \frac{\partial \epsilon_k}{\partial \chi} \end{bmatrix}^T \quad (19)$$

- Compute the update rule:

$$\begin{aligned}\chi &= -(H + \lambda I)^{-1} \mathbf{g} \\ H &= \mathbf{J}^T \mathbf{J} \\ \mathbf{g} &= \mathbf{J}^T \varepsilon(\chi)\end{aligned}\tag{20}$$

- Set $\tilde{\mathbf{s}}_i^{i+1} = \tilde{\mathbf{s}}_i^i - \chi$

Where, the value of λ controls the distance traveled along the gradient direction and controls the convergence of the solution by switching between pure gradient descent and Newton's method.

RESULTS AND DISCUSSION

The experiments are performed on a Pioneer 3-DX Mobile Robot incorporating an 800 MHz intel Pentium processor as shown in Fig. 2A. Motor control is performed on the on-board computer, while a 2.4GHz PC connected to the robot by a wireless link provides the main processing power for stereo vision processing and the SLAM software. A Point Grey Bumblebee Stereo-vision mounted at the front of the robot is used for detecting the landmarks. The test environment is a robot laboratory with limited space in Fig. 2B.

One of the most difficult issues in a SLAM implementation is the data association problem, where the hypothesis of assigning an observation to an existing map location is tested. In our implementation the KD-tree Based SIFT and Mahalanobis distance solves this problem. For illustrating the advantages of our methods over previous approaches, we implement SLAM with our novel RBPF and previous method. This experiment is described as follows.

Firstly, the robot is set at the distance of 4 m from the lab door and the robot orientation is parallel with the door, at the same time, there are some obstacles in the front of the robot. While the robot is moving ahead, the stereo pair image frames are captured and processed, building the map of the indoor environment.

In Fig. 3, the experiment results. In the map, S represents the start point of robot path and 'E' represents the end point of the robot path, the red points represent the robot path particle, the 2D view of 3D landmarks in the map is represented with blue points. As shown in Fig. 3A, if we increase the number of particles, the performance of conventional RBPF will be improved largely, however, the storage requirement and calculation burden is severely aggravated, owing to each particle associated with a view of the map. Figure 3B shows the built map with the novel RBPF, which adopts 50 particles and 100 particles. The performance of the novel RBPF changes a little with increasing the number of particle, so we can build precise map only with few particles.

Another experiment was carried out in our single lab room, where the compact map is built with our novel RBPF method and 186 pair stereo image frames of size 320×240 are captured. The real

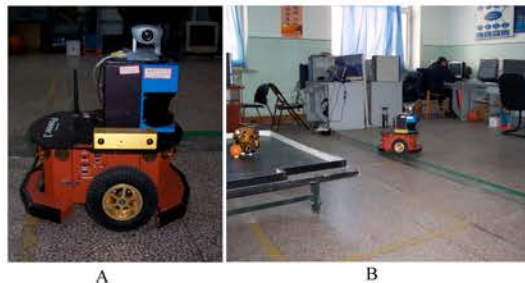


Fig. 2: (A) Pioneer mobile robot and (B) experimental environment

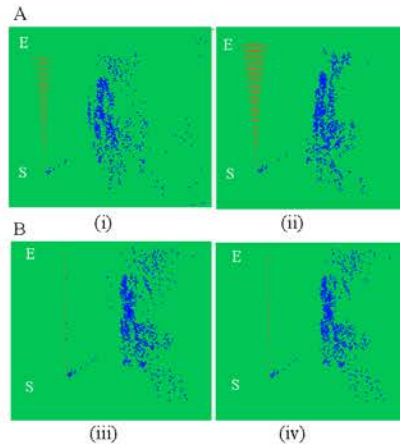


Fig. 3: Experiment results of map building based (A) conventional RBPf: (i) 100 particles, (ii) 500 particles; (B) novel RBPf: (iii) 50 particles, (iv) 100 particles

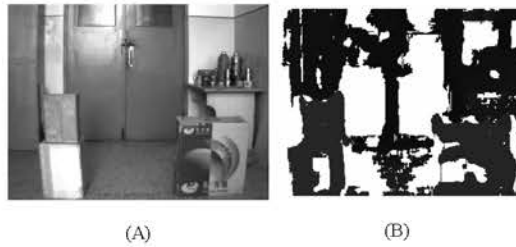


Fig. 4: Part of an image from the stereo camera (A) and the corresponding disparity image (B)

