Millimeter-Wave Coherent Imaging of Moving Targets by Using Complex-Valued Self-Organizing Map and Auto-Encoder

Yuya Arima and Akira Hirose, *Fellow, IEEE*

***Abstract*—Active millimeter-wave imaging systems using complex-valued self-organizing map (CSOM) have potentially wide applications in moving-target imaging by acquiring and classifying complex textures. However, interference in the coherent observation causes harmful effects such as amplitude fluctuation and phase distortion, resulting in the deterioration of the visualization quality achieved by the CSOM. In this article, we propose the introduction of feature refinement using a complex-valued auto-encoder (AE) into the feature extraction process of our imaging system to suppress these effects. We show that the complex-valued AE is extremely useful for extracting features appropriately even under such adverse influences, and improves the clustering performance. We also experimentally investigate the influence of the number of hidden-layer neurons on the AE performance to discuss the robustness of our system.**

***Index Terms*—Auto-encoder (AE), complex-valued neural network, millimeter-wave imaging, security, self-organizing map (SOM).**

I. Introduction

MANY engineering fields employ various imaging techniques to obtain physical information on targets. Most imaging systems assume that a person looks at the output images to identify included targets [1]. For this purpose, it is necessary to represent the features of objects more precisely and clearly. Thus, research has focused on enhancing the resolution and signal-to-noise ratio (SNR). However, such measures of enhancement are limited by the increase in the cost of equipment and physical restrictions of sensor structures. In addition, simple improvement of the resolution causes the growth of the amount of data to be processed and leads to an increase in the processing time.

Human perception has a high ability in visual completion and noise suppression. Therefore, it is possible for us to identify objects and read characters even in low-resolution and/or high-noise images. Neural network systems, which mimic such superior cognitive functions, show a capability in compensating for a lack in an image to identify objects on the basis of features. It is also possible for such systems to realize higher functions by using or adding information that cannot be originally perceived by human beings, namely, phase and polarization. In the use of phase information in addition to signal intensity, there is an approach to treat the signals in the complex domain with neural networks extended to complex representations. For example, there are studies on classifying complex data observed by space-borne synthetic aperture radars (SARs) with complex-valued neural networks [2], [3].

Our group has also proposed neural networks that utilize phase information, in addition to the signal intensity (amplitude) obtained in coherent measurement by using complex-valued convolutional neural networks, for SAR data structurization [4]–[6]. In other words, we introduced adaptive discrimination neural networks that learn the local complex texture (phase and amplitude) as a local distinction feature [7], [8]. By introducing the phase information, we can obtain the transmission/reflection characteristics of a target substance as well as its distance. By evaluating the complex amplitude locally, we can estimate 3-D local unevenness and/or demonstrate high compensation ability [9]. Using complex-valued neural networks, we can propose various techniques that work better than conventional methods to distinguish targets from their silhouettes and/or the fine texture of objects, even in low-resolution and low-SNR observations.

Recently, we have proposed an imaging system using millimeter wave as the sensing medium [10]. Millimeter wave is an electromagnetic wave whose wavelength is on the order of millimeters, as its name implies. Millimeter wave is used for both active and passive imaging because of its high rectilinearity and its frequency lying in the blackbody-radiation band at room temperature [11], [12]. A millimeter wave also has high transparency in various dielectric materials such as plastics, clothing, and paper. Thus, it is also used for whole-body scanners that detect weapons hidden under clothes at airports [1]. The same function is realizable with a backscatter X-ray inspection device [13], but millimeter-wave systems have the major advantage of safety for the human body.

However, both types of systems require a long scanning time in which a target person should stay still. This becomes a problem when we expand the application to a situation, where more people need to be processed, such as a station ticket gate. Therefore, systems for measuring moving pedestrians have been

Manuscript received May 20, 2019; revised January 16, 2020; accepted January 28, 2020. Date of publication February 12, 2020; date of current version May 15, 2020. This work was supported by JSPS KAKENHI under Grants 21300089 and 18H04105. *(Corresponding author: Yuya Arima.)*

The authors are with the Department of Electrical Engineering and Information Systems, The University of Tokyo, Tokyo 113-8656, Japan (e-mail: y_arima@eis.t.u-tokyo.ac.jp; ahirose@ee.t.u-tokyo.ac.jp).

Digital Object Identifier 10.1109/JSTARS.2020.2971275

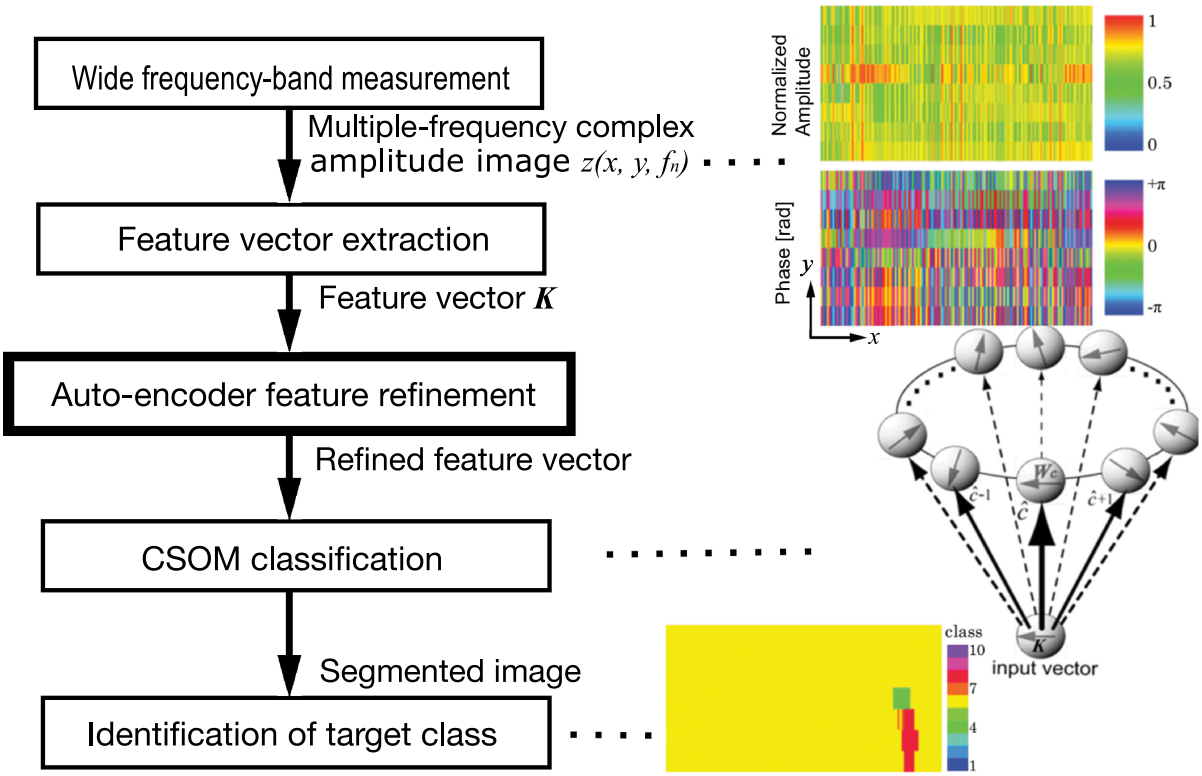


Fig. 1. Total flowchart of scattering acquisition and data processing [18].

developed. For example, there are studies of using W-band radar [14], using UWB Doppler radar [15], etc. However, these systems have the disadvantage that the resolution is lower than that of systems for stationary objects.

The system we proposed [10] solves this problem by introducing a neural network to adaptively identify the complex textures from low-resolution data. This elimination of necessity to acquire high-resolution data can also solve privacy problems. For pedestrian applications and on-vehicle use, we have developed a prototype real-time imaging system composed of a millimeter-wave active radar part for capturing moving objects that pass in front of an array antenna, and a processing part to identify targets by using a complex-valued self-organizing map (CSOM). We have evaluated its basic performance [16]. We have also examined in detail the effect of parameters in the system configuration and image processing on the visualization accuracy, and analyzed how to improve the accuracy [17]. We also reported the improvement of the clustering performance of the CSOM for distinguishing targets as a result of introducing a complex-valued auto-encoder (AE) in the process of extracting the complex texture of a local region [18].

In this article, we systematically investigate in detail the feature refinement using the complex-valued AE by focusing on the number of hidden-layer neurons of the AE, which determines the dimension of the refined features. Our system performs coherent millimeter-wave imaging. Coherent imaging is susceptible to adverse effects of interference in the measurement and suffers from amplitude fluctuation and phase distortion. These effects are particularly serious in measurements of moving objects. The AE is useful for extracting appropriate features under such harmful effects. We show that the AE is extremely effective for millimeter-wave active imaging when the number of hidden-layer neurons is set appropriately.

This article is constructed as follows. Section II outlines the system construction and the total processing flow. Section III explains how the system captures coherent image data. Section IV details the AE in our imaging system. Section V describes the feature refinement in the AE. Section VI presents adaptive imaging results. Then, we discuss the relationship between the number of the hidden-layer neurons and the degree of independence among the neuron weights in Section VII. Section VIII concludes this article.

II. System Construction for Local Texture Classification by CSOM

We use a complex-valued neural network to handle phase and amplitude data obtained by coherent measurement. In our system, we use a CSOM, which is a complex-valued version of a self-organizing map (SOM). A CSOM is effective in ground-penetrating radar systems, which visualize landmines and/or other targets buried underground [9], [19], [20]. We use a ring CSOM, whose network configuration is a ring, to project feature-vector similarity in 1-D space.

The processing flow is shown in Fig. 1. First, coherent measurements are made at several modulation frequencies (f_n). Obtained data at each modulation frequency has amplitude and phase components as shown in the upper right part of Fig. 1. Feature vectors $\boldsymbol{K}$ are extracted from this data. The feature extraction method is described in detail in the next paragraph.

