Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 50-1 (2017) 2621–2626

Bayesian Online Learning for Information-based Multi-Agent Exploration with Unknown Radio Signal Distribution

Jinhong Lim * **J.Hyeon Park** * **H.Jin Kim** *

* *Department of Mechanical and Aerospace Engineering, and Institute of Advanced Aerospace Technology Seoul National University, Seoul, Korea (e-mail:{wlsghd90,ka2hyeon,hjinkim}@snu.ac.kr).*

Abstract: Exploring an unknown environment with multiple robots is an enabling technology for many useful applications. This paper investigates decentralized motion planning for multi-agent exploration in a field with unknown received signal strength (RSS) distribution. The environment is modelled with a Gaussian process using Bayesian online learning by sharing the information obtained from the measurement history of each robot. Then we use the mean function of the Gaussian process to infer the multiple RSS source locations. The inferred source locations are modelled as the probability distribution using Gaussian mixture-probability hypothesis density (GM-PHD) filter. This modelling enables nonparametric approximation of mutual information between source locations and future robot positions. We combine the variance function of the Gaussian process and the mutual information to design an informative and noise-robust planning algorithm for multiple robots. The experimental performance is analyzed by comparing with the variance-based planning algorithm.

Keywords: Bayesian nonparametric methods; online Gaussian process; mutual information; GM-PHD filter; RSS-based localization; active sensing; decentralized multi-agent

1. INTRODUCTION

Planning informative actions is an essential problem for active sensing tasks such as autonomous surveillance, mapping and target localization. Especially, engaging multiple robots to cooperatively perform these active sensing tasks has been well studied. Most of the proposed methods are based on the information theory such as entropy and mutual information, because these quantities well represent the informativeness using the measurement history and future poses of sensors.

Many possible approaches are investigated to implement the information-theory-based methods. Among them, some recent studies employ nonparametric approaches to derive entropy (Viseras et al. (2016)) and mutual information (Dames and Kumar (2015)). In Gaussian process-based terrain learning (Viseras et al. (2016)) algorithm, the motion planning of multiple robots is performed using the variance which is derived from online learning of Gaussian process, where the variance is considered as entropy. Besides, in mutual information-based target localization (Dames and Kumar (2015)) algorithm, mutual information is calculated using the detection model of laser scanner and approximating the integral using probability hypothesis density (PHD).

The information-theory-based method is highly applicable to various kinds of sensors. Charrow et al. (2015) proposed the visual sensor-based 3-D mapping using a single quadrotor. Besides, information-based localization using single or multiple robots is performed using downward facing camera (Dames et al. (2015), Kim et al. (2013)), laser scanner (Dames and Kumar (2015)), and many kinds of sensors such as bearing-only or range-only sensors (Charrow et al. (2014), Hoffmann and Tomlin (2010)). Furthermore, recognition of environmental signal distribution based on measurement history has also been performed (Viseras et al. (2016), Bai et al. (2016), Ouyang et al. (2014)).

Among the various kinds of sensors, we consider a received signal strength (RSS) sensor for learning the unknown RSS distribution in this paper. The distribution of RSS field is essential for RSS-based indoor localization (He and Chan (2016), Ferris et al. (2006)) in case of mobile-phone-based commercial advertisement and RSS-based search-and-rescue in GPS-restricted space. For such applications, the accurate RSS distribution map is important to utilize as fingerprint or propagation-model-based likelihood. Thus, we present a decentralized RSS distribution learning framework for a team of autonomous mobile robots.

The proposed algorithm consists of the following three steps: learning the RSS distribution model, modelling the source locations as probability distribution using PHD, and planning the path by computing mutual information and combining with variance. The corresponding details of the three steps are as follows.

* This work was conducted with the support of Defense Acquisition Program Administration(DAPA) and Agency for Defense Development(ADD), and of the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science, ICT and Future Planning (MSIP) (No. 2014034854).

