IEEE Open Journal of the
Industrial Electronics Society

Received 20 March 2023; revised 15 May 2023; accepted 5 June 2023. Date of publication 8 June 2023;
date of current version 22 June 2023. The review of this paper was arranged by Associate Editor M. Manic.

Digital Object Identifier 10.1109/OJIES.2023.3284014

A Deep Metric Learning-Based Anomaly Detection System for Transparent Objects Using Polarized-Image Fusion

ATSUTAKE KOSUGE [1] (Member, IEEE), LIXING YU[1], MOTOTSUGU HAMADA [1] (Member, IEEE), KAZUKI MATSUO [2], AND TADAHIRO KURODA [1] (Fellow, IEEE)

[1]Graduate School of Engineering, The University of Tokyo, Tokyo 1138656, Japan.
[2]ExFusion Inc., Osaka 541004, Japan.

CORRESPONDING AUTHOR: ATSUTAKE KOSUGE (e-mail: kosuge@dlab.t.u-tokyo.ac.jp)

This work was supported by the New Energy and Industrial Technology Development Organization under Project JPNP20004.

ABSTRACT While visual inspection systems have been widely used in many industries, their use in the food and optical equipment industries has been limited. Transparent and reflective materials are often used in these applications, but existing anomaly detection (AD) systems have low accuracy in their detection due to low visibility. Here, we developed an AD system using a polarization camera for reflective and transparent target objects. Two new techniques are developed. First is the polarized image fusion (PIF) technique which suppresses glare from reflective surfaces while highlighting transparent foreign objects. In PIF, four captured polarized images are fused to synthesize a high-quality image according to calculated weight coefficients. The second new technique is an ArcObj-based deep metric learning technique to improve AD accuracy. The proposed system was evaluated in experiments on three datasets: cookie samples wrapped in transparent plastic bags; transparent plastic bottles; and transparent lenses. High AD accuracies in terms of the area under the receiver operating characteristic curve (AUC) were achieved: 0.88 AUC for the cookie dataset; 0.87 AUC for the bottle dataset; and 0.98 AUC for the lens dataset. Compared to the state-of-the-art AD algorithm (Patchcore), the proposed method improved AD accuracy by 0.09 AUC.

INDEX TERMS Neural networks, polarized image sensor, reflection, sensor fusion, visual inspection.

I. INTRODUCTION

Artificial intelligence (AI) technology has been widely studied to automate various manual tasks. The introduction of AI technology is expected to improve safety, such as in advanced driver assistance systems (ADAS) [1], [2], and factory productivity, such as in factory automation [3].

A task in high demand for automation in factories is visual inspection. There has been research on building an automated visual inspection system [4] capable of inspecting all sorts of production items. The advantage of such a total inspection system is that the overall quality of a factory's production can be fully guaranteed by using it to inspect all products and removing defective ones. A large amount of human resources is typically required for total inspection; a fully automated visual inspection system can significantly reduce the total inspection cost. Here, deep neural network based anomaly detection (AD) systems have been actively studied and have achieved high detection accuracy comparable to that of human experts. The state-of-the-art methods using convolutional neural networks (CNN) [5], [6] have achieved a high performance index (area under receiver operating characteristic curve (AUC) of 0.99) on large industrial product datasets (MVTec dataset [7]).

While there is demand for such visual inspection technologies, their use in the food and optical equipment industries is difficult. The major technical challenges of applying conventional visual inspection systems to such industries are (1) reflective target objects that are common in the food industry, such as transparent plastic wrapping bags [see Fig. 1(a)], and (2) transparent target objects that are common in both the food

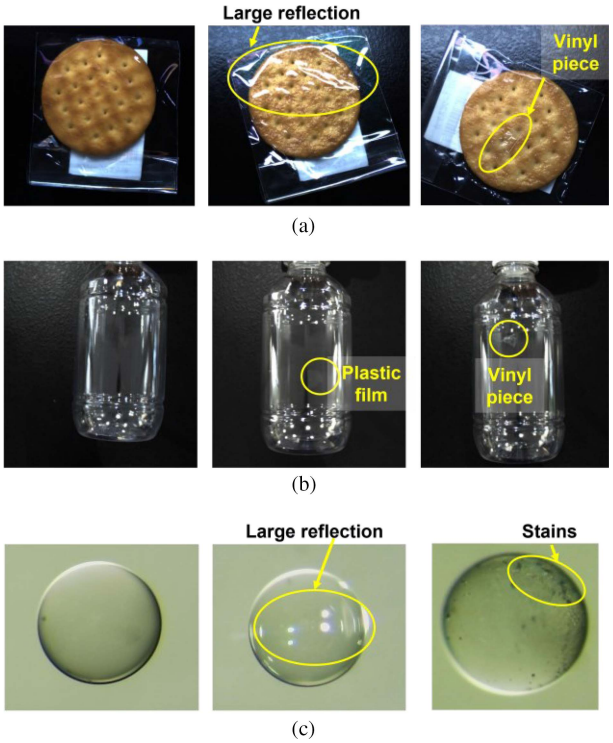


FIGURE 1. Transparent and highly reflective materials commonly used in the food and optical industries have poor visibility. (a) Cookies wrapped in transparent plastic bags. (b) Plastic bottles with and transparent anomalous objects. (c) Transparent lenses with small stains.

and optical industries, such as objects made of plastic and glass [see Fig. 1(b) and (c)]. Such reflective materials cause large glares in captured images, which may be recognized as anomalies and greatly increase the false-negative rate [10].

An emerging application in the optical equipment industry is high-power lasers [8], [9]. In these applications, tiny transparent lenses and transparent laser targets (which are called shells) about 1 mm in diameter need to be inspected. Since the components are tiny, transparent objects, they have few features, such as textures or patterns that can be used for judging between normal and abnormal conditions. Because of reflections, even slight deviations in the camera setup or lighting fixture location can significantly change the appearance of the object. Such transparent objects and reflections are known to degrade the accuracy of conventional AD systems [10].

To realize automated visual inspection for the food and optical industries, we have developed a visual inspection system based on a polarized-image fusion (PIF) technique (a preliminary report was presented in [10]) with the following techniques.

