Privacy-Aware Face Recognition with Lensless Multi-pinhole Camera

Yasunori Ishii[1(✉)], Satoshi Sato[1], and Takayoshi Yamashita[2]

[1] Panasonic Corporation, 1006 Kadoma, Kadoma, Osaka, Japan
{ishii.yasunori,sato.satoshi}@jp.panasonic.com
[2] Chubu University, 1200 Matsumotocho, Kasugai, Aichi, Japan
takayoshi@isc.chubu.ac.jp

Abstract. Face recognition and privacy protection are closely related. A high-quality facial image is required to achieve a high accuracy in face recognition; however, this undermines the privacy of the person being photographed. From the perspective of confidentiality, storing facial images as raw data is a problem. If a low-quality facial image is used, to protect user privacy, the accuracy of recognition decreases. In this paper, we propose a method for face recognition that solves these problems. We train a neural network with an unblurred image at first, and then train the neural network with a blurred image, using the features of the neural network trained with the unblurred image, as an initial value. This makes it possible to train features that are similar to the features trained with the neural network using a high-quality image. This enables us to perform face recognition without compromising user privacy. Our method consists of a neural network for face feature extraction, which extracts suitable features for face recognition from a blurred facial image, and a face recognition neural network. After pretraining both networks, we fine-tune them in an end-to-end manner. In experiments, the proposed method achieved accuracy comparable to that of conventional face recognition methods, which take as input unblurred face images from simulations and from images captured by our camera system.

Keywords: Coded aperture · Lensless multi-pinhole camera · Face recognition · Image debluring

1 Introduction

Face recognition [12,25,46] is an important task in various applications. A facial image is personal information, and so we need to consider privacy when we include face recognition in these applications. The European Union has enforced the General Data Protection Regulation, which requires the protection of personal information. In addition, the regulation of privacy protection may expand worldwide in the future. We must therefore pay close attention to the protection of privacy when using facial images. However, it is difficult to successfully realize both privacy protection and face recognition.

A. Bartoli and A. Fusiello (Eds.): ECCV 2020 Workshops, LNCS 12539, pp. 476–493, 2020.
https://doi.org/10.1007/978-3-030-68238-5_35

Nodari et al. [29] proposed a method of decreasing the visibility of a facial region with a mosaic, in Google Street View. Padilla-Lopez et al. [30] proposed a method of replacing a facial image with a public image, to avoid privacy concerns. Fernande et al. [13] proposed a blurring method for a self-moving robot. Thorpe et al. [44] proposed a method using two types of blurred images, which differ depending on whether they are public or private images. These methods focus on image processing after an image is captured, and must overcome the serious problem of storing the facial image securely, because stored facial images can be leaked, and privacy can be violated through hacking.

To protect privacy, an image should be captured with enhanced security for personal information. As examples, an event camera records only the change in brightness of pixels for each frame [15], and thermal-image-based recognition also records information without detailed personal information [14]. Browarek et al. [4] proposed a method for human detection that uses an infrared sensor. Dai et al. [11] proposed a privacy protection method that uses a low-resolution camera. However, although these methods preserve privacy, they have difficulty in recognizing faces with high accuracy.

Computational photography [33] is another perspective on privacy protection that is worth considering. As an example of such technologies, Cossalter et al. [10] proposed a method that protects privacy by random projection based on compressive sensing. It captures an optically blurred image using a coded-aperture camera [6,17,19,23,24,26,41,45], in which a coded mask is arranged in front of the aperture. Pittaluga et al. [31,32] proposed a method of shooting an optically blurred image with a multi-aperture mask using three-dimensionally printed optics. However, its effectiveness in recognition technology is not clear because image recognition is not evaluated quantitatively. Wang et al. [48] proposed an action recognition method that protects privacy using a coded aperture camera. However, the recognition accuracy is low because it is insufficient as a feature extraction method for improving the accuracy of image recognition. Canh et al. [5] proposed a method in which it is possible to select whether to reconstruct the area excluding the face area or including the face area from the blurred image; however, they do not describe the application that uses the reconstructed image.

It is difficult to identify an individual from a blurred image captured by the coded aperture camera. However, it is possible to obtain a reconstructed image from the blurred image if we know a code pattern of the mask. Inspired by this technology, we propose a multi-pinhole camera (MPC) with a mask that has multiple pinholes. We also propose a face recognition method that achieves the accuracy of non-blurred images even when we use privacy-preserving blurred images. In this paper, an image captured with a normal camera is called an unblurred image, and an image captured with an MPC, in which privacy is protected, is called a blurred image. In general, the features of a blurred image are ineffective for face recognition, and reconstruction methods are employed in prepossessing, to obtain effective features. ConvNet-based methods reconstruct high-quality facial images [8,9,16,21,28,35,37,39,42,43,47,50,51]. However, these methods require substantial memory and have a high computational cost.