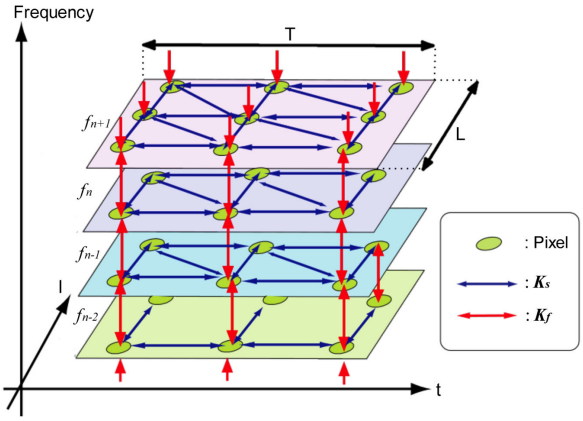


Fig. 2. Spatial and frequency domain correlation feature extraction in local window [17].

In our conventional system [17], we feed the extracted feature vectors $\boldsymbol{K}$ directly to the CSOM to perform self-organization and classification. In this study, we introduce a feature refinement module, namely, a complex-valued AE, before feeding the refined feature vectors to the CSOM. Then, a segmented image as shown in the lower right part of Fig. 1 is obtained by assigning colors corresponding to the CSOM classes to appropriate pixels.

We obtain a set of 3-D data, namely, two spatial dimensions and one frequency dimension, by modulation-frequency stepped observation. We extract feature vectors by calculating correlations among pixels in a local window as shown in Fig. 2. The extraction procedure is as follows.

1) Cut out a local window of $L \times T$ size (L, T: window size).
2) Construct a feature vector $\boldsymbol{K}$ by averaging M, the spatial correlation $K_s(i,j)$, and the interfrequency correlation $K_f(f_n)$ calculated with (2)–(4) as follows:

$$\boldsymbol{K} = \big[M\, K_s(0,0)\, K_s(0,1)\, K_s(1,0)\, K_s(1,1) \times K_f(f_1) \ldots K_f(f_{N_f-1})\big]^{\top} \tag{1}$$

$$M = \frac{1}{LTN_f} \sum_{l=1}^{L} \sum_{t=1}^{T} \sum_{n=1}^{N_f} z(l,t,f_n) \tag{2}$$

$$K_s(i,j) = \frac{1}{(L-i)(T-j)N_f} \sum_{l=1}^{L-i} \sum_{t=1}^{T-j} \sum_{n=1}^{N_f} \times z^*(l,t,f_n) z(l+i,t+j,f_n)\ (i,j \in \{0,1\}) \tag{3}$$

$$K_f(f_n) = \frac{1}{LT} \sum_{l=1}^{L} \sum_{t=1}^{T} z^*(l,t,f_n) z(l,t,f_{n+1}) \tag{4}$$

where $z(l,t,f_n)$ is the complex amplitude measured at antenna l, time t, and modulation frequency f_n, which is normalized such that $0 \le |z| \le 1$; z^* stands for complex conjugate of z; and N_f is the total number of modulation frequency points. The feature vectors are composed of one complex mean, four spatial correlations, and $(N_f - 1)$ interfrequency correlations. That is, the feature vectors have $(N_f + 4)$ dimensions.

III. Coherent Observation

The coherent image data for classification is obtained using our prototype millimeter-wave active radar unit shown in Figs. 3 and 4 [10], [16], [17]. Fig. 3 shows a schematic diagram of the signal processing flow, and Fig. 4 is a photo of the front end including the transmitting horn antenna and the 16-parallel receiving antenna elements. The unit has an array of 16 parallel receiving elements composed of bulk linear tapered slot antennas (bulk LTSAs) [20], [21]. We use a millimeter wave whose amplitude is modulated by a sine wave of approximately 900 MHz to realize active illumination. The wave scattered by targets is received by aligned antenna elements. Each bulk-LTSA is directly connected to an envelope detector circuit, which detects the envelope of the received signal immediately. Then, the ∼900 MHz signal is homodyne-detected with a local oscillator and an in-phase/quadrature-phase mixer (IQ-mixer). This EPD minimizes the circuit to handle millimeter wave directly, and reduces the cost of circuit design. In addition, we can utilize the phase information corresponding to the modulation wavelength of about 30 cm, which is suitable for detection of human- and bottle-size targets, instead of the carrier wavelength of about 1 cm [17].

We target a pedestrian holding two plastic bottles filled with water by assuming a liquid bomb[18]. Figs. 5 and 6 show an optical image and the obtained complex amplitude images, respectively. The amplitude is normalized to the range from 0 to 1 after taking the logarithm. The horizontal and vertical axes correspond to the observation time and the positions of the parallel antenna elements, respectively. Since our proposed system measures moving objects during the coherent imaging, amplitude fluctuation, and phase distortion often appear owing to interference. The amplitude increases at about the center of the image, corresponding to the time when the target passes. The phase also fluctuates correspondingly. Since a radio wave absorber is set at the background of the measurement area, the amplitude should be small and the phase should be constant in areas other than the objects. However, there are amplitude fluctuations and phase changes even in the background part. Therefore, we attempt to reduce these effects by introducing feature refinement using the complex-valued AE.

IV. Refinement of Features by Complex-Valued AE

In our conventional system, we feed the feature vectors extracted by the above method directly to the CSOM. In our proposed method, we introduce a feature refinement process using an AE to improve the robustness of the system. An AE is a type of multilayer feedforward neural network having an hourglass-type structure. An AE has input and output layers of identical size and a smaller hidden layer, as shown in the schematic diagram in Fig. 7. The hidden layer acquires an expression that effectively represents the features in fewer dimensions by backpropagation learning with input signals as teachers. This process can refine the feature vectors, which are newly obtained in the hidden layer of the AE [22]–[24]. This process is similar to general image denoising [25]–[27]. In this case, we use it to exclude

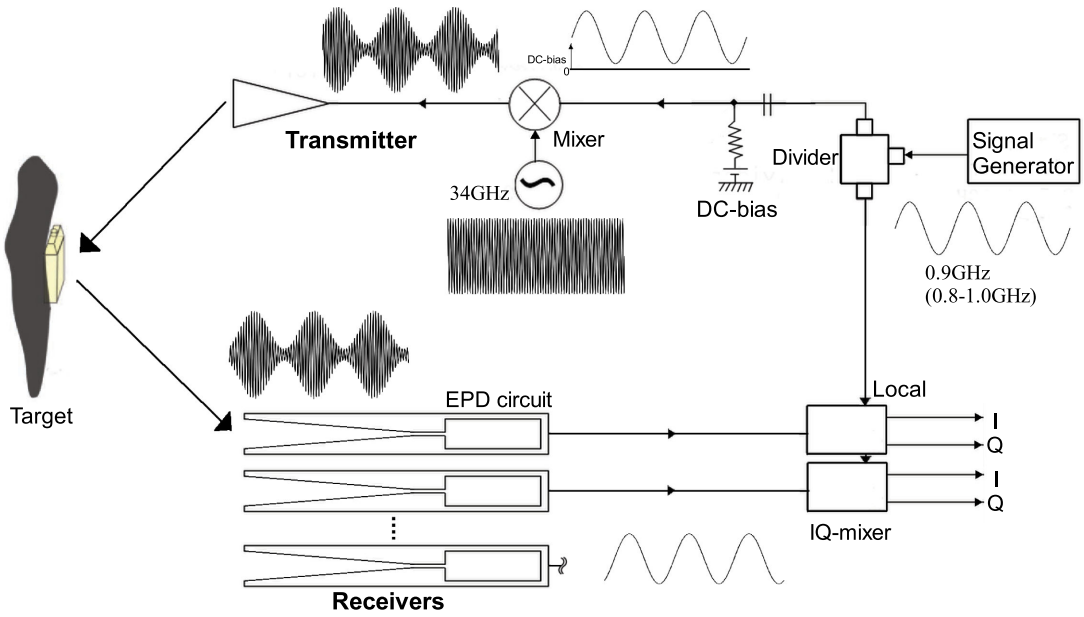


Fig. 3. Schematic diagram of amplitude modulation for active illumination and envelope phase detection (EPD) [17].

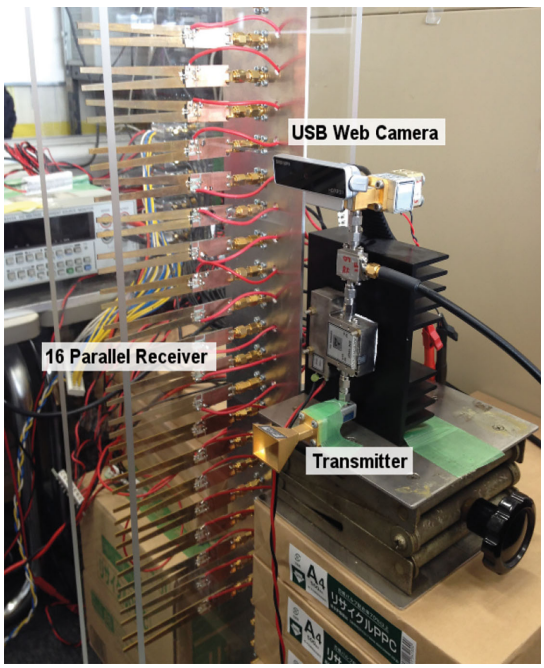


Fig. 4. Photograph of the transmitting unit and the 16 parallel antenna elements of the receiving unit [17].