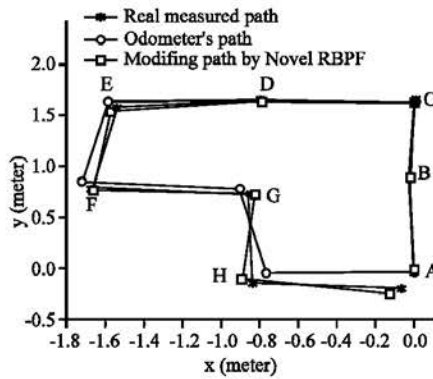


Fig. 5: The Comparison of the three paths, red line present the measured path of robot, green line present the odometer's path, the blue show the modifying path by Novel RBPf

environment of our room is shown in Fig. 4A, this is a typical image of our test environment where the pathways are very narrow with little texture throughout the camera's field of view. Fig. 4B is the disparity image generated by the stereo correspondence algorithm with lighter shades indicating objects closer to camera.

In order to measure the accuracy of our novel RBPF, we have measured ground truth pose data for different locations. The starting location is node A and the last location is node I, the others are labeled nodes B, C, D, E, F, G, H. We have labeled these locations on the learned map shown in the Fig. 5. The comparison of the three paths that comprise odometer path (green dashed line) and modified path by the novel RBPF (blue solid line) and the real measures path (red solid line) showed in the Fig. 5. From the Fig. 5, we can see that the odometer data is less deviation from the real measure data in the short term, however, with the robot moving, the deviation become larger. But for the modified data by novel RBPF, the deviation changes a little with the robot moving.

CONCLUSION

This article described a novel algorithm for SLAM problem using stereo vision. Like many previously published SLAM algorithms, our method calculates posterior probability distributions over 3D SIFT featured maps and robot locations. It does so recursively based on a key property of the SLAM problem: The conditional independence of feature estimates given the robot path. This conditional independence gives rise to a factored representation of the posterior using a combination of particle filters for estimating the robot path and the GMUPF for estimating the map. Furthermore, the number of resampling steps is determined adaptively, which seriously reduces the particle depletion problem. In order to reduce the false correspondence in the data association the Mahalanobis distance is introduced in the KD-tree based feature matching instead of the Euclidean distance. We have shown that we can construct accurate metric maps of 3D point landmarks from dense correlation-based stereo without loop closure. Experiment results on real robot in our indoor environment show that the advantages of our methods over previous approaches.

REFERENCES

- Dailey, M.N. and M. Parnichkun, 2005. Landmark-based simultaneous localization and mapping with stereo vision. In: Proceedings of the Asian Conference on Industrial Automation and Robotics.
- Elinas, P., R. Sim and J.J. Little, 2006. Oslam: Stereo vision slam using the rao-blackwellised particle filter and a novel mixture proposal distribution. Proceedings of the IEEE International Conference on Robotics and Automation, Orlando, FL.
- Gil, A., O. Reinoso, C. Fernandez, M. Vicente, A. Rottmann and O. Martinez-Mozos, 2006. Simultaneous localization and mapping in unmodified environments using stereo vision. In: International Conference on Informatics in Control, Automation and Robotics (ICINCO), pp: 302-309.
- Julier, S., J. Uhlmann and H.F. Durrant-Whyte, 2000. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. On AC*, 45: 477-482.
- Leonard, J., J.D. Tard'os, S. Thrun and H. Choset, 2002. Workshop notes of the icra workshop on concurrent mapping and localization for autonomous mobile robots. In: Proceedings IEEE Int. Conf. Robotics and Autonomous, Washington, DC.
- Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60: 91-110.
- Miaohai, Li, Bingrong, Hong, Zesu, Cai, Ronghua, Luo, 2006. Novel rao-blackwellized particle filter for mobile robot SLAM using monocular vision. *Int. J. Intel. Technol.*, 1: 63-69.
- Murphy, K. and S. Russell, 2001. Rao-Blackwellized Particle Filtering for Dynamic Bayesian Networks, in *Sequential Monte Carlo Methods in Practice*. Springer Verlag.

- Se, S., D. Lowe and J. Little, 2002. Mobile Robot localization and mapping with uncertainty using scale-invariant visual landmarks. *Int. J. Robotics Res.*, 21: 735-758.
- Stachniss, C., G. Grisetti and W. Burgard, 2005. Recovering particle diversity in a rao-blackwellized particle filter for SLAM after actively closing loops. In: *Proceedings IEEE Int. Conf. Robotics and Automation*, Barcelona, Spain, pp: 667-672.
- Wendel, J., J. Metzger, R. Moenikes, A. Maier and G.F. Trommer, 2005. A performance comparison of tightly coupled GPS/INS navigation systems based on extended and sigma point kalman filters. *Proceedings of ION GNSS Long Beach, California*, 13-16: 456-466.