Sampling points optimization deployment for water environment monitoring based on PSO algorithm

To cite this article: Rong Zhang et al 2018 J. Phys.: Conf. Ser. 1074 012169

View the article online for updates and enhancements.
Sampling points optimization deployment for water environment monitoring based on PSO algorithm

Rong Zhang, Jie Zhang, Yan Shen¹ and Qixin Xu

School of Control Engineering, Chengdu University of Information Technology, ChengDu 610225, China

¹ E-mail: sheny@cuit.edu.cn

Abstract. In this paper, the problem of deploying the sampling points for water environment monitoring is discussed. The mean square error (MSE) of reconstructing environmental scalar field is used as the evaluation metric of the monitoring performance, and particle swarm optimization (PSO) algorithm is applied for obtaining the near-optimal set of sampling locations. The water quality characteristics of the whole area is reconstructed by the limited samples. The simulation results show that the novel method is more efficient than the deployment method of random Gaussian distribution for recovering the aquatic phenomenon, and the reconstructed accuracy of scalar field is obviously improved.

1. Introduction

With the exploitation of water resources, the aquatic environment is polluting and it is in danger. Generally, water environment monitoring mainly adopts traditional methods, which have many shortcomings such as wasting resources, poor real-time etc. Obviously, the traditional method cannot meet the requirement of real-time monitoring. Therefore, it is significant to develop the automatic monitoring technology.

With the advance in the integration of robot technology and Wireless Sensor Networks (WSNs), the underwater Robotic Fish with water quality detection sensors can be put into lakes or rivers to form a WSN mobile platform for water environment monitoring [1]. It can provide meaningful information for monitoring tasks through the analysis of the sampled data from these sensors in pollution area.

The physical phenomenon is characterized by scalar field such as temperature, turbidity, pH [2], and the correlation of water quality characteristics decreases with the increasing of the different positions distance. The purpose of water environmental monitoring is not only getting the values of sampling points, but also reconstructing the environmental field through these limited samples, so that we can gain the distribution of water quality characteristics in the whole monitoring area. In the following, how to acquire the most reliable reconstruction performance with the optimal sampling locations deployment is the focus of this research.

At present, the configuration of the sampling locations for aquatic monitoring focused on the uniform deployment of WSNs. For example, some articles adopted an approach which disposed the sampling points uniformly based on a Lawn Mower Sampling (LMS) pattern. In [3], this strategy was used to analyze the adaptive sampling methodologies for AUVs and the authors also discussed the sampling path energy of it. In [4], the method was used to minimize the maximum reconstruction error and the energy consumption. However, the situation which the aquatic spatial phenomena distribution is non-uniform haven’t been considered. If the aquatic characteristics in the spatial distribution is non-
uniform, and the number of samples is limited, the uniform deployment of sampling points in the field may not always be effective.

In order to make up for the shortage of sampling locations uniform distribution, [5] adopted a random Gaussian distribution to determine the location of the sampling points so that more samples could be obtained in the field near the hot spot. [6] proposed a method of choosing sampling points set which maximized the mutual information of the sampled locations. And the sampling values which are obtained from those locations could provide a better recovery of scalar field. The Fuzzy C-Means (FCM) algorithm was used in [7], which can find out the centroid positions of the aquatic environment. Then those positions were used as the next navigation points of Underwater Vehicle, so that the task of monitoring and sampling was completed. In [8], Kriging variance as the selection criteria of sampling points, the prior information of marine characteristics was utilized, and a greedy algorithm was proposed to search the sample points so that the optimal location information was obtained. However, these papers haven’t illustrated the actual distribution of the sampling points in the whole environment explicitly.

In the following sections, Section 2 presents an assumed model of steady-state diffusion environmental scalar field. In section 3, the fitness function is proposed and the PSO algorithm is illustrated for optimizing the sampling points. Then, the simulation results are discussed in section 4. Finally, Section 5 concludes the main work.

2. Assumptions of static aquatic environmental field model

According to the Fick’s law in [9], in the water, the diffusion process of temperature, salinity, oil and other substances is divided into two types: one is the concentration of diffusion material varies only with the diffusion distance rather than the time, which is described as steady-state diffusion; the other is the concentration of diffusion material varies not only with the diffusion distance but also with the time, which is known as non-steady-state diffusion.