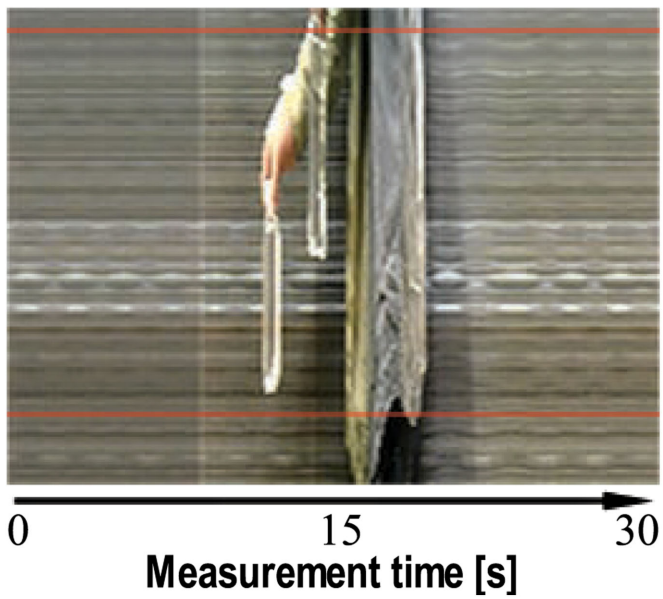


Fig. 5. Optical image of a pedestrian holding two plastic bottles. The two horizontal red lines show the top and bottom heights of the observation area [18].

harmful interference effects such as amplitude fluctuation and phase distortion occurring in coherent active imaging.

Since our system handles complex features, we use a complex-valued AE. The sigmoid function $\tanh u$, commonly used as an activation function in real-valued neural networks, has divergent points in the complex domain. Thus, the sigmoid function $\tanh u$ cannot be used without modification in complex-valued neural networks. Therefore, we use an amplitude-phase-type activation function expressed as

$$f(u) = \tanh(|u|)\exp(i\angle u). \tag{5}$$

This is an operation giving a saturating property only to the absolute value while maintaining the argument of complex numbers. This function does not depend on the direction of the real and imaginary axes, and is suitable for processing information whose rotation about the origin is meaningful. In other words, it is convenient for dealing with the phase and amplitude of a wave as in this case [8].

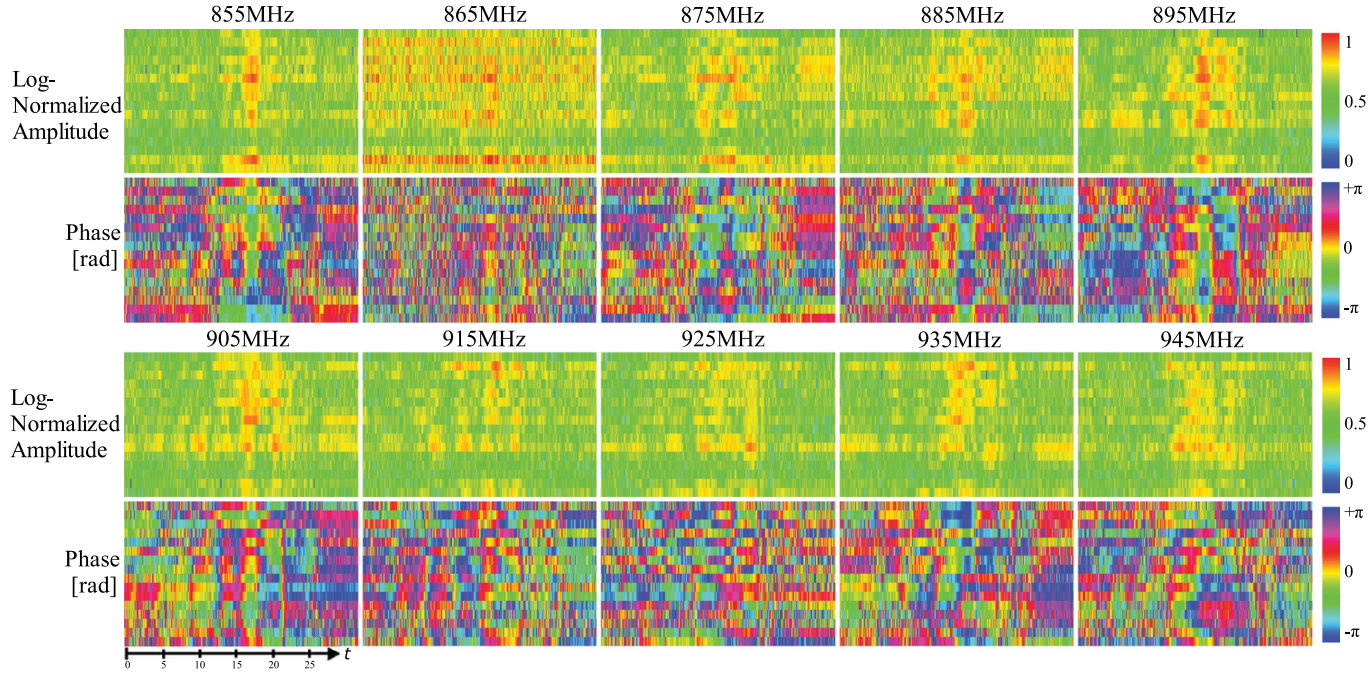


Fig. 6. Captured raw data showing log-normalized amplitude and phase in rad for ten modulation frequency points with 34 GHz carrier wave [18].

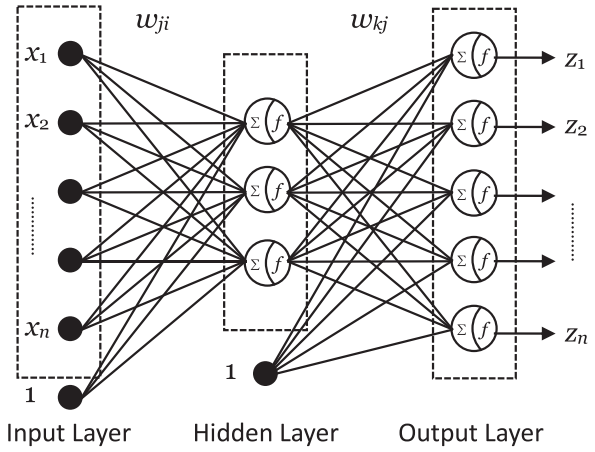


Fig. 7. Schematic diagram of the complex-valued AE [18].

Using the activation function $f(u)$, neural weight matrices $\mathbf{W}_{\text{hidden}}$ and $\mathbf{W}_{\text{output}}$, and neural thresholds $\boldsymbol{w}_{0\,\text{hidden}}$ and $\boldsymbol{w}_{0\,\text{output}}$, we express the output signals in each layer as

$$\boldsymbol{y} = f(\mathbf{W}_{\text{hidden}}\,\boldsymbol{x} + \boldsymbol{w}_{0\,\text{hidden}}) \tag{6}$$

$$\boldsymbol{z} = f(\mathbf{W}_{\text{output}}\,\boldsymbol{y} + \boldsymbol{w}_{0\,\text{output}}). \tag{7}$$

We update the weight matrices $\mathbf{W}_{\text{hidden}}$ and $\mathbf{W}_{\text{output}}$, to bring the output signals $\boldsymbol{z}$ close to the input signals $\boldsymbol{x}$ for the set of feature vectors extracted in Section II, and feed to the CSOM the output signals $\boldsymbol{y}$ in the hidden layer as a set of refined feature vectors instead of the raw $\boldsymbol{x}$.

We describe the learning process of the complex-valued AE below. We denote the teacher signals of each layer as $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$. In the learning process of the AE, the teacher signals of the output layer are input signals, that is, $\hat{\boldsymbol{z}} = \boldsymbol{x}$. On the other hand, the teacher signals of the hidden layer $\hat{\boldsymbol{y}}$ are obtained by backpropagation as

$$\hat{\boldsymbol{y}} = \{f(\hat{\boldsymbol{z}}^*\,\mathbf{W}_{\text{output}})\}^* = f(\mathbf{W}^*_{\text{output}}\,\hat{\boldsymbol{z}}). \tag{8}$$

The components in $\mathbf{W}_{\text{hidden}}$ and $\mathbf{W}_{\text{output}}$, expressed as

$$w_{\text{hidden}\,ji} = |w_{\text{hidden}\,ji}|\exp(\theta_{\text{hidden}\,ji})$$
$$w_{\text{output}\,kj} = |w_{\text{output}\,kj}|\exp(\theta_{\text{output}\,kj})$$

are updated as

$$\begin{aligned}|w_{\text{hidden}\,ji}(t+1)| = |w_{\text{hidden}\,ji}(t)| - \varepsilon\Big\{&(1-|y_j|^2)\\ &\times(|y_j| - |\hat{y_j}|\cos(\theta_j - \hat{\theta}_j))|x_i|\cos\theta^{\text{rot}}_{ji}\\ &-|y_j||\hat{y_j}|\sin(\theta_j-\hat{\theta}_j)\frac{|x_i|}{|u_j|}\sin\theta^{\text{rot}}_{ji}\Big\}\end{aligned} \tag{9}$$

$$\begin{aligned}\theta_{\text{hidden}\,ji}(t+1) = \theta_{\text{hidden}\,ji}(t) - \varepsilon\Big\{&(1-|y_j|^2)\\ &\times(|y_j| - |\hat{y_j}|\cos(\theta_j - \hat{\theta}_j))|x_i|\sin\theta^{\text{rot}}_{ji}\\ &+|y_j||\hat{y_j}|\sin(\theta_j-\hat{\theta}_j)\frac{|x_i|}{|u_j|}\cos\theta^{\text{rot}}_{ji}\Big\}\end{aligned} \tag{10}$$

$$\begin{aligned}|w_{\text{output}\,kj}(t+1)| = |w_{\text{output}\,kj}(t)| - \varepsilon\Big\{&(1-|z_k|^2)\\ &\times(|z_k| - |\hat{z_k}|\cos(\theta_k - \hat{\theta}_k))|y_j|\cos\theta^{\text{rot}}_{kj}\\ &-|z_k||\hat{z_k}|\sin(\theta_k-\hat{\theta}_k)\frac{|y_j|}{|v_k|}\sin\theta^{\text{rot}}_{kj}\Big\}\end{aligned} \tag{11}$$