10.1016/j.ifacol.2017.08.137

- First, using the RSS measurement history of all robots, the RSS distribution model of the entire space is obtained, including the unexplored area, by learning the Gaussian process model. The model is described in terms of means and variances over the entire space (Rasmussen (2006)), and lowering the variances is used as one of the objectives considered in this paper.
- Then, we modelled the signal source location as a probability distribution using PHD by predicting the source location from mean function of the Gaussian process model (Ferris et al. (2006)). Among various kinds of implementations of propagating the PHD, which is called PHD filter (Vo et al. (2005)), the Gaussian mixture-probability hypothesis density (GM-PHD) filter is one of the most suitable forms for real-time implementation, because the GM-PHD filter assumes the prior distribution of the target sample to be Gaussian (Vo and Ma (2006), Clark et al. (2006)). Thus, we engage the idea of approximating the prior to Gaussian distribution, and applying measurement-based spatial prior on the birth process to enhance the multi-target localization performance (Houssineau and Laneuville (2010)).
- Finally, the local optimal movements of the robots are derived from the policy using combined two objectives. Using the estimated PHD of source location and the future poses of robots, mutual information between them is calculated in nonparametric way (Charrow et al. (2014), Cover and Thomas (2012)). By maximizing the sum of the two normalized objectives, variance and mutual information, the robots select the most informative actions from randomly sampled action set.

Using single quantity among the two, i.e. entropy and mutual information, has both pros and cons. Entropy, which is equal to the variance function of Gaussian process model, is an useful objective to find the unexplored region, so the variance is reduced the most by measuring data at the region. However, the variance-based policy shows poor performance when the amount of data is not enough for noisy signals. Besides, the importance of the data at each position is not reflected in the robots' motion. On the other hand, mutual information between the RSS source location and future movements of the robots focuses on detecting the sources, where the measurements around the sources are the most important data in regression settings. Also, the mutual information tends to lead mobile robots to move toward the sources with even less amount of data. The mutual information, however, is not desirable if all samples of GM-PHD filter is pruned. The problems of each objective are compensated by the advantages of the other by fusing them.

The main contribution of this paper can be summarized as follows. (1) We fuse variance and mutual information to design noise-robust and efficient policy for multi-agent exploration. Unlike most previous research that use a single quantity among entropy and mutual information, this paper combines them for the settings where Gaussian process is used to model the environmental distribution. (2) Moreover, we engage GM-PHD filter which uses the diffuse spatial prior of the birth process, so the samples of the filter are generated over the entire space. Although Dames and Kumar (2015) also engaged GM-PHD filter

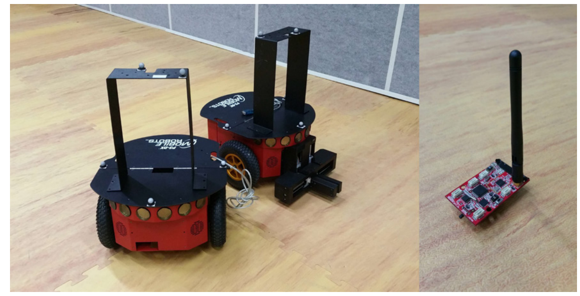

Fig. 1. Experimental components (Pioneer P3-DX (left) and UBee430 mote (right)).

to compute mutual information, the used traditional GM-PHD filter is not suitable when inference of the source locations is uncertain or fluctuates. An actual hardware experiment is performed to find out how the fused two objectives compensate the problems of each other.

The rest of the paper is composed as follows. Section 2 shows the estimation process of the environmental model and source location with details of problem formulation. Then, we describe the decentralized multi-agent control laws in Section 3. Finally, the experimental results are presented in Section 4, which is followed by the conclusions.

2. ESTIMATION OF ENVIRONMENTAL MODEL AND SOURCE LOCATION

In this section, we present the estimation process of the environmental RSS distribution and the source location. Section 2.1 shows the online learning of RSS distribution using Gaussian process, and Section 2.2 presents the RSS source location estimation using GM-PHD filter.

2.1 Gaussian Process Model

Using the data sampled by robots, we build Gaussian process model (Rasmussen (2006)) that represents the environmental distribution. In our case, let $\mathbf{x}_r^k \in X^k$ be the 2-D position vector of the r-th robot at iteration k, $\mathbf{p}_r^k \in P^k$ be the packet which contains RSS data of m access points in the order of their ID received by r-th robot at iteration k, and $y_r^k \in \mathbf{y^k}$ be the total sum of RSS data measured by r-th robot at iteration k, where $r = 1, \cdots, R$. Then, by continuously sharing and saving data, the training dataset at iteration k can be made as $D = \{(X^1, \mathbf{y}^1), (X^2, \mathbf{y}^2), \cdots, (X^k, \mathbf{y}^k)\}$, which contains entire RSS data history measured by all robots.