It is assumed that the aquatic area $S$ is a two-dimensional plane and usually meshed, and a release of $M$ kg pollutants occur at location $(\mu_x, \mu_y)$. In general, each grid point carries information $(x, y, Z)$, where $(x, y)$ denotes the position coordinate and $Z$ denotes the data value corresponding to the position. The molecular diffusion is isotropic; that is to say, the dispersion degree of each direction of the contaminant is the same and variance is $\sigma^2$. In the case of the steady-state diffusion environment, the field $Z$ can be formed, where probability density function $f(x, y)$ is approximately Gaussian [10] as formula (1):

$$Z = f(x, y) = \frac{M}{2\pi\sigma^2} \exp \left[ -\frac{1}{2} \left( \frac{L^2}{\sigma^2} \right) \right]$$

(1)

Where $L = \sqrt{(x - \mu_x)^2 + (y - \mu_y)^2}$, means the Euclidean distance of any point $(x, y)$ to the source location $(\mu_x, \mu_y)$ on the surface of the field.

3. Optimization Deployment of Sampling Points

3.1. Determination of Fitness Function

In order to reconstruct environmental scalar field, it is necessary to use the numerical estimation algorithm to estimate the values of the un-sampled locations by the existing samples. For example, when we have acquired the sampling data $(x_d, y_d, Z_d), d = 1, 2, \cdots, D$ from the locations that had been deployed, the griddata interpolation algorithm would be adopted to reconstruct the environment scalar field [11]. Then the scalar field $Z$ is comprised of the sampling values $Z_d$ and the the estimated values $\hat{Z}_d$. The more samples are near the location of being estimated, the smaller the reconstructed error is. Thus a performance metric must be utilized to evaluate the quality of reconstructed field.
In this paper, Mean Square Error (MSE) is adopted, which described the relationship between the estimated value and the true value. The smaller MSE is, the higher accuracy of estimation can be obtained. That is to say, the reconstructed scalar field can reflect the water quality information more accurately.

This work takes the MSE of reconstructed scalar field as the evaluation metric of the monitoring performance. Meanwhile, the fitness function of the PSO algorithm is formulated, such as the formula (2):

$$fitness = \frac{1}{I} \sum_{i=1}^{I} (Z_i - Z_e)^2$$  \hspace{1cm} (2)

Where $Z_e$ is the actual value of estimated position $(x_e, y_e)$ in the environmental scalar field. $Z_i$ is the estimated value form the position point $(x_i, y_i)$. $I$ is the number of position points for error analysis.

It is a NP problem, which selects a certain number of sampling locations from the aquatic environment and minimizes the MSE of estimation at the un-sampled positions by using the values of these sampled locations. PSO algorithm is a swarm intelligence optimization and it is derived from Kennedy's research [12]. It has been widely used in optimization problems such as Neural Network Training [13], parameters optimization [14] and sensor deployment [15]. In our work, we use PSO algorithm to find the near-optimal set of sampling points to minimizing the scalar field reconstruction error.

### 3.2. Optimization of PSO algorithm

It is assumed that $D$ sampling points are deployed on the two-dimensional plane $S$ of the aquatic area, and each sample point contains information of $(x_d, y_d, Z_d), d = 1, 2, \cdots, D$. In the PSO algorithm, each particle represents a set of sampling points $\overrightarrow{X}_i = (X_{i1}, X_{i2}, \cdots, X_{iD}), \text{where } X_{id} = (x_{id}, y_{id}), i = 1, 2, \cdots, N, d = 1, 2, \cdots, D$. The position $X$ and velocity $V$ of the $N$ particles are shown as follows:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} X_{i1} & X_{i2} & \cdots & X_{iD} \\ X_{21} & X_{22} & \cdots & X_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{ND} \end{pmatrix} \quad V = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{pmatrix} = \begin{pmatrix} V_{i1} & V_{i2} & \cdots & V_{iD} \\ V_{21} & V_{22} & \cdots & V_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ V_{N1} & V_{N2} & \cdots & V_{ND} \end{pmatrix}$$  \hspace{1cm} (3)