1) PIF technique removes glare from captured images by combining images from four different polarizer angles. In addition, by applying a brightness normalizing filter, transparent foreign objects are highlighted to improve the detection accuracy.
2) ArcObj-based deep metric learning technique further improves AD accuracy. While unsupervised learning techniques are often applied to visual inspection systems because they do not require data collection and data labeling, the lack of data labels makes it difficult to improve accuracy. We have developed a method to improve the accuracy by applying deep metric learning. Since the feature vectors of reflective and transparent materials have complex values, the simple L2 norm-based deep metric learning [27] degrades AD accuracy (see Section IV-C, Table 3). In this article, an ArcObj-based deep metric learning method is developed. Since feature vectors are distributed in angular space, the accuracy can be improved by applying angular margin penalty-based loss function during the training process.

By the combination of the two techniques, AD accuracy is improved. Compared with Patchcore [6], which has the highest accuracy on the standard MVTec dataset used in benchmarking AD, the accuracy is improved by 0.09 points with the reflective object dataset of cookies wrapped in transparent plastic bags [see Fig. 1(a)].

II. RELATED WORK

A. ANOMALY DETECTION ALGORITHM

Many unsupervised CNN techniques [5], [6], [12], [13] have been actively studied for AD systems. In such unsupervised CNN techniques, feature extraction is performed using CNNs that are pretrained with large datasets, such as ImageNet. AD is performed by measuring the distance between the distribution of features of normal products obtained during training and those obtained during inference. If the distance is shorter than the predetermined threshold, the target object is recognized as a normal product. A well-known example of this technique is deep one-class classification (DoC) [13]. To improve the accuracy, the CNN is trained using two datasets (normal and unrelated datasets) to localize the feature distribution of the normal product. More advanced methods are student-teacher feature pyramid matching (STFPM) [5] and Patchcore [6]. Patchcore uses feature vectors output from midlayers of CNN. Combined with a nearest neighbor method, Patchcore has achieved the highest accuracy on the MVTec dataset as of 2023 [14].

On the other hand, these techniques target images without reflections. When applied to AD in transparent or reflective objects, the accuracy deteriorates as described in Section IV.

B. REFLECTION AND GLARE REDUCTION

Reflection reduction techniques have been widely studied to improve image quality [15], [16], [17], [18], [19]. In these methods, the transmitted and reflected components are separated, then reflections are removed by using CNN. However, most existing reflection removal methods require strict assumptions about reflections, such as that reflections occur on flat, smooth glass surfaces only [15], [16], [17]. These techniques degrade reflection removal capabilities when applied to scenes with non-uniform reflective surfaces or diffused

reflections, such as in food packaging. For example, since many techniques assume reflection images are out of focus, they may not remove reflection properly when the reflection is in focus therefore sharp and strong [17].

To enhance recognition capability under reflections and glare, dataset synthesis method has been proposed to improve object detection capability with strong reflections (e.g., toilets) in indoor scenes and output their location in the image [18]. In addition, a new dataset was proposed for sign detection in outdoor scenes with a lot of glares [19]. To recognize transparent objects, a method was developed that combines images obtained from multiple viewpoints, detects the surface, and estimates the object outline using CNN [20]. However, since lost information due to reflections and glare was reproduced by CNN supposing that there are no defects, it is not able to detect defects that exist near reflection and glare areas.

C. CONTRIBUTION OF THIS WORK

In this article, an AD system is developed for reflective and transparent objects that have nonuniform surfaces. Food packages, bottles, and lens surfaces are curved and have a wide variety of reflections. Therefore, reflection removal techniques using fixed polarization filters cannot be used to remove the reflections on nonuniform surfaces. In this article, PIF technology, which reduces reflections by synthesizing four polarization images, is developed to reduce reflections on nonuniform surfaces.

III. VISUAL INSPECTION SYSTEM USING POLARIZATION CAMERA

A. POLARIZATION CAMERA

A key component to suppress the reflection is a polarization camera. When specular reflection occurs on the surface of a transparent material, the reflected light is partially or fully polarized. This kind of light can be reduced by using a polarizer. According to Malus's law (1), a polarizer absorbs polarized light depending on the angle α between the vibration direction of the light and the polarization direction of the polarizer, where I_0 is the intensity of the incident light, and I is the intensity of the transmitted light

$$I = I_0 \cos^2 \alpha. \tag{1}$$

Polarizers completely absorb light when the angle α is 90°. Therefore, if a polarizer is put at a certain angle, it will absorb specular reflections in a chosen direction and allow light polarized in other directions to pass through and thereby reduce the effect of glare on an image. (Light in other directions is diffuse light that is less absorbed.)

A recent polarization camera [21] integrates multiple polarizers into the surface of the image sensor [11]. In this way, four different polarization images can be obtained in a single shot by simultaneously integrating polarizers with different directions of 0°, 45°, 90°, and 135°. The polarizers are simply formed by using the metal wires of the image sensor, so the additional cost is negligibly small compared with that of conventional image sensors (see Fig. 2). In fact, the camera [21] is comparable in cost to traditional RGB cameras.

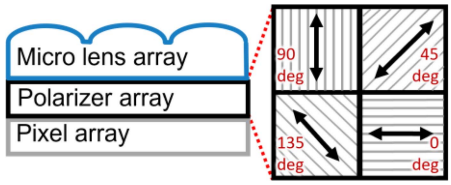


FIGURE 2. Polarizers are formed on the image sensor surface by using metal wires [11].

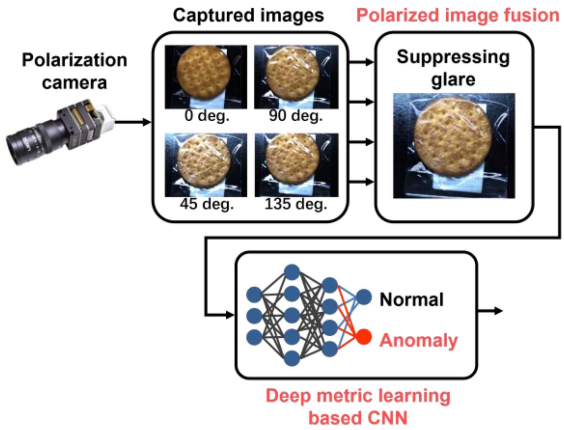


FIGURE 3. Proposed PIF-based AD system applied to the cookie dataset (see Fig. 1).