From the dataset, Gaussian process is used to obtain the predictive distribution over the arbitrary positions $\mathbf{x}_{i*} \in X_*$, where $i = 1, \cdots, n$. We set the arbitrary positions to be the fixed entire workspace of robots, where the number of points is n, so that the source location can be estimated by maximum likelihood principle. Although the large size of the arbitrary position set causes more computational cost on calculating kernel matrix, the matrix can be computed in advance because the matrix is fixed for the whole procedure. For the matrix, Gaussian kernel function is used, function that is defined as (1), with Kronecker's delta $\delta_{\mathbf{xx'}} = 1$ $(if\ \mathbf{x} = \mathbf{x'})$.

$$k(\mathbf{x}, \mathbf{x'}) = \sigma_f^2 \exp\left(-\frac{|\mathbf{x} - \mathbf{x'}|^2}{2l^2}\right) + \sigma_n^2 \delta_{\mathbf{xx'}}. \tag{1}$$

Assuming zero mean and variance of σ_n^2 for noise of target function, let the target function be $s = f(\mathbf{x}) + \epsilon$, and a prediction $s_* = f(\mathbf{x}_*)$. Then the joint distribution is,

$$\begin{bmatrix} s \\ s_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(\bar{X},\bar{X}) & K(\bar{X},X_*) \\ K(X_*,\bar{X}) & K(X_*,X_*) \end{bmatrix}\right), \tag{2}$$

where $\bar{X} = \left[X^1, X^2, \cdots, X^k\right]$, $\left[K(X,X')\right]_{uv} = k(x_u, x_v^{'})$ and $x_u \in X$, $x_v \in X^{'}$, i.e. $K(\bar{X},\bar{X})$ is the $kR \times kR$ covariance matrix computed for all pairs between training positions, $K(\bar{X},X_*)$ is the $kR \times n$ covariance matrix computed for all pairs of training positions and arbitrary positions, and $K(X_*,X_*)$ is the $n \times n$ covariance matrix computed for all pairs between arbitrary positions. The predictive distribution of each access point is then given as (3) and (4) with $m \times kR$ matrix $\bar{P} = [P^1, P^2, \cdots, P^k]$ which contains all shared RSS packet history.

$$\mu_* = K(\bar{X},X_*)^T K(\bar{X},\bar{X})^{-1} \bar{P}^T \tag{3}$$

$$\Sigma_* = K(X_*,X_*) - K(\bar{X},X_*)^T K(\bar{X},\bar{X})^{-1} K(\bar{X},X_*) \tag{4}$$

Let $\theta = [\sigma_f^2, l^2, \sigma_n^2]^T$ be the hyperparameter, which is the only tunable set in the Gaussian process. The optimal hyperparameter can be obtained by maximizing the log-marginal likelihood,

$$\log p(\mathbf{y}|X,\theta) = \\ -\frac{1}{2}\mathbf{y}^T K(\bar{X},\bar{X})^{-1}\mathbf{y} - \frac{1}{2}\log K(\bar{X},\bar{X}) - \frac{kR}{2}\log 2\pi \, . \tag{5}$$

The log-marginal likelihood can be maximized using gradient method, and by optimizing the hyperparameter, the best model that fits observed data the most is obtained as (3). The mean function (3) is also used as the likelihood distribution of the RSS source location which is discussed in Section 2.2. The variance function (4) is used as a part of the control law of robots which is discussed in Section 3.2.

2.2 GM-PHD Filter with Observation-based Spatial Prior

In this section, we model the RSS source location as the probability distribution using GM-PHD filter. The process of GM-PHD filter consists of prediction and measurement update like other typical filters such as Kalman filter or particle filter. At the measurement update step, we obtained measurements from the mean RSS distribution function (3) of Gaussian process model. Let $\mathbf{z}_j^k \in Z^k$ be the inferred location of j-th RSS source at iteration k. Since the model of RSS versus distance can be described as the inverse proportion curve in log-scale, the location of sources can be inferred by finding the location that has the maximum predicted mean RSS among n arbitrary positions.

$$\mathbf{z}_j^k = \arg\max_{\mathbf{x}_*} \left[\mu_*(\mathbf{x}_*)\right]_j \quad for\ \mathbf{x}_* \in X_* \tag{6}$$

At the prediction step, the traditional GM-PHD filter propagates several predicted intensities, which are called birth, spawn, and survive samples respectively. Let $\hat{\mathbf{x}}_j^k \in \hat{X}^k$ be the estimated position vector of j-th source at iteration k, and the system update and measurement update functions be,

$$g(\hat{\mathbf{x}}_j^k) = N\left(\hat{\mathbf{x}}_j^k; G\hat{\mathbf{x}}_j^{k-1}, Q^k\right) \tag{7}$$