$$\theta_{\text{output}\,kj}(t+1) = \theta_{\text{output}\,kj}(t) - \varepsilon\Big\{(1-|z_k|^2) \times (|z_k| - |\hat{z_k}|\cos(\theta_k - \hat{\theta_k}))|y_j|\sin\theta_{kj}^{\text{rot}} + |z_k||\hat{z_k}|\sin(\theta_k - \hat{\theta_k})\frac{|y_j|}{|v_k|}\cos\theta_{kj}^{\text{rot}}\Big\} \tag{12}$$

where the components of column vectors $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$, $\boldsymbol{u}$, and $\boldsymbol{v}$ are, respectively, denoted as $|x_i|\exp(\theta_i)$, $|y_j|\exp(\theta_j)$, $|z_k|\exp(\theta_k)$, $|u_j|\exp(\theta_j)$, and $|v_k|\exp(\theta_k)$ in polar coordinates; θ_{ji}^{rot} and θ_{kj}^{rot} are defined as

$$\theta_{ji}^{\text{rot}} \equiv \theta_j - \theta_i - \theta_{\text{hidden}\,ji},\ \theta_{kj}^{\text{rot}} \equiv \theta_k - \theta_j - \theta_{\text{output}\,kj} \tag{13}$$

and ε is the learning coefficient. The column vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are weighted sums in each layer and are expressed as

$$\boldsymbol{u} = \mathbf{W}_{\text{hidden}} \cdot \boldsymbol{x},\ \boldsymbol{v} = \mathbf{W}_{\text{output}} \cdot \boldsymbol{y}.$$

V. Learning Process of the Complex-Valued AE and Refined Feature Vectors

To configure the AE, there are some parameters to be set such as the number of hidden layers and the number of neurons in each hidden layer. The dimension of the converted feature vector after feature refinement, in other words, the input dimension of the CSOM in the next stage, is determined by the number of hidden-layer neurons. If the number of hidden-layer neurons is too large, the features of the input vector are completely reproduced, which means that the harmful amplitude fluctuations and phase distortions are not removed. In contrast, when the number of hidden-layer neurons is too small, the refined features are insufficient to reproduce the features meaning that they cannot be classified by the CSOM. To investigate appropriate parameters experimentally, we constructed some AEs with different parameters.

First, to investigate the effect of the number of hidden layers, we prepared AEs with one and three-hidden layers and compared them. However, there was almost no difference in the results. Thus, we examined the distribution of the output values of all layers and found that most of the output values were in the linear region of the activation function used for the AEs. Therefore, even if there are multiple hidden layers, it is possible to linearly combine the weights of the hidden layers, and there is almost no difference in the results from those obtained using a single-hidden layer. We decided to use an AE with a single-hidden layer in consideration of calculation time.

In the experiment, we use a set of 14-D feature vectors extracted from the data obtained at ten modulation frequency points by the method explained in Section II. The window size is $T \times L = 7 \times 4$. We set the size of the hidden layer to 2–15. We use the mean square error (MSE) of output signals from teachers as an index of the degree of reproduction of features in the AEs.

Fig. 8 shows the MSE of the AEs versus learning iteration. As the iteration proceeds, the MSE decreases, but stops at a certain value, showing a floor. The number of iterations at which the reduction of MSE stops depends on the size of the hidden layer, but the reduction stops sufficiently at approximately 100 iterations with any size. Then, in the subsequent experiments, the number of learning iterations is fixed at 100. We found in a series of trials that the learning is not so sensitive to the learning coefficient ε. Then, we also fix the learning coefficient ε at 0.01.

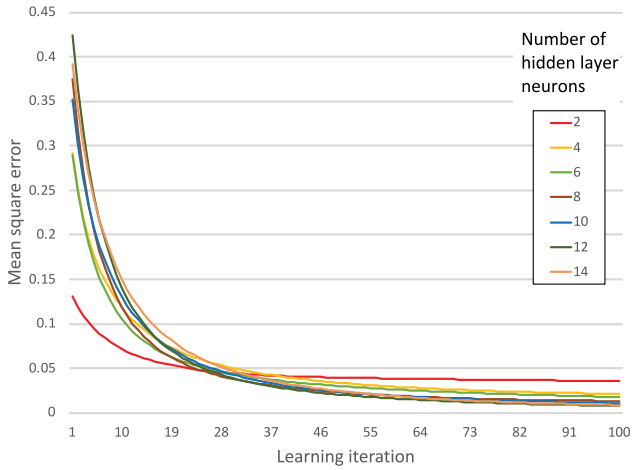


Fig. 8. Mean square error versus learning iterations.

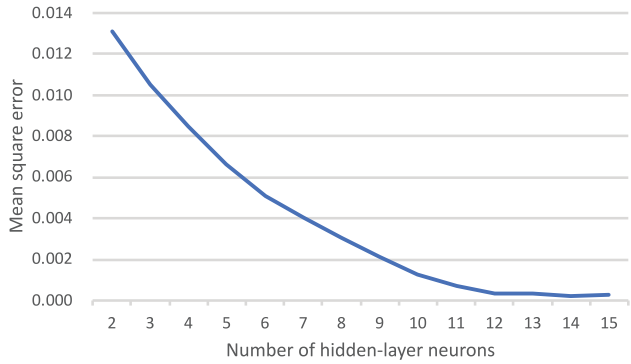
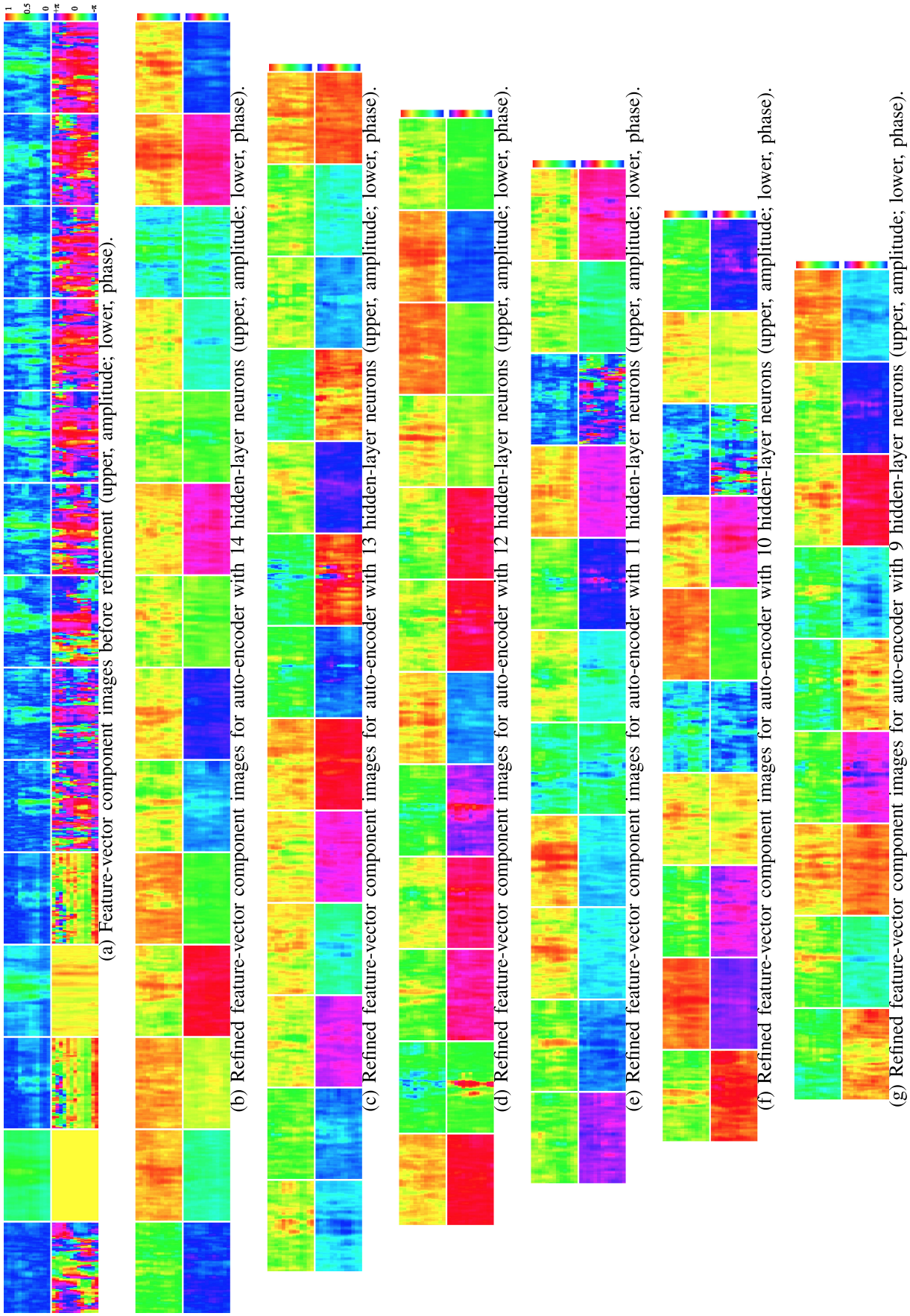


Fig. 9. Mean square error versus number of hidden-layer neurons in the AEs (ε is 0.01 and number of learning iterations is 100).

Fig. 9 shows the MSE of the AEs with different hidden-layer sizes. The horizontal axis shows the number of hidden-layer neurons and the vertical axis shows the MSE after 100 learning iterations for the dataset. The MSEs are the average of ten AEs for each hidden-layer size. As the hidden-layer size increases, the MSE decreases, but the gradient becomes gentler toward the right of the curve. A lower MSE is not necessarily desirable since the reproduced signal includes noise that should be eliminated. We predict that a hidden-layer size of approximately 12, where the MSE stops decreasing, should be appropriate.