Two optimal solutions can be gained in each iteration: one is the optimal solution of each individual particle in the history searching process, $\overrightarrow{p_{best}} = (p_{i1}, p_{i2}, \cdots, p_{iD}), i = 1, 2, \cdots, N$, which is called the individual historical optimal solution; the other is the optimal solution of all the particles in the past searching process, $\overrightarrow{g_{best}} = (g_1, g_2, \cdots, g_D)$, which is called the global optimal solution. Meanwhile, each particle updates its position and velocity with these two optimal solutions continuously. The velocity and position of the particles are updated by the following formula (4) and (5):

$$V^T_i = w * V^T_i + c_1 * r_1 (\overrightarrow{p_{best}} - \overrightarrow{X}_i^T) + c_2 * r_2 (\overrightarrow{g_{best}} - \overrightarrow{X}_i^T)$$  \hspace{1cm} (4)

$$\overrightarrow{X}_i^T = \overrightarrow{X}_i^T + V^T_i$$  \hspace{1cm} (5)

Where $i = 1, 2, \cdots, N, w$ is inertia weight, $c_1$ and $c_2$ are learning factor and they are constant, $r_1$ and $r_2$ are the uniform random numbers for $[0, 1]$.

$\overrightarrow{g_{best}}$ is the globally optimal resolution, and it is searched continuously by the iterative optimization process of PSO algorithm. The solution is the best vector which meets with the formula (2).

Then, the process of sampling points optimization deployment based on PSO algorithm is illustrated as follows:
① The position and velocity of the particle are initialized based on the random Gaussian distribution. The number of particles is $N$, the number of sampling points is $D$ and the matrix of position and velocity is described with formula (3).

② The fitness value $\text{fitness}_i$ of each particle $i = 1, 2, \cdots, N$ is calculated by equation (2), individual historical optimal solution of particles $p_{\text{best}_i}$ and its corresponding fitness value $\overline{\text{fitness}}_{p_{\text{best}_i}}$ are initialized, meanwhile, the global optimal solution of particles $g_{\text{best}}$ and the corresponding fitness value $\overline{\text{fitness}}_{g_{\text{best}}}$ are initialized.

③ the maximum number of iterations $G_{\text{max}}$ is assumed.

④ According to the formula (4), (5), the position and velocity of each particle are adjusted.

⑤ According to the formula (2), the fitness values $\overline{\text{fitness}}_{x_i}$ of all the new particles $\{x_i\}_{i=1}^N$ are calculated.

⑥ Compare $\overline{\text{fitness}}_{x_i}$ with $\overline{\text{fitness}}_{p_{\text{best}_i}}$, if $\overline{\text{fitness}}_{x_i} < \overline{\text{fitness}}_{p_{\text{best}_i}}$, make $\overline{\text{fitness}}_{p_{\text{best}_i}} = \overline{\text{fitness}}_{x_i}$, $p_{\text{best}_i} = \overline{x}_i$.

⑦ Compare $\overline{\text{fitness}}_{x_i}$ with $\overline{\text{fitness}}_{g_{\text{best}}}$, if $\overline{\text{fitness}}_{x_i} < \overline{\text{fitness}}_{g_{\text{best}}}$, make $\overline{\text{fitness}}_{g_{\text{best}}} = \overline{\text{fitness}}_{x_i}$, $g_{\text{best}} = \overline{x}_i$.

⑧ If $G_{\text{max}}$ is reached, the process of algorithm optimization is terminated, the optimal particle, i.e. the coordinate set of sample points $g_{\text{best}}$ and the corresponding fitness value $\overline{\text{fitness}}_{g_{\text{best}}}$ are acquired. Otherwise, returned ④.

4. Simulation Results

In the experiment, we evaluate the proposed method by being compared with the deployment method of random Gaussian distribution for deploying sampling points. Assuming that the area of static aquatic environmental field is 100m×100m and a release of $M = 10000$kg pollutants occur at a location $(50, 50)$. We use $\mu_x = \mu_y = 50$, $\sigma^2 = 50$ in a Gaussian function to generated a 2D environment scalar field based on the steady-state diffusion model in formula (1), which is shown in figure1.

![Figure 1. 2D environment scalar field with pollutant distribution.](image)

Figure 2 shows the method of random Gaussian distribution, in which (a) depicts that 100 sampling points are placed by random Gaussian distribution whose parameters are $\mu_x = \mu_y = 50$ and $\sigma^2 = 400$, and (b) represents the reconstructed scalar field which is generated by the samples from (a).
Comparing with the true field as shown in figure 1, there is a lot of reconstruction uncertainty in peripheral blue circle because the sampling points may be not distributed rationally.