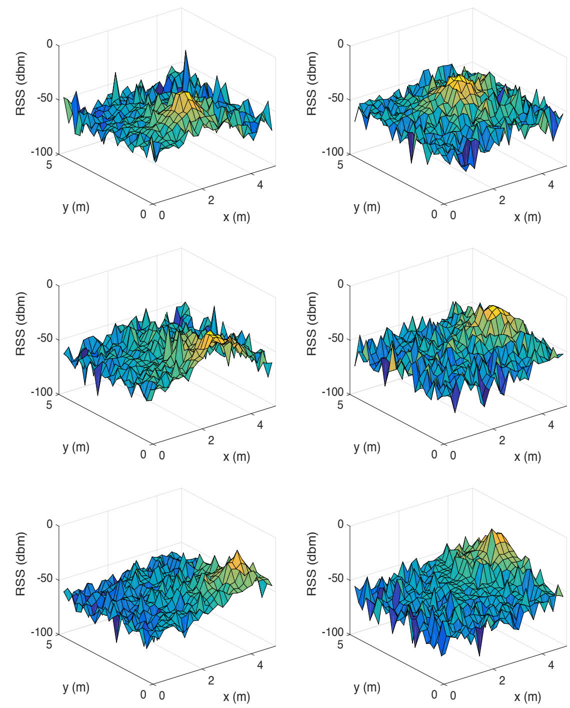

Fig. 2. Ground truth RSS distributions of six RSS nodes.

$$h(\mathbf{z}_j^k) = N\left(\mathbf{z}_j^k; H\hat{\mathbf{x}}_j^k, R^k\right), \tag{8}$$

where G and H are 2×2 identity transition and observation matrices respectively, and Q^k and R^k are the 2×2 covariance matrices. Then the intensities of each sample are computed as,

$$v_{\gamma,k|k-1} = \sum_{(j)} w_\gamma^{(j)} N(m_\gamma^{(j)}, \mathbf{P}_\gamma) \tag{9}$$

$$v_{\beta,k|k-1} = \sum_{(j)}\sum_{(l)} \{w_{k-1}^{(j)} w_\beta^{(j)} N(m_{\beta,k|k-1}^{(j)(l)}, \mathbf{P}_\beta)\} \tag{10}$$

$$v_{s,k|k-1} = p_s \sum_{(j)} w_{k-1}^{(j)} N(m_{s,k|k-1}^{(j)}, \mathbf{P}_s), \tag{11}$$

where γ, β, s stand for birth, spawn, survive samples respectively, $m_{s,k|k-1}^{(j)} = G\hat{\mathbf{x}}_{(j)}^{k-1}$, $m_{\beta,k|k-1}^{(j)}$ is the position vector obtained by adding Gaussian random noise to $m_{s,k|k-1}^{(j)}$, $\mathbf{P}$ is the covariance matrix of the distribution of each sample, m_γ is the matrix which contains position vectors extracted from the preset spatial prior, and w is the weight of corresponding samples. Then, the final estimated target posterior intensity is updated as,

$$v_{k|k-1} = v_{\gamma,k|k-1} + v_{\beta,k|k-1} + v_{s,k|k-1} \tag{12}$$

$$v_k = (1 - p_{D,k}) v_{k|k-1} + \sum v_{D,k}, \tag{13}$$

where $v_{D,k}$ is the term reflecting that the measurements are correctly detected. Detailed explanation on propagation of PHD can be found in Section III-B of (Vo and Ma (2006)) including mixture and pruning steps. The traditional method that presets spatial prior on birth samples is not suitable for our case, because the locations of the RSS sources are unknown in advance, and the estimates of the locations are highly perturbing. Thus, we apply observation-based spatial prior to the birth process. Then the intensities (12) and (13) are changed to,

$$v_{k|k-1} = v_{\beta,k|k-1} + v_{s,k|k-1} \tag{14}$$

$$v_k = (1 - p_{D,k})v_{k|k-1} + \sum v_{D,k} + \sum_{(j)} w_\gamma^{(j)} N(\mathbf{z}_{(j)}^k, \mathbf{P}_\gamma), \tag{15}$$

by introducing a normalizing factor w_0. Details about usage of the normalizing factor can be found in Section 4 of (Houssineau and Laneuville (2010)). Using the posterior intensities (15), which are composed of weights and mean positions of the sources, estimated locations can be obtained. Also, the final estimates are used for nonparametric approximation of mutual information, which is discussed in Section 3.1.