Figs. 10 and 11 show the feature-vector component images before and after refinement using the complex-valued AE. Each image of the feature-vector components is shown as pairs of the amplitude (magnitude) in the upper row and the phase (argument) in the lower row. Each image corresponds to a single scan in Fig. 5. In the feature vector before refinement shown in Fig. 10(a), we find large phase fluctuations at places in the background part. In contrast, in the feature vector after refinement, we find small fluctuations. Amplitude and phase changes at the object position are emphasized in some images



(a) Feature-vector component images before refinement (upper, amplitude; lower, phase).

(b) Refined feature-vector component images for auto-encoder with 14 hidden-layer neurons (upper, amplitude; lower, phase).

(c) Refined feature-vector component images for auto-encoder with 13 hidden-layer neurons (upper, amplitude; lower, phase).

(d) Refined feature-vector component images for auto-encoder with 12 hidden-layer neurons (upper, amplitude; lower, phase).

(e) Refined feature-vector component images for auto-encoder with 11 hidden-layer neurons (upper, amplitude; lower, phase).

(f) Refined feature-vector component images for auto-encoder with 10 hidden-layer neurons (upper, amplitude; lower, phase).

(g) Refined feature-vector component images for auto-encoder with 9 hidden-layer neurons (upper, amplitude; lower, phase).

Fig. 10. Feature vectors before and after refinement by the complex-valued AE (1/2).

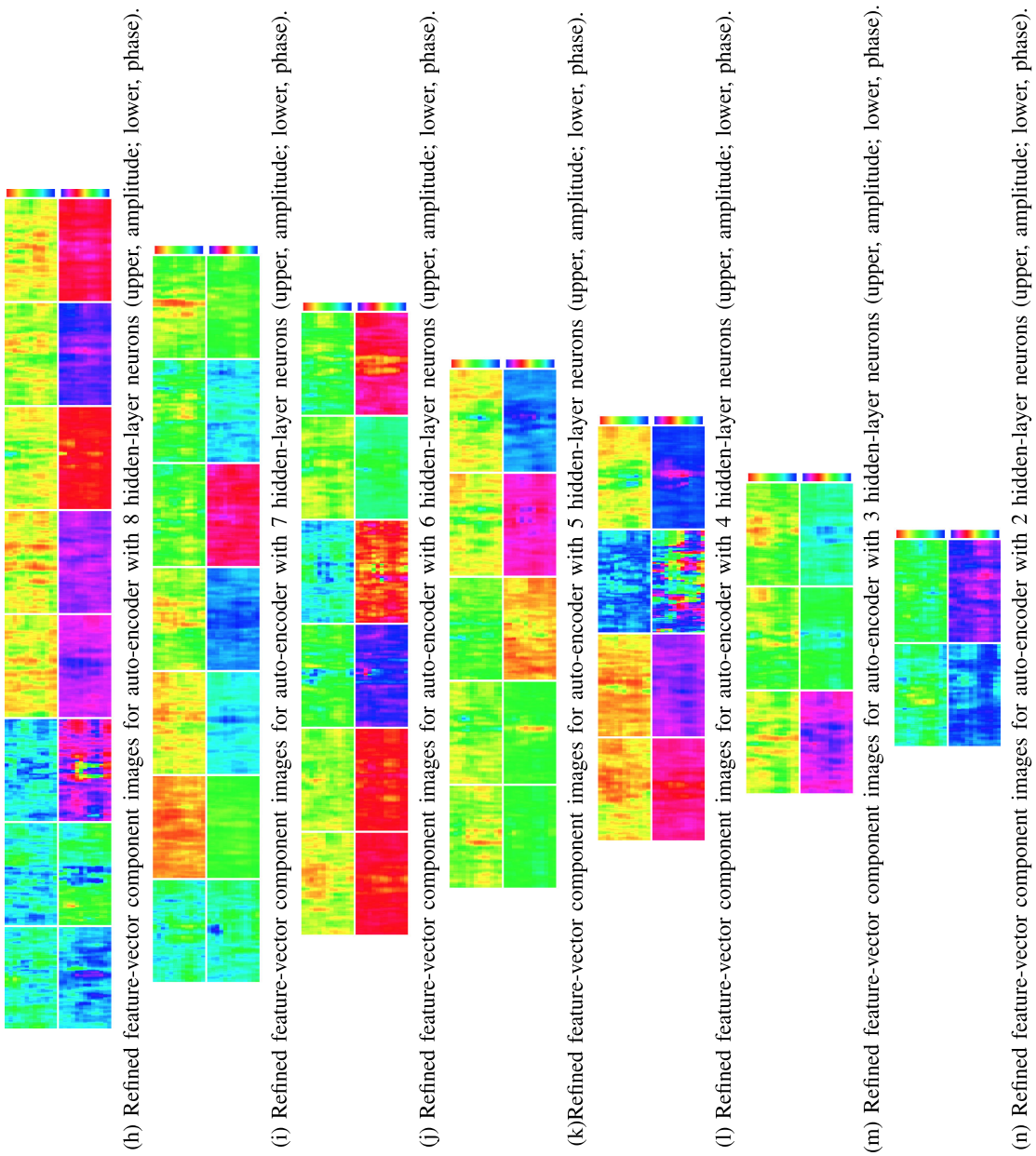


Fig. 11. Feature vectors before and after refinement by the complex-valued AE (2/2).

such as the second component image of Fig. 10(d). In addition, in the feature vectors before refinement shown in Fig. 10(a), the component images have greatly different tendencies from one another because their values are essentially different (mean, spatial correlations, and interfrequency correlations) as described in Section II. This may lead to inappropriate initial values and deviation in updates to reduce the adverse effect on learning accuracy in the CSOM. On the other hand, in the refined feature vectors (b)–(n), there is no such concern because the components of the refined feature vectors are mixed and uniform in their nature.

VI. Classification Results With the CSOM

We fed and classified the above feature vectors to the CSOM. Figs. 12 and 13 show (a) a sketch of the total measurement area, (b) a classification result without feature refinement, and (c)–(o) the results with refinement using AEs with various hidden-layer sizes. In the result without feature refinement shown in Fig. 12(b), we find that some areas without objects, especially in the first-half of the measurement (ahead of the objects), are classified into multiple classes as artifacts. On the other hand, areas where the objects exist are classified into only a few classes, and we cannot distinguish the human body and the plastic bottles from one another. It is considered that the artifacts in the first-half are generated by the influence of phase distortions, which are found in the obtained complex amplitude data (see Fig. 6) and the extracted feature vectors [see Fig. 10(a)]. In the results shown in Fig. 13(l) and (n) for the case of classification with feature refinement, the artifacts seen in Fig. 12(b) are so reduced that we can distinguish the background area and the areas in which objects exist. In addition, in the results shown in Fig. 12(c),

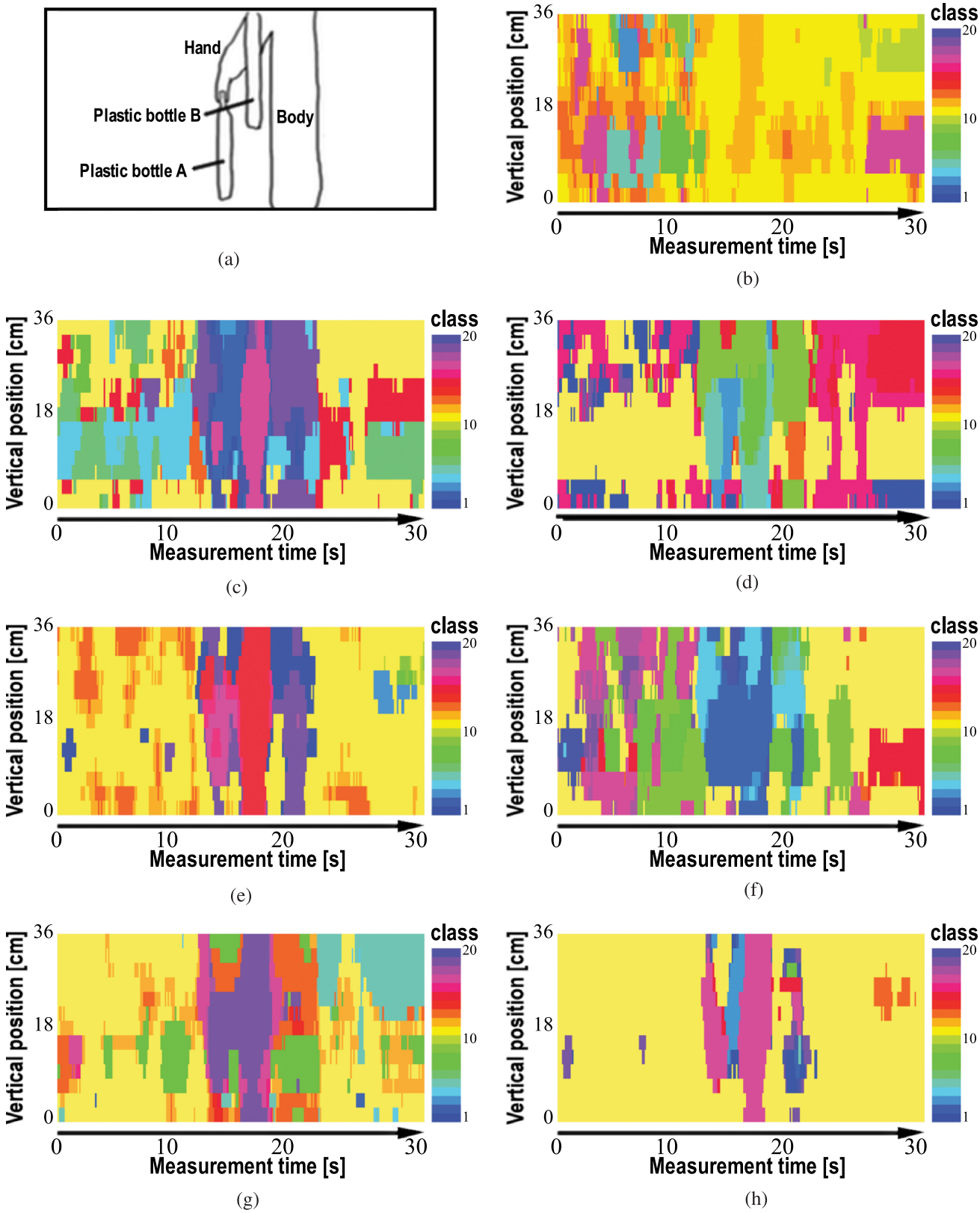
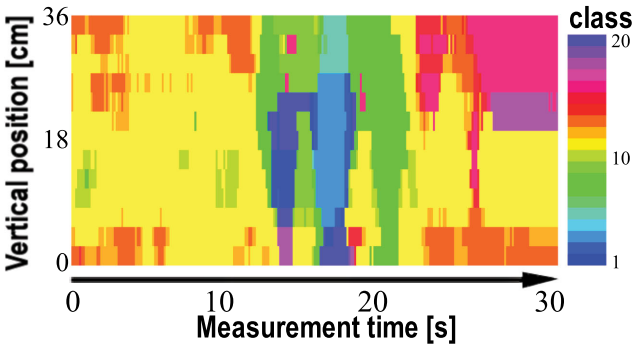
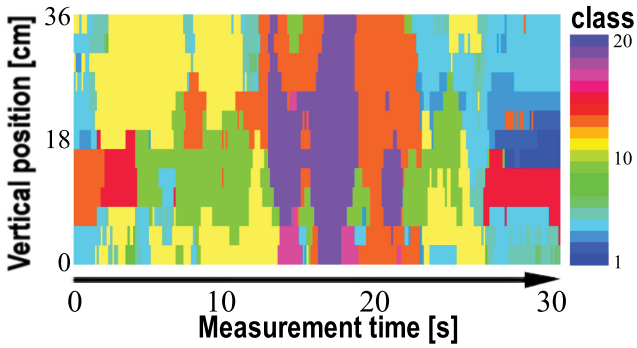