B. POLARIZED-IMAGE FUSION (PIF) BASED ANOMALY DETECTION ALGORITHM

The proposed PIF-based AD algorithm is shown in Fig. 3. It synthesizes high-quality images from four polarized images to reduce the reflections from transparent surfaces and improves the visibility of transparent foreign objects. Furthermore, the ArcObj technique, which is our application of the Arcface deep metric learning technique used for face recognition [22], is used to improve the accuracy of AD. The details of each technique are described below.

A technical issue in suppressing reflection is that a fixed polarizer angle cannot suppress reflections well under all conditions. This is because the optimal polarizer angle depends on the position and shape of the object and the angle of the background light. The best angle may be 0° for some cases but 90° for others depending on the position and orientation of the object. However, the polarization angle is usually difficult to adjust once the camera is set up on the production line in a factory.

The PIF technique can mitigate this problem. The technique synthesizes a high-quality image that has reduced reflection from four polarized input images. Its AUC is improved compared with that of a method where one best image is selected from among four polarized images.

The key idea is similar to the one behind high dynamic range (HDR) techniques that are used by cameras with limited

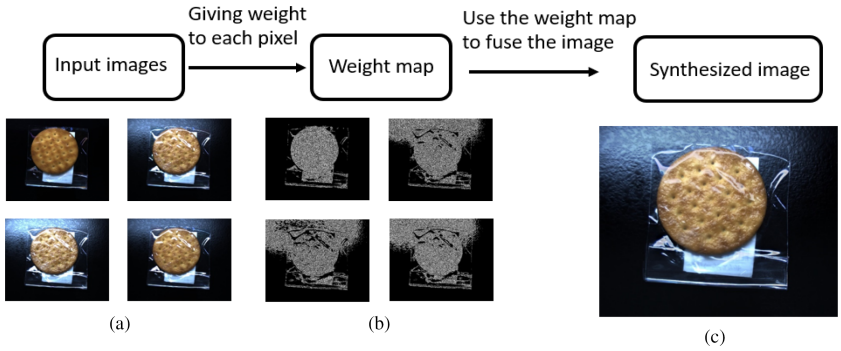


FIGURE 4. PIF technique: fusing four images captured at different polarizer angles to eliminate glare.

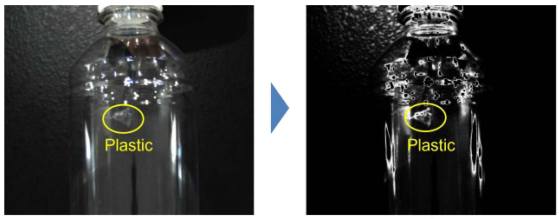


FIGURE 5. Transparent foreign objects are highlighted by applying the brightness normalization filter *B*. (a) Input image. (b) Brightness normalization filter *B*.

TABLE 1. Pseudocode for Polarized-Image Fusion Algorithm

Algorithm 1 Polarized Image Fusion
Input: 4 polarized images ($\boldsymbol{I}(i, j, deg)$) (deg = 0, 45, 90, 135)
Output: 1 image $\boldsymbol{O}$(i, j)
for all (i, j) do
$\boldsymbol{L}(i, j, deg) = Laplacian(\boldsymbol{I}(i, j, deg))$
$\boldsymbol{C}(i, j, deg) = Median(\boldsymbol{L}(i, j, deg))$
$\boldsymbol{S}(i, j, deg) = Std(\boldsymbol{I}(i, j, deg))$
$\boldsymbol{B}(i, j, deg) = \prod Gauss\ Curve(I_{deg\_CH}(i, j, deg))$ (CH = Red, Green, Blue)
$\boldsymbol{W}(i, j, deg) = \boldsymbol{C}(i, j, deg) \times \boldsymbol{S}(i, j, deg) \times \boldsymbol{B}(i, j, deg)$ **(Eq. (2))**
$\boldsymbol{W_N}(i, j, deg) = \boldsymbol{W}(i, j, deg) / \sum_{k=0}^{3} \boldsymbol{W}(i, j, k \times 45)$ **(Eq. (3))**
$\boldsymbol{O}(i, j) = \sum_{k=0}^{3} \boldsymbol{W_N}(i, j, k \times 45) \times \boldsymbol{I}(i, j, k \times 45)$
Return $\boldsymbol{O}(i, j)$

dynamic ranges compared with that of the human eye to create high-quality images. In HDR, to cover the dynamic range of the eye, multiple images with different exposure times are captured. By fusing them on the basis of pixel intensity and physical camera information, such as exposure time, saturated pixels (both blackout and whiteout) can be removed and a high-quality image synthesized. Different fusion techniques have been proposed, but most of them use exposure time as the parameter for fusion [23].

PIF is similarly designed to remove glare caused by reflection through the fusion of multiple images. To cover multiple reflection angles from the target object surface, four images are captured with a polarization camera using polarizers at four different angles 45° apart. Reflections can be suppressed by fusing the images on the basis of their pixel intensity values. Unlike HDR, the polarization camera captures the four images with different polarization angles simultaneously, so all four have the same exposure time. This means that most image fusion algorithms for HDR cannot be applied to PIF as they fuse the images on the basis of exposure time.

To mitigate this problem, a weight map-based fusion algorithm is developed that does not use exposure time. Corresponding scalar-valued weight maps that represent the important elements of each image are calculated and used for fusion (see Fig. 4). This weight map-based fusion algorithm not only suppresses reflections, but also highlights transparent foreign objects (see Fig. 5). Therefore, it improves the accuracy of detecting transparent, abnormal objects. Pseudocode for the proposed PIF technology is given in Table 1.