**Figure 2.** The sampling points distribution and reconstructed scalar field of random Gaussian distribution.

Figure 3 shows the method of PSO algorithm, whose parameters are given as [15]: $c_1 = 2$, $c_2 = 2$, $w = 0.5$, particle number $N = 10$, iterations $G = 100$. The graph of the sampling points distribution is shown as (a) and the corresponding reconstructed scalar field is acquired in (b). Comparing with the result in figure 2, our method provides better estimated values of un-sampled locations by changing the pattern of deployment. It achieves a better reconstructed accuracy of environmental field, where more sampling points are deployed in regions of high gradient.

**Figure 3.** The sampling points distribution and reconstructed scalar field of PSO algorithm.

More experiments were carried out on these two methods separately to compare the performance of reconstructing the field and the MSE comparison diagram is shown in figure 4, in which $n$ is the number of experiments. As a result, the better estimation can be achieved with optimal sampling locations which are redeployed by PSO algorithm, and the reconstructed accuracy is obviously improved.
5. Conclusions
In this paper, it takes the steady-state diffusion environment as the application background, the MSE of environment scalar field reconstruction is used as the performance metric, and the related fitness function is established. The sampling points are initially deployed by the random Gaussian distribution and these points are reconfigured by PSO algorithm again, so that the near-optimal set of sampling points is obtained by updating the sampling locations continuously in the process of algorithm iteration. In this way, the distribution of water quality characteristics can be reconstructed based on limited samples. The simulation results show that the proposed method is more efficient than the deployment method of random Gaussian distribution to recovery the aquatic phenomenon, and the reconstructed accuracy of scalar field is obviously improved. It is of great application value to the water environmental monitoring task with non-uniform aquatic phenomena distribution.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (61472050), the Key project of Sichuan Province (2017GZ0004) and the key project of Education Department of Sichuan Province (17ZA0057).

References
[1] Shen Y, Xu Q, Zhang J, Zhang R and Li J 2016 Target tracking with energy efficiency using robotic fish-based sensor networks Int. Conf. on Embedded Software & System. IEEE Computer Society pp 24-9
[2] Wang Y, Tan R, Xing G, Wang J and Tan X 2014 Profiling aquatic diffusion process using robotic sensor networks J. IEEE Transactions on Mobile Computing 13 880-93
[3] Mora A, Ho C and Saripalli S 2013 Analysis of adaptive sampling techniques for underwater vehicles J. Autonomous Robots 35 111-22
[4] Chen B, Pandey P and Pompili D 2012 An adaptive sampling solution using autonomous underwater vehicles Ifac Proc 45 352-56
[5] Wang Y, Wilkerson M and X Yu 2011 Hybrid sensor deployment for surveillance and target detection in wireless sensor network Proc. on 7th Wireless Communication and Mobile Computing Conf. (IWCMC) (Istanbul, Turkey) pp 326-30
[6] Jonsson P B, Wang J and Kim J 2017 Scalar field reconstruction based on the gaussian process and adaptive sampling Int. Conf. on Ubiquitous Robots and Ambient Intelligence pp 442-45
[7] Cococcioni M, Lazzerini B and Lermusiaux P F J 2015 Adaptive sampling using fleets of underwater gliders in the presence of fixed buoys using a constrained clustering algorithm
Conf. Oceans IEEE pp 1-6

[8] Zhu X, Yu J, Ren S and Wang X 2010 Near-optimal collecting data strategy based on ordinary kriging variance Conf. Oceans IEEE pp1-6

[9] Kaufmann R S 1998 Fick’s law Springer (Netherlands)

[10] Lu YM, Dragotti PL and Vetterli M 2011 Localizing point sources in diffusion fields from spatiotemporal samples Proc. of Int. Conf. on Sampling Theory and Applications (SampTA)

[11] VK Hombal 2009 Dissertations & Theses – Gradworks

[12] Kennedy and James 2011 Particle swarm optimization Encyclopedia of machine learning Springer (US) pp 760-66

[13] Taormina R and Chau K W 2015 Neural network river forecasting with multi-objective fully informed particle swarm optimization Journal of Hydroinformatics 17(1) pp 99-113

[14] Liu J, Liu Z and Xiong Y 2013 Method of parameters optimization in svm based on pso Transactions on Computer Science & Technology 2(1) pp 9-16

[15] Yu X Y, Sun Q, Wang X Y, Xu J P, Wang L and Zhang H Y 2016 Research on optimization deployment of water quality sensor based on particle swarm optimization algorithm Transducer & Microsystem Technologies