3. DECENTRALIZED MULTI-AGENT CONTROL

3.1 Nonparametric Computation of Mutual Information

One of the objectives considered for cooperative multi-agent exploration in this paper is mutual information between estimated RSS source locations and future measurement of robots. The mutual information is defined as,

$$\begin{aligned} I\left[\hat{\mathbf{X}}^k;\mathbf{B}\right] &= \int\int p(\hat{X}^k, B)\log\frac{p(\hat{X}^k, B)}{p(\hat{X}^k)p(B)}\delta\hat{X}^k\delta B \\ &= H\left[\mathbf{B}\right] - H\left[\mathbf{B}|\hat{\mathbf{X}}^k\right], \end{aligned} \tag{16}$$

where, $H[\mathbf{B}]$ is the entropy, $H[\mathbf{B}|\hat{\mathbf{X}}^k]$ is the conditional entropy of the future measurements, and $b \in B$ is the binary event,

$$b = \begin{cases} 0 & (Z^k = \phi) \\ 1 & (else) \end{cases}, \tag{17}$$

where the binary event denotes whether each robot detects at least one RSS source or not. As shown in (17), $b = 0$ denotes that a robot does not detect any of the RSS source, and $b = 1$ represents the other case. By introducing such event, the future positions of each robot can be modelled as random samples with probability distribution, so the mutual information becomes the measure of the difference between two probability distributions, future robot positions and estimated RSS source locations (15). In this paper, we set probability of detection to be,

$$p_d(\hat{\mathbf{x}}^k) = \min\left(1, \frac{r_t + 1}{r_{rs} + 1}\right) \times \mathbf{1}\left(r_{rs} < r_{\max}\right), \tag{18}$$

where r_{rs} is the distance between a robot and an estimated RSS source location, r_t is the preset threshold range, and r_{max} is the preset maximum range. Such setting has physical meaning for RSS propagation model, since the RSS tends to steeply increase near the source. Furthermore, the setting is also expected to be applied to other sensors (Dames and Kumar (2015), Viseras et al. (2016)) such as laser scanner, and magnetic sensors, because it is important to sample the information near the peaks in estimating an unknown model using regression. Then, the entropy and conditional entropy can be derived as,

$$H\left[\mathbf{B}\right] = -\left\langle p(b^{1:R}), \ln p(b^{1:R})\right\rangle \tag{19}$$

$$H\left[\mathbf{B}|\hat{\mathbf{X}}^k\right] = -\int p(\hat{X}^k)\sum_{b\in\{0,1\}} p(b^{1:R}|\hat{X}^k)\ln p(b^{1:R}|\hat{X}^k)\delta\hat{X}^k, \tag{20}$$

where,

Algorithm 1 Multi-Agent Exploration

$\bar{X} \leftarrow NULL$, $\bar{Y} \leftarrow NULL$, $\bar{P} \leftarrow NULL$, $\hat{X} \leftarrow NULL$
compute $K(X_*, X_*)$
for r = 1,2,⋯,R (*parallel implementation*) **do**
  while StopCriteria **do**
    $X_{team}, \mathbf{y}_{team}, P_{team} \leftarrow GetInformation$
    $\mathbf{p}_r^k \leftarrow ReceivePacket$
    $y_r^k \leftarrow sum(\mathbf{p}_r^k)$
    $\bar{X} \leftarrow [\bar{X}, \mathbf{x}_r^k, \bar{X}_{team}]$
    $\bar{Y} \leftarrow [\bar{Y}, y_r^k, \mathbf{y_{team}}]$
    $\bar{P} \leftarrow [\bar{P}, \mathbf{p}_r^k, P_{team}]$
    $\theta^* \leftarrow OptimizeHyperparameter(\bar{X}, \bar{Y})$
    $\mu_*, \Sigma_* \leftarrow PredictiveDistributioin(X_*, \bar{P}, \theta^*)$
    $\mathbf{z}_j^k \leftarrow \arg\max_{\mathbf{x}_*}[\mu_*(\mathbf{x}_*)]_j \quad (for j = 1, \cdots, m)$
    $\hat{X}^k = GM - PHDfilter(\hat{X}^{k-1}, Z^k)$
    $Q_r \leftarrow SampleNextPosition(\mathbf{x_r^k})$
    $\mathbf{q}_r^* = \arg\max_{\mathbf{q}_r}\{\sigma(\mathbf{q}_r) + I(\mathbf{q}_r)\}$
    $BroadcastInformation(\mathbf{x}_r^k, \mathbf{q}_r^*, y_r^k, P_r^k)$
    $\mathbf{x}_r^{k+1} \leftarrow \mathbf{q}_r^*$

$$p(b = 0|\hat{X}^k) = \prod_{\hat{X}^k}\left(1 - p_d(\hat{\mathbf{x}}_j^k)\right)$$

$$\begin{aligned} p(b^1, b^2, \cdots, b^R) = \int &\prod_{r\in C_0} p(b^r = 0|\hat{X}^k) \\ &\times \prod_{r\in C_1}\left(1 - p(b^r = 0|\hat{X}^k)\right) \times p(\hat{X}^k)\delta\hat{X}^k. \end{aligned}$$