Each weight map is calculated from three values: contrast C; saturation S; and brightness B. C is the output after applying a Laplacian filter [23] to the grayscale input image from each angle. A median filter [25] is also applied to reduce noise. C tends to have a high value for important elements such as edges and texture. S is defined by the standard deviation [26] within the RGB channel to express the saturation of each pixel. Since images captured at different polarizer angles have different levels of brightness due to the different amounts of attenuation of reflected light, a Gaussian curve [25] is applied to each image for normalization. Data from each RGB channel is normalized by a Gaussian curve and then multiplied together. (Note that in [23], normalization is by the exposure time instead of the brightness B.) Then, the weight map W is calculated as:

$$W_{ij,k} = C_{ij,k} \times S_{ij,k} \times B_{ij,k} \tag{2}$$

where the subscripts i, j, k identify the (i, j) pixel in the kth image. In each grayscale weight map, the intensity of each pixel indicates its quality.

Once all the weight maps W are computed, they are normalized as follows.

$$W_{N_{ij,k}} = W_{ij,k} / \sum_{k'=1}^{4} W_{ij,k'}. \tag{3}$$

The normalized weight maps W_{N} are used to fuse the input images I, where the output image O is calculated as follows:

$$O_{ij} = \sum_{k=1}^{4} W_{N_{ij,k}} I_{ij,k}. \tag{4}$$

Fig. 4(a) shows photos of the same object at different polarizer angles, Fig. 4(b) shows the corresponding weight maps, and Fig. 4(c) shows the synthesized image. Note that the upper left image in Fig. 4(a) is an input image at a polarizer angle of 0°. While it has a small amount of glare, strong reflections occur in the other three images (taken at a different object angle or with the plastic bag deformed into a different shape). Most reflections are suppressed in the synthesized image, which is clear and high quality.

In the brightness normalization, the brightness B has a high value when the pixel intensity is close to the average intensity value of the image, and a low value when the intensity is higher or lower than the average owing to the Gaussian curve. For example, strongly reflected areas, such as white areas and black backgrounds, have values close to zero. Conversely, pixels with intermediate brightness, such as those showing the surface of a plastic bottle, have high values. Notably, transparent objects, such as the piece of vinyl inside the plastic bottle shown in Fig. 5(b), also have high values. As a result, such transparent foreign objects are highlighted by this brightness normalization operation, which contributes to a more accurate AD.

C. ARCOBJ BASED DEEP METRIC LEARNING

Unsupervised deep learning techniques are used by AD systems to reduce the cost of building a dataset that has good coverage. There have been many studies to improve accuracy. One sort is deep metric learning, which increases the distance between samples of the same object class (intraclass). It has been studied for face recognition applications where it is difficult to know in advance how many people need to be distinguished. Feature vectors that are easily distinguishable are required for such face identification tasks.

On the other hand, as discussed in the experimental results section (see Section IV), the use of a simple deep metric learning such as with the L2 norm will degrade the AUC more than the use of a conventional CNN-based method [13]. This is because the feature vectors output from the feature extractor will not be precisely distinguished in Euclidean space. This issue has led to many distance calculation methods having been studied in the field of face recognition. A prominent method is Sphere face [22] in which the output of a feature extractor for face recognition trained with the Softmax loss has a unique angular distribution. This suggests that it is more useful to use the angular distance for the distance calculation than the Euclidean distance.

In this article, we developed the ArcObj method from the ArcFace method for face recognition [24] as a way to improve the accuracy of AD for transparent object targets (see Fig. 6). In ArcFace, the face recognition feature extractor is pretrained with the Softmax loss and outputs a unique distribution in the angle space [22]. We found that the usefulness of this unique angular distribution is not limited to face recognition; it can also be applied to industrial products. In particular, when the feature extractor is trained using the Softmax loss on a large dataset containing industrial products, such as lenses and

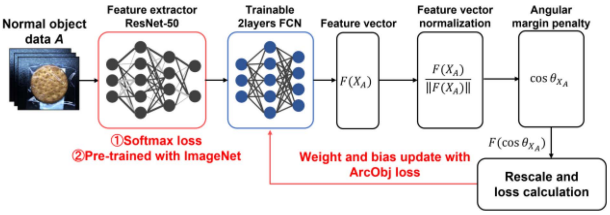


FIGURE 6. ArcObj-based unsupervised deep metric learning where the weights are updated using the angular margin-based loss function.

bottles, the ArcFace method can be applied to visual inspection of products, such as cookies and bottles in the food industry and lenses in the optical industry. Here, the feature vector $F(X)$ is converted to $\cos\theta_X$, and training is performed using the loss function calculated with the angular distance $F_L(X)$. Note that X is the input data, $F(X)$ is the feature extractor function, and $F_L(X)$ is the angular margin penalty function defined as follows [24]:

$$F_L(X) = s * \cos(\theta_X + m). \tag{5}$$

As can be seen from the above equation, the feature vector is mapped onto the angle space. The parameter s is a scaling factor and m is a margin penalty factor.

The network configuration used in this article is based on DoC [13]. Resnet-50 was used for the feature extractor and pre-trained with ImageNet dataset and the SoftMax loss. Two-trainable layers, fully connected layers were added to extract feature vectors. However, while the original DoC updates the weights of trainable layers with a cross-entropy loss function, this article updates the weights with ArcObj loss function as follows:

$$\text{Loss} = \frac{1}{2}\sum_{i=0}^{1}\frac{\exp(s * \cos(\theta_X + m))}{\exp(s * \cos(\theta_X + m)) + \sum_{k\neq X} s * \cos(\theta_k)} \tag{6}$$

where s is set to 30 and m is set to 0.5. The task of this article is binary classification (normal or abnormal). Therefore, the number of classes is set to 2. The implementation is mostly the same as that of the ArcFace work described in [24].