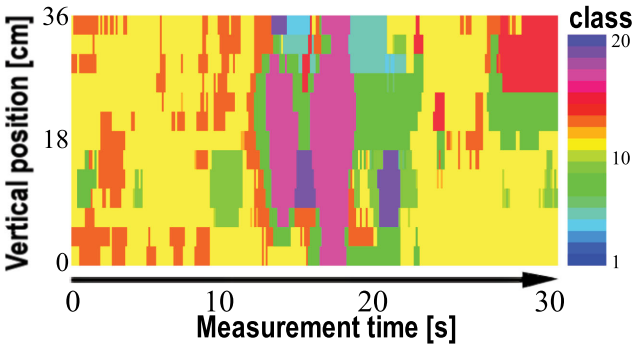
Fig. 12. (a) Sketch of the target area and visualization results obtained using the CSOM (b) without and (c)–(o) the feature refinement using the complex-valued AE (1/2). (a) Sketch of the target area (b) Visualization result without feature refinement. (c) Visualization result with 14-hidden-layer-neuron AE. (d) Visualization result with 13-hidden-layer-neuron AE. (e) Visualization result with 12-hidden-layer-neuron AE. (f) Visualization result with 11-hidden-layer-neuron AE. (g) Visualization result with 10-hidden-layer-neuron AE. (h) Visualization result with 9-hidden-layer-neuron AE.



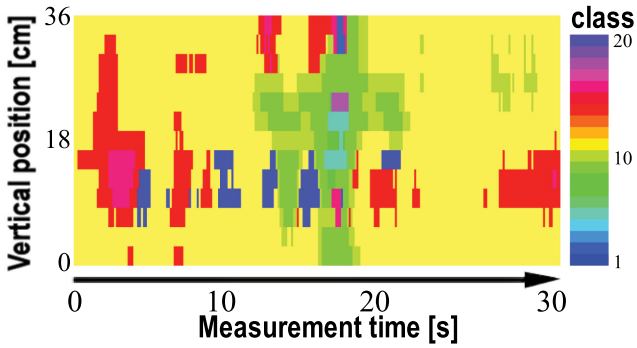
(i) Visualization result with 8-hidden-layer-neuron AE.



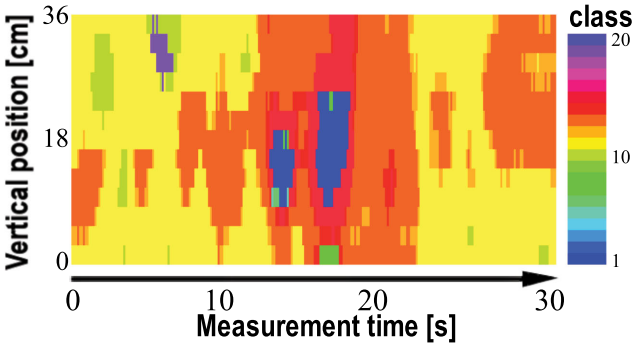
(j) Visualization result with 7-hidden-layer-neuron AE.



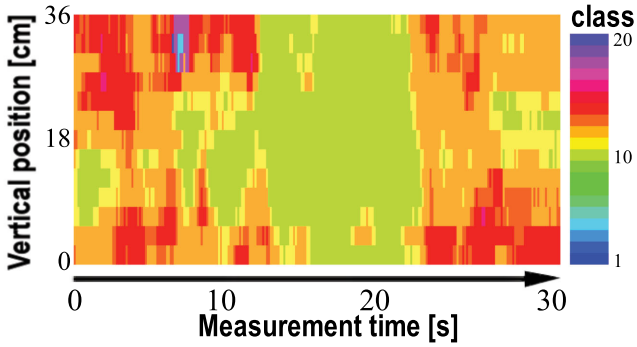
(k) Visualization result with 6-hidden-layer-neuron AE.



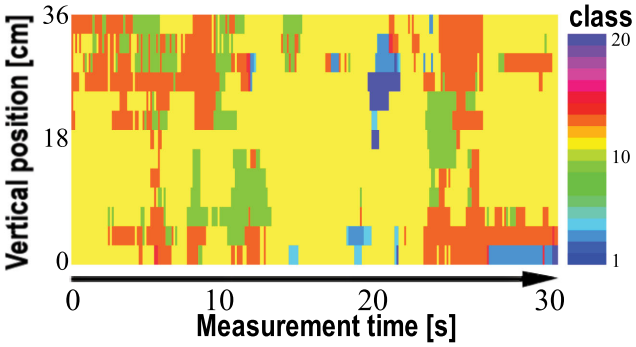
(l) Visualization result with 5-hidden-layer-neuron AE.



(m) Visualization result with 4-hidden-layer-neuron AE.



(n) Visualization result with 3-hidden-layer-neuron AE.



(o) Visualization result with 2-hidden-layer-neuron AE.

Fig. 13. (a) Sketch of the target area and visualization results obtained using the CSOM (b) without and (c)–(o) the feature refinement using the complex-valued AE (2/2). (i) Visualization result with 8-hidden-layer-neuron AE. (j) Visualization result with 7-hidden-layer-neuron AE. (k) Visualization result with 6-hidden-layer-neuron AE. (l) Visualization result with 5-hidden-layer-neuron AE. (m) Visualization result with 4-hidden-layer-neuron AE. (n) Visualization result with 3-hidden-layer-neuron AE. (o) Visualization result with 2-hidden-layer-neuron AE.

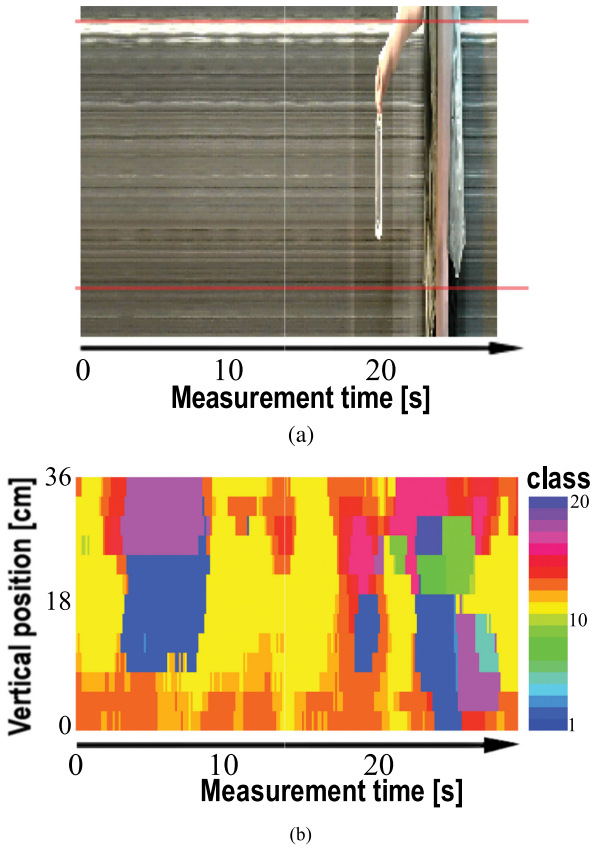


Fig. 14. Visualization result of another scene with 6-hidden-layer-neuron complex-valued AE.

(e), (g), (h), (i), (j), (k), and (m), we can distinguish target plastic bottles from the pedestrian because they are classified into different classes, and we can conclude that the visualization accuracy is improved. However, even with feature refinement, there are also results where the visualization accuracy was not improved, for example, the results shown in Fig. 13(o).

In the results with small number of hidden-layer neurons, as shown in Fig. 13(m)–(o), a tendency is seen that, as the number of hidden-layer neurons increases, the objects are classified in greater detail and distinguished as the human body and plastic bottles. This is because, if the number of hidden-layer neurons is too small, the refined feature vectors do not express sufficient feature components, meaning that the objects cannot be classified using the CSOM. In contrast, when the number of hidden-layer neurons is larger than five, there seems to be no correlation between the visualization accuracy and the number of hidden-layer neurons. This does not match the prediction that, if the number of hidden-layer neurons is too large, the dimension reduction has very little effect on reducing the noise that the visualization accuracy does not improve. The reason is that, when the AEs have sufficient hidden-layer neurons, some neurons become linear dependent on other neurons, and the number of effective dimensions decreases to within an appropriate range as shown below.

