1P2-13

Statistical Generation of Echo Waveform Obtained by a Single Acoustic Transmitter and Receiver at Unmeasured Locations

Atsushi Tsuchiya[1†], Naoto Wakatsuki[1*], Tadashi Ebihara[1], Keiichi Zempo[1] and Koichi Mizutani[1] ([1]Univ. Tsukuba)

1. Introduction

Autonomous mobile robots are being introduced to automate factory automation and construction. Autonomous robots need to estimate their position in order to plan their path. Light detection and ranging (LiDAR), millimeter wave radar, and ultrasonic sensors have been proposed as typical sensors for self-position estimation1[1]. In particular, ultrasonic sensors have been adopted as sensing devices for various autonomous mobile applications because they can be installed at a lower cost than other methods. On the other hand, ultrasonic sensors have poorer angular resolution than other sensors. They are also prone to false detection of reflected waves due to multipath interference. These problems make it challenging to use ultrasonic sensors when building advanced autonomous mobility applications.

Therefore, we proposed a method to measure reflected waves and estimate self-position using the Doppler effect[2]. This method estimates self-position by matching echo images obtained by Doppler shift and impulse response measurements with map images obtained by pre-measurement. The map image is the sum of the echo images at the position coordinates accurately measured during the mapping phase. This map can predict the echo images measured at arbitrary position coordinates for which no echo images have been measured. This mapping method can use the features of single reflections and retroreflections, but it loses the features of multiple reflections. This paper proposes a statistical method to generate predictive echo images with high similarity to measured echo images. The finite difference time domain (FDTD) method is used to generate training and evaluation data for the echo images, and the similarity between the predicted echo image output by the image generation model and the echo image obtained by the FDTD method is evaluated.

2. Proposed method

Figure 1 shows a schematic diagram of the proposed method. The proposed method trains an

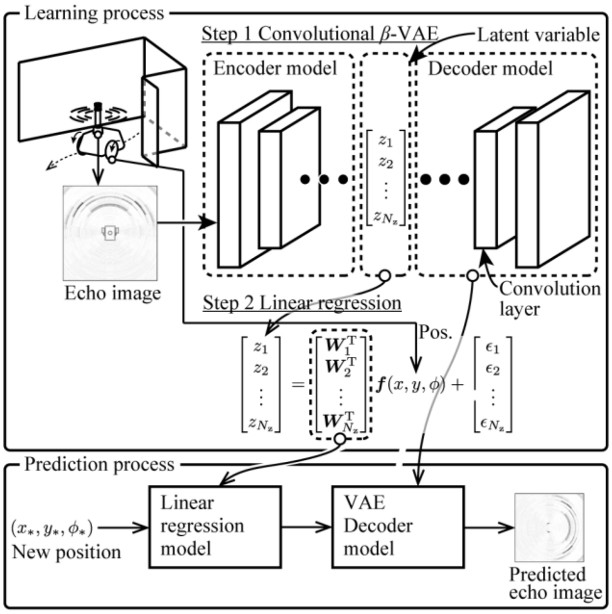

Fig. 1 Schematic diagram of the proposed method.

image generation model using as input a set of data corresponding to the echo images measured by the Doppler-compensated cross-correlation[2] method and the position coordinates at which the echo images were measured. The image generation model comprises a convolutional variational autoencoder (Convolutional β-VAE)[3] and a linear regression model with a polynomial basis function. The training of the proposed method is performed in two steps. First, unsupervised convolutional β-VAE encoder and decoder training is performed on echo images. This unsupervised dimensionality compression method maps the features of the input image H to a low-dimensional latent variable $\boldsymbol{z} \in \mathbb{R}^d$, where $\boldsymbol{z}$ is constrained to behave as a random variable following a multivariate normal distribution, and the model is trained. After the training of the convolutional β-VAE is completed, supervised learning of the linear regression model is performed. The basis function is denoted by

$$\boldsymbol{f}(x, y, \phi) = [1 \quad x \quad \cdots \quad x^Q \quad y \quad \cdots \quad y^Q \quad \phi \quad \cdots \quad \phi^Q]^{\mathrm{T}}. \quad (1)$$

The objective variable of the linear regression model is the latent variable $\boldsymbol{z}$, which is output when the echo image is input to the β-VAE model, and the explanatory variables are the position coordinates x,

E-mail: †tsuchiya@aclab.esys.tsukuba.ac.jp,
*wakatuki@iit.tsukuba.ac.jp

y, and ϕ at which the echo image was measured.