IV. EXPERIMENTAL RESULTS

A. SYSTEM SETUP

The proposed visual inspection system using PIF was constructed as shown in Fig. 7 using a polarized light camera [22]. Cookie samples, plastic bottles, and lenses were photographed to construct the datasets. Both the cookie and the plastic bottle are about 50 mm in diameter and were photographed with a 16 mm fixed-focus lens set at a height of 95 mm. The camera is connected to Nvidia's Jetson Xavier via the Power over Ethernet port. The Xavier is used as the host and controls the timing of the shots. There are two operating modes: data collection and inference. In the data collection mode, the obtained data is transferred to the server after the required number of images are obtained and used to train the proposed

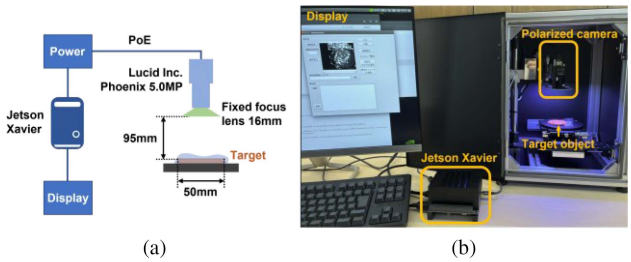
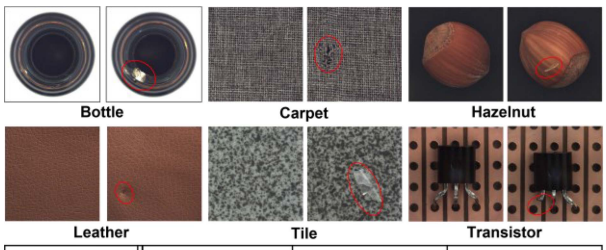
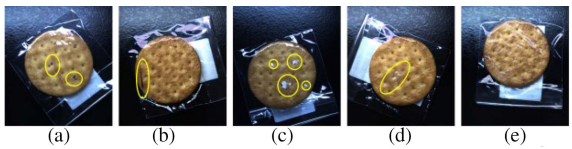


FIGURE 7. Experimental setup using polarized camera and Jetson Xavier for the real time validation of the AD system. (a) Experimental system setup. (b) Photograph of experimental setup.



Method	Patchcore-1 [6]	STFPM [5]	This work DoC+ArcObj+PIF
Mean average of all classes	0.99	0.87	0.85
Bottle	1.00	0.99	0.99
Carpet	0.98	0.96	0.89
Hazelnut	1.00	0.99	0.94
Leather	1.00	0.98	0.98
Tile	0.99	0.97	0.96
Transistor	1.00	0.87	0.84

FIGURE 8. Experimental results on dataset I "industrial standard dataset MVTec [7]" showing that our algorithm has competitive accuracy to the latest algorithm STFPM [5].



Method	Input		Anomaly pattern (a)	(b)	(c)	(d)	Average
STFPM [5]	Single RGB	AUC	0.76	0.70	0.77	0.84	0.77
		F1 score	0.75	0.80	0.76	0.87	0.79
Patchcore-1 [6]	Single RGB	AUC	0.73	0.94	0.74	0.80	0.79
		F1 score	0.75	0.92	0.76	0.89	0.83
Ours	PIF	AUC	0.91	0.87	0.84	0.89	0.88
		F1 score	0.90	0.91	0.90	0.91	0.90

FIGURE 9. Experimental results on dataset II "cookie wrapped in a transparent plastic bag" showing that our algorithm has good accuracy for reflective target objects. (a) Black stains. (b) Black foreign objects. (c) White stains. (d) Transparent vinyl piece. (e) Normal object.

Method	Input		Anomaly pattern (a)	(b)	(c)	Average
STFPM [5]	Single RGB	AUC	0.98	0.71	0.82	0.80
		F1 score	0.94	0.54	0.65	0.71
Patchcore-1 [6]	Single RGB	AUC	0.98	0.70	0.87	0.84
		F1 score	0.88	0.53	0.75	0.72
This work	PIF	AUC	0.98	0.79	0.93	0.89
		F1 score	0.91	0.92	0.92	0.91

FIGURE 10. Experimental results on dataset III "transparent plastic bottles" showing that our algorithm has good accuracy for transparent target objects. (a) Colored film. (b) Transparent vinyl piece. (c) Transparent thin film. (d) Normal object.

network structure. After training, the network is implemented on Jetson Xavier and the system moves into the inference mode. In the inference mode, inference is performed on all captured images using the trained model.

B. DATASETS AND PERFORMANCE SUMMARY

Experiments were conducted on three datasets. One dataset is an industrial RGB standard dataset (MVTec) that is the benchmark dataset [7] to measure the AD accuracy [14]. The other two datasets consist of images of transparent and reflective objects on which conventional RGB camera-based AD CNNs would have low accuracy. The first set was of a cookie in a transparent plastic bag, and the second one was of a transparent plastic bottle. In the cookie dataset, there were 2000 normal samples (1600 for training and 400 for testing) and 200 abnormal samples (50 for each anomaly type). In the plastic bottle dataset, there were 4000 normal samples (3200 for training and 800 for testing) and 400 abnormal samples (100 for each anomaly type). The input image size is resized to 224 × 224.

Fig. 8 shows the evaluation results on the MVTec dataset. We compared the results with two state-of-the-art AD methods: Patchcore [6], which has the highest AD accuracy on this dataset as of 2023 [14], and STFPM [5]. Our method is based on DoC [13] combined with the ArcObj method and has competitive accuracy to STFPM [5].

While Patchcore [6] and STFPM have good accuracy with the normal dataset (MVTec), the accuracy of AD in reflective and transparent objects is degraded (see Figs. 9 and 10).

Fig. 9 shows some examples of the cookie dataset. The cookie was wrapped in a transparent plastic bag, so there were many specular reflections. Noise was added to the surface of some samples to represent the abnormal class. There were four kinds of anomaly: small black stains simulating mold and small foreign objects;long black acrylic plates simulating elongated foreign objects; small white stains simulating mold and small foreign objects; and transparent objects. The cookie was ϕ 55 mm in size. The bag was 60 mm × 80 mm. This dataset was for assessing the ability to detect anomalies against a strong reflective surface. The task was to distinguish between samples with anomalies (in the yellow circled areas) and the normal ones.