The computations for integral and the operator $\langle \bullet, \bullet \rangle$ are done by the means of nonparametric approximations, i.e. total sum of the weighted sample sets, superscript of b is the indicator of each robot, C_0 represents the estimated samples that are not detected, C_1 denotes the estimated samples that are detected by at least one robot, and $b^{1:R}$ represents all possible combinations of binary event b.

3.2 Concatenated Objective-based Control Policy

In this section, we discuss the concatenated control policy for multiple robots considering both variance and mutual information, which are discussed in Section 2.1 and Section 3.1 respectively. These objectives are widely used in many related studies such as (Viseras et al. (2016)), (Dames and Kumar (2015)), (Fink and Kumar (2010)), but most of the studies utilized only one objective of the two. Although (Fink and Kumar (2010)) estimated the RSS source location, only one RSS source is considered in the algorithm and a single optimal policy is randomly selected between variance-based one and source location estimation-based one. In real-world applications, the number of RSS source is usually unknown and the received packet length is not fixed because of the varying accessibility to RSS sources at the specific position. We consider such issues using the random sample-based nonparametric methods.

Let the future position of a robot be $\mathbf{q}_r \in Q_r$ and d be the preset radius. Then the future positions are sampled from the set

$$Q_r = \left\{\mathbf{q}_r \in Q_r \;\middle|\; \begin{array}{c} \left\|\mathbf{x}_r^k - \mathbf{q}_r\right\|_2 = d, \\ \mathbf{q}_r \in \{collision\ free\} \end{array}\right\}. \tag{21}$$

We sample six candidates of local goal positions from the set (21). To define the control policy to select the

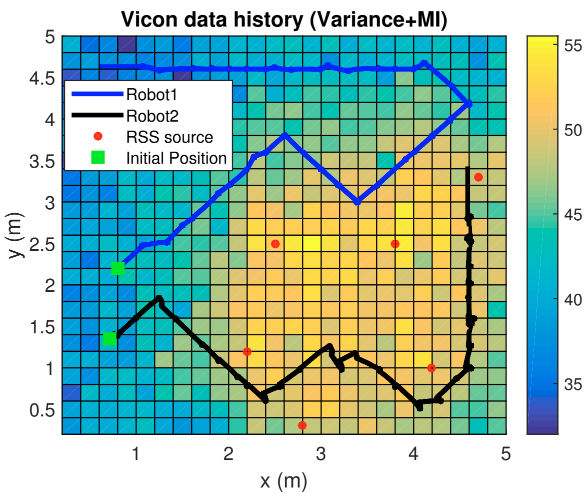

Fig. 3. Trajectories of each robot in the experiment.

optimal one from the candidates, let the normalized form of variance and mutual information be $\sigma(\mathbf{q}) \in \bar{\Sigma}$ and $I(\mathbf{q}) \in \bar{\mathbf{I}}$ respectively. Then the optimal future position is,

$$\mathbf{q}_r^* = \arg\max_{\mathbf{q}_r} \left\{ \sigma(\mathbf{q}_r) + I(\mathbf{q}_r) \right\}. \tag{22}$$

The variance term leads to the largest reduction of the variance over the entire field, and the mutual information term contributes to selecting the most informative position that increases the probability of detecting RSS sources. The algorithm is set to be terminated if the mean normalized variance is converged. The criteria, however, can be interrupted by occasional bad results of hyperparameter optimization. In such cases, the maximum number of iterations is set according to the approximate size of the exploring space.

4. EXPERIMENTAL RESULTS

We apply the proposed algorithm to the experiment that learns the environmental RSS distribution and finds the locations of RSS sources. The experimental setup is composed of two Pioneer P3-DX's which are equipped with a laptop; UBee430 motes including six source nodes and two base nodes; and a Vicon system which provides the poses of each robot. The robots share the information through TCP/IP communication using Wi-Fi, and receive Vicon data from the host PC. Each robot is controlled using ROS program with the main algorithm running on Matlab program. The UBee430 motes are programmed using TinyOS software.