Fig. 14 shows the result of another data different from the above using the 6-hidden-layer-neuron AE. The measurement conditions are detailed in [17]. In this classification result, while a large false image is generated in front of the traveling direction, it is found that the position of the human body and the plastic bottle are classified.

Figs. 12, 13 and 14 show relatively good results in several trials for each number of hidden-layer neurons. Even if the input data and parameters were same, the learning results of the AEs and the classification results could differ from one another. Fig. 15 shows the classification results with feature refinement by multiple 7-hidden-layer-neuron AEs learned under the same conditions except the initial values of the weight matrices. The used data are the same as that of Figs. 12 and 13. Depending on the initial values, there are some results in which the feature refinement works advantageously to suppress the artifacts as shown in, for example, Fig. 15(b) and (e), and in which we distinguish the plastic bottles from the human body as shown in Fig. 15(c), i.e., the body in blue and the plastic bottles in red. However, there are other results whose visualization accuracy does not improve, such as those shown in Fig. 15(g) and (h). Since the implemented AE and CSOM learn by a deterministic method, the difference in the results is due to these difference in the randomly set initial value of the weight vector. In other words, there is nonnegligible possibility to fall into local solutions depending on the initial value. It is necessary to consider introducing a method of setting the initial values or a learning algorithm that does not easily fall into local solutions. In the next section, we examined the possibility that the hidden layer neurons would not work effectively because the weight vectors of multiple hidden layer neurons become linear dependent, which may be one of the causes of the AE falling into a local solution.

VII. Independence Among the Hidden-Layer Neurons

We investigate the degeneracy of the hidden-layer neurons of the AE. We define the weighting matrix $\mathbf{W}_{\text{hidden}}$ of the hidden layer in the trained AE as

$$\mathbf{W}_{\text{hidden}} \equiv \begin{bmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_n \end{bmatrix} \equiv \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1(m+1)} \\ w_{21} & w_{22} & \cdots & w_{2(m+1)} \\ \vdots & & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{n(m+1)} \end{bmatrix} \tag{14}$$

where m and n are the numbers of terminals in the input layer and neurons in the hidden layer, respectively. Here, when the input vector is $\boldsymbol{x}$, the output of a hidden-layer neuron y_j $(1 \leq j \leq n)$ is expressed as

$$y_j = f(\boldsymbol{w}_j \boldsymbol{x}) = f\left(\sum_{i=1}^{m} w_{ji} x_i + w_{j(m+1)} \cdot 1 \right). \tag{15}$$

To investigate the degree of independence of the hidden-layer neurons, we define the correlation d_{jk} $(1 \leq j \leq n,\ 1 \leq k \leq n,\ j \neq k)$ between weight vectors $\boldsymbol{w}_j$ and $\boldsymbol{w}_k$ in $\mathbf{W}_{\text{hidden}}$ as

$$d_{jk} \equiv \frac{|\boldsymbol{w}_j \boldsymbol{w}_k^*|}{\|\boldsymbol{w}_j\| \|\boldsymbol{w}_k\|} = \frac{1}{\|\boldsymbol{w}_j\| \|\boldsymbol{w}_k\|} \left| \sum_{i=1}^{m+1} w_{ji} w_{ki}^* \right|. \tag{16}$$

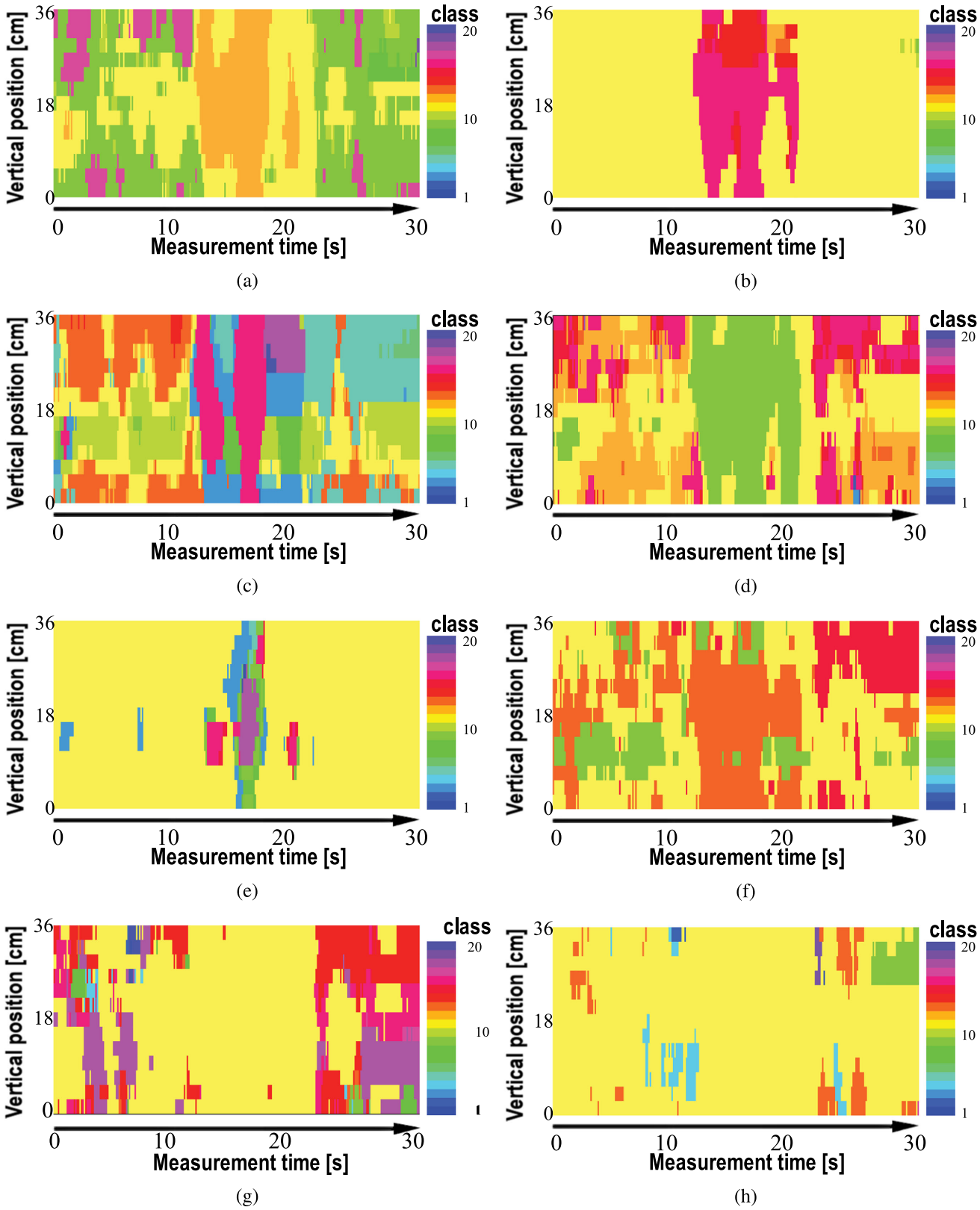


Fig. 15. Various visualization results to show the initial-value dependence for 7-hidden-layer-neuron complex-valued AE.

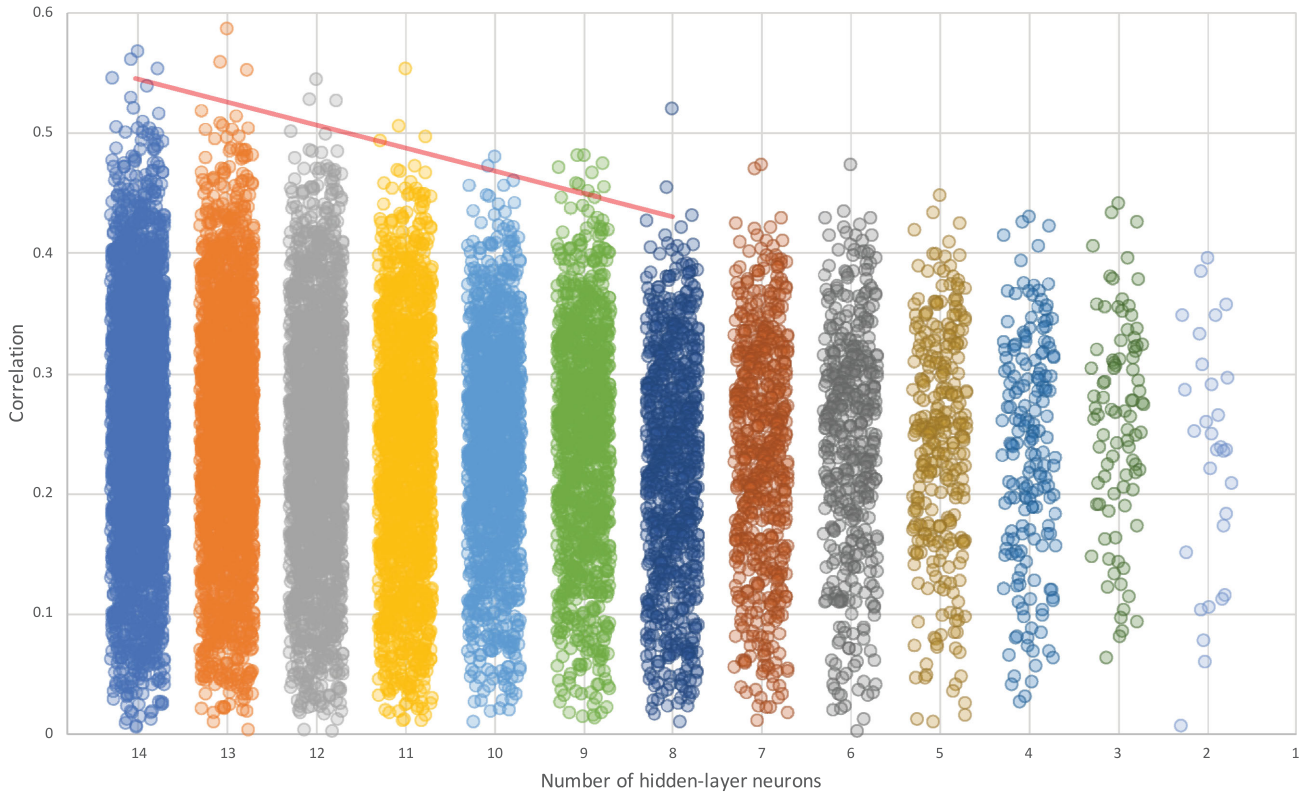


Fig. 16. Correlations d_{jk} of every pair of weight vectors in $\mathbf{W}_{\text{hidden}}$.