Fig. 10 shows samples from the second dataset. As in the first set, there were three abnormal classes: a 5-mm-square piece of color film; a 5-mm-square piece of transparent vinyl; and a 5-mm-square piece of transparent thin film. These foreign objects were placed in plastic bottles and were not fixed in position. They could freely move about inside the bottles. The bottle had a cylindrical curved surface, for which it is

TABLE 2. Ablation Study Results Showing That PIF Improved the AD Accuracy by 0.04 AUC With the Cookie Dataset

Polarized images used for evaluation	CNN	Anomaly pattern of cookie dataset				Average
		(a)	(b)	(c)	(d)	
Single, 0 deg	DoC [13]	0.84	0.84	0.81	0.77	0.81
Single, 45 deg		0.78	0.83	0.83	0.75	0.79
Single, 90 deg		0.76	0.79	0.79	0.69	0.76
Single, 135 deg		0.77	0.77	0.76	0.70	0.75
4 polarizers, fusion (PIF)		0.83	0.89	0.83	0.86	0.85

TABLE 3. Ablation Study Results Showing That ArcObj Improved the AD Accuracy by 0.03 AUC With the Cookie Dataset

Polarized images	CNN and learning strategy	Anomaly pattern of cookie dataset				Average
		(a)	(b)	(c)	(d)	
Four images fusion (PIF)	DoC [13]	0.83	0.89	0.83	0.86	0.85
	Metric learning L2-norm [27]	0.76	0.80	0.79	0.84	0.80
	This work: Metric learning ArcObj	0.91	0.87	0.84	0.89	0.88

difficult to completely eliminate specular reflection through polarizers. The plastic bottle was about 150 mm × 60 mm.

As shown in Fig. 9, our method achieved a higher accuracy score than the state-of-the-art methods for all anomaly patterns in the cookie dataset. The AUC was improved by 0.09 points on average compared with Patchcore [6]. Similarly, its AUC improvement was as high as 0.05 in the case of the plastic bottle dataset (see Fig. 10).

C. ABLATION STUDY

Ablation studies were conducted on the PIF technique and ArcObj-based deep metric learning technique to evaluate their effectiveness. The effectiveness of the image fusion was evaluated by comparing two types of one-class classification, that is, training and testing with only single-angle polarization images without using image fusion, and training and testing with polarization images that were fused using the PIF technique on the cookie sample dataset.

The test results of the ablation study for PIF are given in Table 2. The transparent bag wrapping the cookie caused strong specular reflection, so the accuracy varied depending on the polarizer angle. The proposed PIF technique greatly improved the accuracy from 0.81 to 0.85 AUC compared with an AD using single-angle polarization images.

Table 3 gives the results of the ablation study of the deep metric learning. It compares DoC (which is an unsupervised CNN), the simple deep metric-learning method with the L2-norm distance [27], and our ArcObj based deep metric learning method. DoC had high discrimination capability and achieved the best AUC on some anomaly patterns. However, the ArcObj-based method achieved the highest AUC on most anomaly patterns; it had an average improvement of 0.03 AUC over DoC. Moreover, the results show that simply applying deep metric learning does not contribute to an AUC improvement; the L2-norm deep metric learning degraded the AUC compared with that of DoC. For visual inspection applications, it is better to use the angular distance as in face identification tasks. The above results are for the first dataset only, but similar trends were found for the second and third datasets.

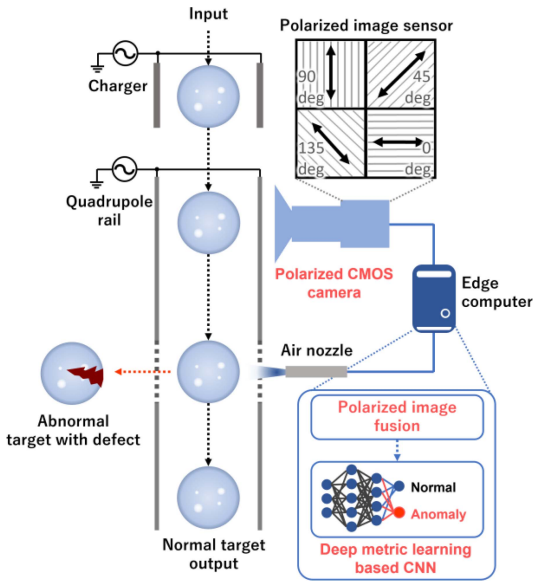


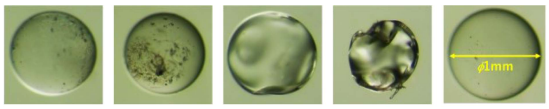
FIGURE 11. Conceptual sketch of a visual inspection system for tiny lenses for optical industry applications using a quadrupole.

V. FUTURE APPLICATION

The technology we propose is useful not only in the food industry but also in the optical equipment industry, such as in high-power lasers inspection [8], [9]. In these applications, tiny transparent lenses and transparent laser targets (which are called shells) about 1 mm in diameter need to be inspected. In particular, the optical components need to be fed through the inspection system at a high throughput of 10 Hz or more.

Fig. 11 shows an example of a visual inspection system for the optical industry application. The proposed AD algorithm is implemented in an edge computer, which is integrated into the production line in the factory. For the industrial production line application, the camera is used to take photos of products that are automatically transported on a conveyor belt. Abnormal products are removed from the conveyor belt by a robot on the basis of processing results of the photos by the AD system. In optical industrial applications such as high-power lasers, mechanical transport methods, such as conveyor belts are not suitable for handling tiny lenses of 1 mm in diameter. As shown in Fig. 11, a quadrupole is used to electrically transport the lenses in a noncontact manner [10]. The lenses are given static electricity by the charger and moved by electrostatic induction at a constant speed along the quadrupole rail, as if they were on a conveyor belt. However, since there is no physical contact during the movement, there is no risk of scratches or dust adhering to the lenses. An air jet ejects abnormal products from the rail.

Our AD system can detect anomaly patterns of the lens. Fig. 12 shows samples from the lens dataset we collected. There were four abnormal classes: a small black stain; large



(a) (b) (c) (d) (e)

Method		Anomaly pattern				Average
		(a)	(b)	(c)	(d)	
This work	AUC	0.97	0.98	0.99	0.99	0.98
	F1 score	0.98	0.99	1.0	1.0	0.99

FIGURE 12. Experimental results on dataset IV "tiny lenses" showing that our system can be applied to optical industry applications. (a) Small stain. (b) Large stain. (c) Small abnormal shape. (d) Large abnormal shape. (e) Normal object.

black stain; small abnormal shape; and large abnormal shape. The object in this dataset was very small, only 1 mm [see Fig. 12(e)]. Therefore, only this dataset was photographed in a static environment using a zoom lens. The background was changed from being black to transparent to simulate a quadrupole rail (see Fig. 11). The background color was greenish due to the lighting setup of the environment. Our proposed AD system achieved a high AUC for all patterns. The processing of our AD system has a fast enough inference throughput such as 10 fps or more with the edge computer (Jetson Xavier). This technology has sufficient AD capability and throughput (10 fps) to meet the high throughput requirements of high-power laser applications (10 Hz).