We set the validation area to be a 5m × 5m square region, and step size of the robot movement to be 20cm. Also, we obtain the ground truth RSS distributions with the resolution of 20cm for each node, which are extremely noisy as shown on Fig. 2. The ground truth data is obtained by averaging 10 measurements at each positions. We provide the known ground truth for the robots as the measurements in the experiment. The covariance matrices of GM-PHD filter including the system model and observation model are set to be $[0.2^2, 0;\ 0, 0.2^2]$, and the probabilities are set to be $p_s = 0.8$, $p_D = 0.9$, $w_\gamma = 0.6$, $w_\beta = 0.4$, and $w_\gamma^0 = 0.7$ which is the normalizing factor for observation-

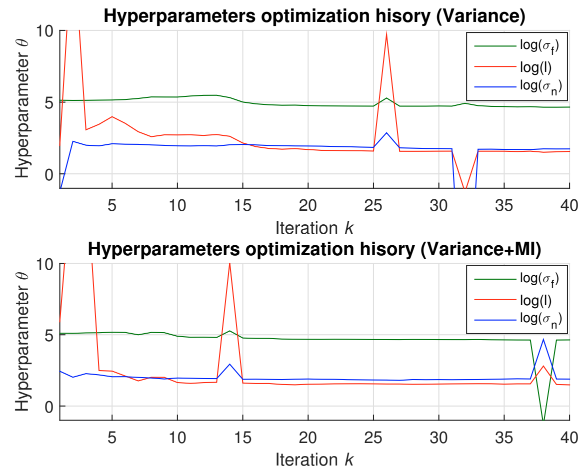

Fig. 4. Hyperparameters optimization histories of variance-based algorithm (upper) and the proposed method (lower).

based birth prior. Also, the thresholds used in computing mutual information are set to be $r_t = 0.2$ and $r_{max} = 6$.

By comparing with the entropy-based method, the performance of the proposed algorithm is validated in two aspects: the accuracy of RSS source localization and RSS distribution learning; and convergence speed of hyperparameters and RMS error. In Fig. 4, the hyperparameters converge faster when both objectives are used. The algorithm based on concatenated policy converges after visiting about 10 sampling points, while the entropy-based algorithm converges after visiting about 17 waypoints. This result suggests that the mutual information term leads to growth in robustness against the noisy measurement. The absence of prior knowledge about the environment makes it hard to find the optimal position when only the entropy is used, although it is more important to sample measurements in the most informative position at the beginning. Otherwise, the concatenated policy finds the informative position by maximizing mutual information.

In Figs. 5 and 6, we perform 10 validations using each policy, and show mean RMS error histories of the RSS source localization and RSS distribution learning. In average, the RMS error reduces faster and converges to a lower value in the concatenated policy case. Although the converged values seem to be similar in Fig. 6, the source localization performance indicates that the distribution estimation result of the concatenated policy is more desirable. The larger source localization error means that the distribution estimation for the peaks is not desirable, where the peak area is the most informative and dominant part for RSS model. The major portion of the converged RMS error in the concatenated policy comes from the extremely noisy ground truth. During the experiment, some undesirable situations are observed, for example, all the estimated samples of GM-PHD filter are eliminated at the pruning step because of very unsteady prediction results. The robots, however, still find the suboptimal action by selecting the position with maximum entropy. Overall, the proposed algorithm well performs the task, as the two objectives compensate the issues of each other.

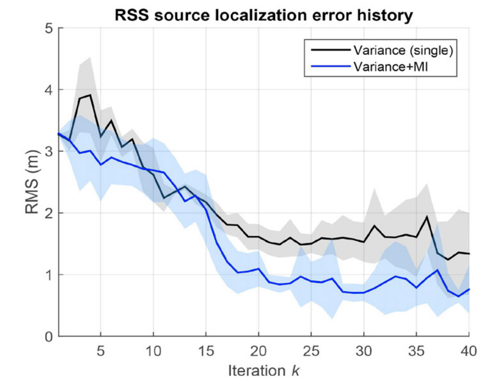

Fig. 5. Comparison of RSS source localization error histories with 10 trials.