The trained image generation model is used in the following steps. First, the position coordinates x_*, y_*, and ϕ_* to be predicted are input into the linear regression model. The predicted echo image is the resulting output image when the latent variable $\boldsymbol{z}_*$ is input to the decoder model.

3. Simulation setup

The accuracy of the proposed method was verified by FDTD simulation[4]: echo images were measured while moving the transmitter and receiver point positions within the FDTD simulation, and the echo images corresponding to the transmitter and receiver point positions were used as the data set. **Figure 2(a)** shows the path of movement of the transmitter/receiver used to create the training data and the position of the wall surface. Figure 2(b) shows the path of movement of the transmitter/receiver and the positions of the wall surface used to create the evaluation data. **Table I** shows the simulation conditions and signal processing parameters.

4. Results and discussions

We evaluated the similarity of the predicted echo images generated by calculating the mutual information content between the actual and predicted echo images observed in the simulation. When echo images are generated for self-position estimation, the similarity should be high when the actual position where the echo image was measured and the position where the predicted echo image was generated match. Therefore, we define an index to evaluate the ability of the generated images to identify space. It is denoted by

$$I_{\mathrm{diff}} = I[H_{\mathrm{t}}(x_*, y_*, \phi_*); H(x_*, y_*, \phi_*)] - I[H_{\mathrm{t}}(x_* + w, y_* + w, \phi_*); H(x_*, y_*, \phi_*)], \quad (2)$$

where I is the mutual information content, $H_{\mathrm{t}}(x_*, y_*, \phi_*)$ is the actual echo image, $H(x_*, y_*, \phi_*)$ is the predicted echo image, and w is a random variable according to $\mathcal{N}(0,1)$. **Figure 3** shows a histogram of I_{diff}. The shape of the distribution of the proposed method extends to larger values compared to the existing methods; the size of I_{diff} indicates the spatial identification capability of the generated images. This capability is crucial for accurate self-position estimation in autonomous mobile robotics, and our method's superior spatial identification capability suggests its potential to improve such systems' performance significantly.

5. Conclusion

This paper proposes a statistical method for generating predictive echo images for self-position estimation. The proposed method is based on convolutional β-VAE and linear regression model. The validation results using FDTD simulations suggest that the proposed method has a better spatial identification capability than the existing methods.

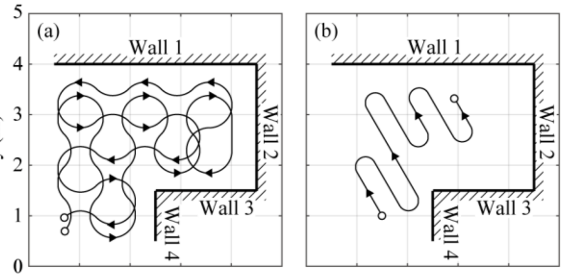

Fig. 2 (a) Path of movement of the transmitter/receiver used to create the training data, (b) path of movement of the transmitter/receiver used to create the evaluation data.

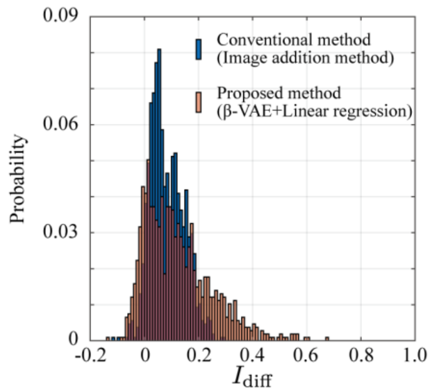

Fig. 3 Histogram of I_{diff}.

Table I Signal processing parameters.

Sampling frequency (kHz)	f_{s}	40
Sequence length	L	1024
Carrier frequency (kHz)	f_{c}	10
Chip rate (kHz)	$1/T_{\mathrm{c}}$	10
Number of iterations	M	6
Number of arrival directions	N	29
Dimension of latent variables	d	16
Degree of polynomial	Q	6

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 22KJ0431.

References

1) A. Tsuchiya, N. Wakatsuki, T. Ebihara, K. Zempo and K. Mizutani: *Jpn. J. Appl. Phys*. **61** (2022) SG1037.
2) A. Tsuchiya, N. Wakatsuki, T. Ebihara, K. Zempo and K. Mizutani: *IEEE J. Indoor and Seamless Pos. and Navi*., 2, pp. 193-204, 2024.
3) Irina Higgins et al., *Int. Conf. Learning Representations*, 2017.
4) T. Tsuchiya and M. Kanamori: *Jpn. J. Appl. Phys*. 60 (2021) SDDB02.