VI. CONCLUSION

An automatic visual inspection system was proposed for the food and optical equipment industries to alleviate the labour shortage problem. While transparent and reflective materials are often used in these applications, existing AD systems have low accuracy in their detection due to low visibility. To mitigate this problem, two techniques are developed in this article. The first is PIF which suppresses glare from reflective surfaces and transparent foreign objects, and the second is ArcObj-based deep metric learning to improve AUC. These techniques were validated on image datasets of cookies in plastic bags and transparent plastic bottles simulating a food industry application, and lenses simulating an optical industry application.

As discussed in Section IV-C, the ablation study confirmed that our proposed techniques successfully improve AUC (0.04 points improvement by PIF and 0.03 points improvement by ArcObj-based deep metric learning). By using both techniques, AD accuracy is improved. Compared with the state-of-the-art AD algorithm Patchcore [6], the accuracy is improved by 0.09 points with the reflective object dataset of cookies wrapped in transparent plastic bags.

The most significant feature of the developed technology is its excellent capability in detecting anomalies in transparent objects, which is difficult even for the human eye. We expect our technology to be applicable to not only visual inspection systems, but also a wide range of computer vision applications, such as industrial robots for smart logistics and smart factory applications where transparent object detection is required.

The technical challenge of the proposed method is the detection of small anomalies. Our proposed polarization camera implements four polarizers on the same image sensor surface, which reduces the resolution to 1/4 that of RGB cameras. In addition, detailed optimization at the system level, such as automatic adjustment of the distance between the target and the camera, is required.

ACKNOWLEDGMENT

This research is based on results obtained from a project, JPNP20004, subsidized by the New Energy and Industrial Technology Development Organization. The authors would like to thank Prof. S. Fujioka and Dr. K. Yamanoi, Osaka University, Osaka, Japan, for providing the lens dataset.

REFERENCES

[1] V. Crescitelli, A. Kosuge, and T. Oshima, "POISON: Human pose estimation in insufficient lightning conditions using sensor fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 2504408, doi: 10.1109/TIM.2020.3043872.

[2] A. Kosuge, S. Suehiro, M. Hamada, and T. Kuroda, "mmWave-YOLO: A mmwave imaging radar-based real-time multi-class object recognition system for ADAS applications," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 2509810, doi: 10.1109/TIM.2022.3176014.

[3] A. Kosuge, K. Yamamoto, Y. Akamine, and T. Oshima, "An SoC-FPGA based Iterative closest point accelerator enabling faster picking robots," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3567–3576, Apr. 2021.

[4] L. Ruff, "A unifying review of deep and shallow anomaly detection," *Proc. IEEE*, vol. 109, no. 5, pp. 756–795, May 2021.

[5] G. Wang et al., "Student-teacher feature pyramid matching for anomaly detection," 2021, *arXiv:2103.04257*.

[6] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *Proc. IEEE/CVF Comput. Vis. Pattern Recognit.*, 2022, pp. 14318–14328.

[7] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD — A comprehensive real-world dataset for unsupervised anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 9584–9592.

[8] C. Radier et al., "10 PW peak power femtosecond laser pulses at ELI-NP," *Cambridge Univ. Press High Power Laser*, vol. 10, pp. 1–5, May 2022.

[9] J. Bromage et al., "MTW-OPAL: A technology development platform for ultra-intense optical parametric chirped-pulse amplification systems," *Cambridge Univ. Press High Power Laser*, vol. 9, pp. 1–12, Oct. 2021.

[10] L. Yu, A. Kosuge, M. Hamada, and T. Kuroda, "An anomaly detection system for transparent objects using polarized-image fusion technique," in *Proc. IEEE Sensors Application Symp.*, 2022, pp. 1–6.

[11] T. Yamazaki et al., "Four-directional pixel-wise polarization CMOS image sensor using air-gap wire grid on 2.5-μm back-illuminated pixels," in *Proc. IEEE Int. Electron Devices Meeting*, 2016, pp. 8.7.1–8.7.4.

[12] S. Akcay et al., "GANomaly: Semi-supervised anomaly detection via adversarial training," in *Proc. AFCV Asian Conf. Comput. Vis.*, 2018, pp. 622–637.

[13] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5450–5463, Nov. 2019.

[14] Papers with Code, "Anomaly detection on MVTec AD," Accessed: Feb. 12, 2023. [Online]. Available: https://paperswithcode.com/sota/anomaly-detection-on-mvtec-ad

[15] Q. Fan, J. Yang, G. Hua, B. Chen, and D. Wipf, "A generic deep architecture for single image reflection removal and image smoothing," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 3258–3267.

[16] X. Zhang, R. Ng, and Q. Chen, "Single image reflection separation with perceptual losses," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4786–4794.

[17] C. Lei, X. Huang, M. Zhang, Q. Yan, W. Sun, and Q. Chen, "Polarized reflection removal with perfect alignment in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1747–1755.

[18] S. Hartwig and T. Ropinski, "Training object detectors on synthetic images containing reflecting materials," 2019, *arXiv.1904.00824*.
[19] N. Gray et al., "GLARE: A dataset for traffic sign detection in sun glare," 2022, *arXiv.1904.00824*.
[20] S. Sajjan et al., "Clear grasp: 3D shape estimation of transparent objects for manipulation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 3634–3642.
[21] Lucid Vision Lab., "Phoneix GiGE camera 050S," Accessed: Feb. 12, 2023. [Online]. Available: https://thinklucid.com/ja/product/phoenix-5-mp-imx264/
[22] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *Proc. IEEE/CVF Comput. Vis. Pattern Recognit.*, 2017, pp. 212–220.
[23] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," in *Proc. IEEE Pacific Conf. Comput. Graph. Appl.*, 2007, pp. 382–390.
[24] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Comput. Vis. Pattern Recognit.*, 2019, pp. 4690–4699.
[25] OpenCV, "Smoothing images," Accessed: Feb. 12, 2023. [Online]. Available: https://docs.opencv.org/3.4/dc/dd3/tutorial_gausian_median_blur_bilateral_html
[26] Numpy, "Standard deviation," Accessed: Feb. 12, 2023. [Online]. Available: https://numpy.org/doc/stable/reference/generated/numpy.std.html
[27] R. Ranjan et al., "L2-constrained softmax loss for discriminative face verification," 2017, *arXiv:1703.09507*.