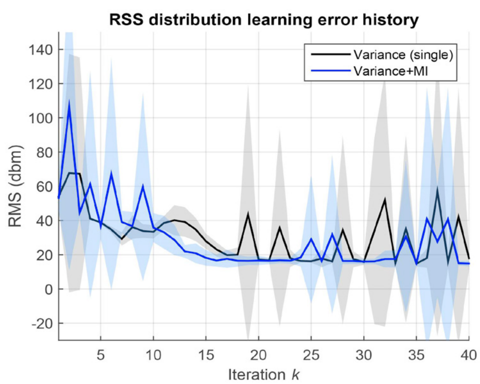

Fig. 6. Comparison of RSS distribution model error histories against the ground truth with 10 trials.

5. CONCLUSION

In this paper, a noise-robust policy for multi-agent exploration in an unknown RSS field is proposed a concatenated policy by fusing two objectives, i.e. variance and mutual information. Also, nonparametric methods are properly applied by using measurement-based spatial prior on the birth proces of GM-PHD filter. The algorithm is applied to decentralized mobile robots to find the most informative positions, while learning the unknown environment and localizing RSS sources. We compared the obtained policy with variance-based one for performance validation. The experimental result shows that the proposed algorithm provides more accurate estimations, and that the estimations converge faster. Moreover, the experiment suggests that the concatenated policy shows more robustness to the noisy signal with two objectives compensating problems of each other. For the future work, adding the self-localization step and applying the algorithm to various kinds of signals may make the algorithm more flexible.

REFERENCES

Bai, S., Wang, J., Chen, F., and Englot, B. (2016). Information-theoretic exploration with bayesian optimization.

Charrow, B., Kahn, G., Patil, S., Liu, S., Goldberg, K., Abbeel, P., Michael, N., and Kumar, V. (2015). Information-theoretic planning with trajectory optimization for dense 3d mapping. In *Proceedings of Robotics: Science and Systems*.

Charrow, B., Michael, N., and Kumar, V. (2014). Cooperative multi-robot estimation and control for radio source localization. *International Journal of Robotics Research*, 33(4), 569–580.

Clark, D.E., Panta, K., and Vo, B.N. (2006). The gm-phd filter multiple target tracker. In *2006 9th International Conference on Information Fusion*, 1–8. IEEE.

Cover, T.M. and Thomas, J.A. (2012). *Elements of information theory*. John Wiley & Sons.

Dames, P. and Kumar, V. (2015). Autonomous localization of an unknown number of targets without data association using teams of mobile sensors. *IEEE Transactions on Automation Science and Engineering*, 12(3), 850–864.

Dames, P., Tokekar, P., and Kumar, V. (2015). Detecting, localizing, and tracking an unknown number of moving targets using a team of mobile robots. *Intl. Sym. Robot. Research*.

Ferris, B., Haehnel, D., and Fox, D. (2006). Gaussian processes for signal strength-based location estimation. In *In proc. of robotics science and systems*. Citeseer.

Fink, J. and Kumar, V. (2010). Online methods for radio signal mapping with mobile robots. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, 1940–1945. IEEE.

He, S. and Chan, S.H.G. (2016). Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Communications Surveys & Tutorials*, 18(1), 466–490.

Hoffmann, G.M. and Tomlin, C.J. (2010). Mobile sensor network control using mutual information methods and particle filters. *IEEE Transactions on Automatic Control*, 55(1), 32–47.

Houssineau, J. and Laneuville, D. (2010). Phd filter with diffuse spatial prior on the birth process with applications to gm-phd filter. In *Information Fusion (FUSION), 2010 13th Conference on*, 1–8. IEEE.

Kim, W., Lee, H., and Kim, H.J. (2013). Predictive modeling of time-varying environmental information for path planning. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 3639–3644. IEEE.

Ouyang, R., Low, K.H., Chen, J., and Jaillet, P. (2014). Multi-robot active sensing of non-stationary gaussian process-based environmental phenomena. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 573–580. International Foundation for Autonomous Agents and Multiagent Systems.

Rasmussen, C.E. (2006). Gaussian processes for machine learning.

Viseras, A., Wiedemann, T., Manss, C., Magel, L., Mueller, J., Shutin, D., and Merino, L. (2016). Decentralized multi-agent exploration with online-learning of gaussian processes. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 4222–4229. IEEE.

Vo, B.N. and Ma, W.K. (2006). The gaussian mixture probability hypothesis density filter. *IEEE Transactions on signal processing*, 54(11), 4091–4104.

Vo, B.N., Singh, S., and Doucet, A. (2005). Sequential monte carlo methods for multitarget filtering with random finite sets. *IEEE Transactions on Aerospace and electronic systems*, 41(4), 1224–1245.