Here, d_{jk} satisfies $0 \leq d_{jk} \leq 1$. A smaller value of d_{jk} means greater independence, whereas $d_{jk} = 1$ means that the hidden-layer neurons represented by $\boldsymbol{w}_j$ and $\boldsymbol{w}_k$ are completely degenerate.

Fig. 16 shows a graph plotting d_{jk} of all combinations of i and j for every AE, we generated (30 AEs for every hidden-neuron number n). We plot $\frac{n\times(n-1)}{2}$ points for every AE versus n. When the number of the hidden-layer neurons n is in the range from 2 to 8, the upper ends of the plots are almost constant. In contrast, when it is from 8 to 14, the upper ends of the plot tend to rise as the number of hidden-layer neurons increases. That is, where the number of hidden-layer neurons is larger than a certain value, the correlation between hidden neurons becomes stronger as the number of hidden-layer neurons increases. This result suggests that the learning dynamics leads not to the extraction of random noise but to the representation of meaningful features even if the weights become similar to some extent. This matches the visualization result in Fig. 12 that there is no correlation between the number of hidden-layer neurons and the visualization accuracy when the number of hidden-layer neurons is more than a certain value.

The values of d_{jk} are fairly widely distributed, as shown in Fig. 16, since the degree of independence is not taken into account in the learning process in the implementation of the complex-valued AEs explained in Section IV. This should also lead to a large effect of the initial values on the resulting independence. A regularization component to increase the independence of the hidden-layer neurons may work effectively in the learning process of the AEs so that they are less effect by randomly chosen initial values.

VIII. Conclusion

We proposed a moving-target visualization technique classifying complex textures by using CSOM in coherent imaging. We developed a prototype millimeter-wave radar system and verified the effectiveness of this visualization technique. We showed that the refinement of features by using a complex-valued AE introduced in the feature extraction process improves the robustness and discrimination performance of the CSOM. Since coherent imaging is susceptible to the harmful effects of interference in measurement, it often suffers from amplitude fluctuation and phase distortion. We showed that the AE is extremely useful in extracting features appropriately even under such adverse influences.

References

[1] D. Sheen, D. McMakin, and T. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 9, pp. 1581–1592, Sep. 2001.

[2] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Dec. 2017.

[3] D. Gleich and D. Sipos, "Complex valued convolutional neural network for TerraSAR-X patch categorization," in *Proc. 12th Eur. Conf. Synthetic Aperture Radar*, VDE, 2018, pp. 1–4.
[4] A. Hirose, S. Tsuda, and R. Natsuaki, "Structurization of synthetic aperture radar information by using neural networks," in *Proc. SAR Big Data Era, Models, Methods Appl.*, Nov. 2017, pp. 1–4.
[5] Y. Sunaga, R. Natsuaki, and A. Hirose, "Proposal of complex-valued convolutional neural networks for similar land-shape discovery in interferometric synthetic aperture radar," in *Proc. Int. Conf. Neural Inf. Process.*, 2018, pp. 340–349.
[6] Y. Sunaga, R. Natsuaki, and A. Hirose, "Land form classification and similar land-shape discovery by using complex-valued convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 24–38, Oct. 2019.
[7] A. Hirose and S. Yoshida, "Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 541–551, Apr. 2012.
[8] A. Hirose, *Complex-Valued Neural Networks.* 2nd ed., Berlin, Germany: Springer, 2012.
[9] T. Hara and A. Hirose, "Plastic mine detecting radar system using complex-valued self-organizing map that deals with multiple-frequency interferometric images," *Neural Netw.*, vol. 17, no. 8–9, pp. 1201–1210, Nov. 2004.
[10] S. Onojima, Y. Arima, and A. Hirose, "Millimeter-wave security imaging using complex-valued self-organizing map for visualization of moving targets," *Neurocomputing*, vol. 134, pp. 247–253, 2014.
[11] K. Mizuno *et al.*, "New applications of millimeter-wave incoherent imaging," in *Proc. Microw. Symp. Dig., IEEE MTT-S Int.*, 2005, pp. 4–pp.
[12] K. Watanabe, M. Yoneyama, and K. Mizuno, "Millimeter-wave active imaging using neural networks for signal processing," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 5, pp. 1512–1516, May 2003.
[13] A. Chalmers, "Three applications of backscatter x-ray imaging technology to homeland defense," *Proc. SPIE*, vol. 5778, 2005, pp. 989–993.
[14] S. Hantscher *et al.*, "Security pre-screening of moving persons using a rotating multichannel w-band radar," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 870–880, Mar. 2012.
[15] K. Saho, T. Sakamoto, and T. Sato, "Imaging of pedestrians with UWB doppler radar interferometry," in *Proc. Int. Symp. Electromagn. Theory*, 2013, pp. 29–32.
[16] Y. Arima and A. Hirose, "Millimeter-wave active imaging using multiple frequencies with complex-valued self-organizing map image processing," in *Proc. Asia-Pacific Microw. Conf.*, 2014, pp. 1031–1033.
[17] Y. Arima and A. Hirose, "Performance dependence on system parameters in millimeter-wave active imaging based on complex-valued neural networks to classify complex texture," *IEEE Access*, vol. 5, pp. 22 927–22 939, 2017.
[18] Y. Arima and A. Hirose, "Improvement of texture clustering performance in complex-valued SOM by using complex-valued auto-encoder for millimeter-wave coherent imaging," in *Proc. Int. Conf. Neural Inf. Process.*, 2017, pp. 722–730.
[19] S. Masuyama, K. Yasuda, and A. Hirose, "Multiple mode selection of walled-LTSA array elements for high resolution imaging to visualize antipersonnel plastic landmines," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 745–749, Oct. 2008.
[20] Y. Nakano and A. Hirose, "Taper-walled linearly tapered slot antenna," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 4, pp. 779–784, Dec. 2011.
[21] D. Radenamad, T. Aoyagi, and A. Hirose, "Low impedance bulk LTSA," *Electron. Lett.*, vol. 46, pp. 882–883, 2010.
[22] G. W. Cottrell, P. Munro, and D. Zipser, "Image compression by back propagation: An example of extensional propagation," *ICS Report*, vol. 8702, 1987, Art. no. 321.
[23] G. W. Cottrell and P. Munro, "Principal components analysis of images via back propagation," *Proc. SPIE*, vol. 1001, pp. 1070–1077, 1988.
[24] D. DeMers and G. W. Cottrell, "Non-linear dimensionality reduction," *Adv. Neural Inform. Process. Syst.*, vol. 5, pp. 580–587, 1993.
[25] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. Dec, pp. 3371–3408, 2010.
[26] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 833–840.
[27] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Adv. Neural Inform. Process. Syst.*, 2012, pp. 341–349.

Yuya Arima received the Ph.D. degree in electrical engineering and information systems from the University of Tokyo, Tokyo, Japan, in 2018.

He is currently studying at the National Institute of Information and Communications Technology. His research interests include neural networks and remote sensing.

Akira Hirose (Fellow, IEEE) received the Ph.D. degree in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1991.

In 1987, he joined Research Center for Advanced Science and Technology (RCAST), the University of Tokyo, as a Research Associate. In 1991, he was appointed as an Instructor with RCAST. From 1993 to 1995, on leave of absence from the University of Tokyo, he joined the Institute for Neuroinformatics, University of Bonn, Bonn, Germany. He is currently a Professor with the Department of Electrical Engineering and Information Systems, the University of Tokyo. In the fields, he has authored or coauthored several books such as *Complex-Valued Neural Networks*, 2nd Edition (Springer, 2012). His research interests include wireless electronics and neural networks.

Dr. Hirose served previously as the Founding President of Asia-Pacific Neural Network Society (APNNS) (2016), President of Japanese Neural Network Society (JNNS) (2013–2015), Vice President of the IEICE Electronics Society (ES) (2013–2015), Editor-in-Chief of the IEICE Transactions on Electronics (2011–2012), Associate Editor of journals such as the IEEE TRANSACTIONS ON NEURAL NETWORKS (2009–2011), IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2009–2012), Chair of the Neurocomputing Technical Group in the IEICE, and General Chair of International Conference on Neural Information Processing (ICONIP) 2016 Kyoto and Asia-Pacific Conference on Synthetic Aperture Radar (APSAR) 2013 in Tsukuba. He currently serves as a member of IEEE Computational Intelligence Society (CIS) Neural Networks Technical Committee (NNTC) (2009–present), Founding Chair of the NNTC Complex-Valued Neural Network Task Force (2010–present), Governing Board Member of APNNA/APNNS (2006-), IEEE GRSS Tokyo Chapter Chair (2013–2015), and General Chair of International Geoscience and Remote Sensing Symposium (IGARSS) 2019 Yokohama. He Hirose is a Fellow of the IEICE, and a member of JNNS and APNNS.