ATSUTAKE KOSUGE (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 2012, 2014, and 2016, respectively.

From 2014 to 2017, he was a JSPS Research Fellow at Keio University. From 2017 to 2020, he held research positions at Hitachi Ltd. and Sony Corporation. In 2021, he was with The University of Tokyo. He is currently an Assistant Professor of Systems Design Lab (d.lab) and a Researcher of Research Association for Advanced Systems. His research interests include energy efficient computing, computational sensing, and 3-D integration technologies.

Dr. Kosuge was a Member for the Technical Program Committee of IEEE A-SSCC (Asian Solid-State Circuits Conference) from 2021, IEICE ICD (Integrated Circuit and Devices), and was a Member for the Organizing Committee of IEEE COOL Chips (Symposium on Low-Power and High-Speed Chips and Systems). He was the recipient of the 2013 Nikkei Electronics Japan Wireless Technology Best Award, 2020 IEICE Young Researcher's Award, and co-recipient of the ASP-DAC'15 Special Feature Award.

LIXING YU received the B.S. degree from Nanjing University, Nanjing, China, and the M.S. degree from the University of Tokyo, Tokyo, Japan, in 2023.

Since 2021, he has been engaged in research on the computer vision system using polarized image sensors for industrial applications.

KAZUKI MATSUO (Member, IEEE) received the Ph.D. degree in physics from Osaka University, Osaka, Japan, in 2020.

From 2018 to 2020, he was a JSPS Research Fellow at Osaka University. In 2021, he was a Postdoctoral Researcher with the University of California San Diego. He is currently the Chief Executive Officer of EX-Fusion Inc. His research interests include Inertial Confinement Fusion for energy production, high energy density plasma physics, and developing X-ray diagnostics.

Dr. Matsuo was the recipient of the 2023 Young Scientist Award of the Physical Society of Japan.

MOTOTSUGU HAMADA was born in Nara, Japan, in 1968. He received the B.S., M.S., and Ph.D. degrees in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1991, 1993, and 1996, respectively.

In 1996, he joined Toshiba Corporation and was engaged in wireless and low-power electronic circuits design with Toshiba's Center for Semiconductor Research and Development, Kawasaki, Japan. From 2002 to 2004, he was a Visiting Scholar with Stanford University. From 2011 to 2016, he was with Mixed Signal IC Division as a Group Manager of Power Analog IC Design Group to lead the development of analog mixed signal ICs. In 2016, he joined Keio University and was a Project Professor. In 2020, he was with the University of Tokyo, where he is currently a Project Professor of Systems Design Laboratory. His research interests include low-power, high-speed CMOS design, low-power wireless systems and circuits design, and power management systems design.

Dr. Hamada was the recipient of the 2007 IEEE International Conference on Computer Design Best Paper Award and the recipient of the Design Automation Conference 2010 Best User Track Poster Award. He was also recognized in the list of "AUTHORS of TEN OR MORE PAPERS IN THE PAST TEN YEARS" at the International Solid-State Circuits Conference 2013(ISSCC2013). He was a Member of the technical program committee of International Solid-State Circuits Conference (2003–2009, 2011) and VLSI Circuits Symposium (2018–2023), and Asian Solid-State Circuits Conference (2005–2012,2017–2022) where he was the RF Subcommittee Chair, Digital Subcommittee Chair, Student Design Contest Chair, and Technical Program Committee Chair. He is a Senior Fellow of Research Association for Advanced Systems.

TADAHIRO KURODA (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1999.

In 1982, he was with Toshiba Corporation. From 1988 to 1990, he was a Visiting Scholar with the University of California, Berkeley, where he conducted research in the field of VLSI CAD. In 1990, he was back to Toshiba, and got engaged in the research and development of BiCMOS ASICs, ECL ASICs, and high-speed low-power CMOS LSIs. He invented a Variable Threshold-voltage CMOS technology to control V_{TH} through substrate bias, and applied it to a DCT core processor in 1995. He also developed a Variable Supply-voltage scheme to control V_{DD} by an embedded dc–dc converter, and employed it to a microprocessor core and an MPEG-4 chip in 1997. He left Toshiba to join Keio University in 2000, and became a Full Professor in 2002. He was the Mackay Professor with the University of California, Berkeley, in 2007. He invented a ThruChip Interface (TCI) by using magnetic coupling for communications among stacked chips in 2008, and a Transmission Line Coupler by using electromagnetic coupling for communications among stacked PCBs in 2012. He left Keio to join the University of Tokyo in 2019. He is the Director of Systems Design Lab (d.lab) at the University of Tokyo, and the Chairperson of Research Association for Advanced Systems (RaaS). He has authored or coauthored more than 500 papers, including 40 ISSCC papers, 30 VLSI Symposia papers, 19 CICC papers and 19 A-SSCC papers. He wrote 30 books/chapters and filed more than 200 patents.

Dr. Kuroda is an IEICE Fellow, and the Chair of Symposium on VLSI Technology and Circuits. He was an elected AdCom member of two terms. He was a recipient of the 2005 P&I Patent of the Year Award, the 2007 ASP-DAC Best Design Award, the 2009 IEICE Achievement Award, and the 2011 IEICE Society Award. He served as a Steering Committee Chair for A-SSCC, a Vice Chair for ASP-DAC, sub-committee chairs for IEEE Asian Solid-State Circuits Conference, International Conference on Computer-Aided Design, SSDM, and International Symposium on VLSI Design, Automation and Test, and TPC members for ISSCC, Symposium on VLSI Circuits, CICC, DAC, ASPDAC, ISLPED, SSDM, ISQED, and other international conferences. He was a Distinguished Lecturer and a representative of Region 10 for the IEEE Solid-State Circuits Society.