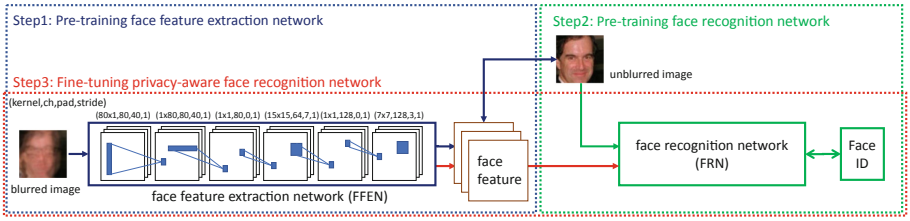

Fig. 1. Training steps of the proposed face recognition method. The training of the proposed method comprises three steps. First, we pretrain the face feature extraction network, using blurred images and unblurred images, for the extraction of face features. Second, we pretrain the face recognition network using unblurred images. Finally, we fine-tune the privacy-aware face recognition network

As an alternative approach, Xu et al. [49] and Ren et al. [36] proposed restoration methods for slight blur using shallow neural networks. These methods are based on the fact that a blurred image is constructed by convoluting an unblurred image with a point spread function (PSF). The methods can both reduce the calculation cost and suppress blurring in the blurred image, even when face reconstruction is difficult. Therefore, the methods are suitable for face feature extraction, while privacy is protected because face reconstruction is difficult, because of their use of a shallow neural network.

Face recognition and privacy protection are inseparable. It is difficult to protect privacy when effective features can be extracted from an unblurred image. However, face recognition accuracy decreases when the restored face features are protected by privacy because the blurred features remain. To solve this dilemma, we focus on the extraction of effective features, rather than the face image. Additionally, it is possible to recognize a face in an extremely blurred image.

We propose a method of improving the face recognition accuracy in a privacy-protected image to solve the above dilemma. Our method involves capturing an extremely blurred facial image using a lensless MPC, for privacy protection. The training of the proposed method comprises three steps, as shown in Figure 1. First, we pretrain the face feature extraction network (FFEN), which extracts face features from a blurred image. Second, we pretrain the face recognition network (FRN) using unblurred images. Finally, we train a privacy-aware FRN, which is the connected FFEN and FRN, as one network, in an end-to-end manner.

The contributions of the paper are as follows.

- New attempts are made to achieve both privacy protection and high recognition accuracy.
- Effective features are extracted from a restored facial image captured by a lensless MPC.
- High facial recognition accuracy is achieved even if the facial image is extremely blurred.

(a) (b) (c)

Fig. 2. Proposed MPC: (a) a lensless MPC, (b) an enlarged image of the coded mask captured with a microscope, and (c) a PSF of the coded mask (b)

Table 1. Specifications of the coded mask

PSFid	Diameter (μm)	Num. of pinholes	Distance (mm)	PSF size (width, height) (pixels)
3–025	120	3	0.25	(111,113)
3–050	120	3	0.50	(191,174)
9–025	120	9	0.25	(140,140)
9–050	120	9	0.50	(271,272)

2 Lensless Multi-pinhole Camera

The MPC captures a blurred image such that privacy is protected, even if a raw image stored in memory is leaked. Conventional coded apertures are intended to record light rays. Their image is therefore only slightly blurred, and privacy protection is not possible. In contrast, in the MPC, a coded mask with multiple pinholes is arranged in front of the aperture. Multiple blurs can be superimposed because light rays from the same object are incident on each pinhole. It is difficult to identify the individual in such a multiple-blurred facial image.

2.1 Design of the Proposed Lensless Multi-pinhole Camera

Various methods have been proposed to reconstruct deblurred images from images captured by a lensless camera [2,20,22,27,40]. These methods have a high cost because the imaging systems must be changed substantially. In contrast, the setup of the lensless MPC only requires a coded mask to be attached in front of the aperture. Therefore, the lensless MPC is more versatile and practical. Figure 2 shows our lensless MPC and coded mask. We employ an FLIR Blackfly BFLY-U3-23S6C-C for the body of the lensless MPC, as shown in Fig. 2(a). Figure 2(b) is an example of an enlarged image of the coded mask captured with a microscope. Figure 2(c) is the PSF corresponding to the coded mask in Fig. 2(b). The blurred image is a spatial convolution of an unblurred image and the PSF. Therefore, if the PSF is unknown, it is difficult to reconstruct the unblurred image from the blurred image. Conversely, if the PSF is known in advance, as it is for our system, the unblurred image can be reconstructed easily, by inversely convoluting the blurred image with the PSF. Because the PSF is different for each camera, a hacker would need to steal the camera to discover the PSF. The hacker would then need to install the camera in its original location after measuring the PSF. Therefore, the PSF leakage risk is lower than the image leakage risk in an actual scene. Even if a blurred image is leaked from the network or data storage, the risk of image restoration is small.

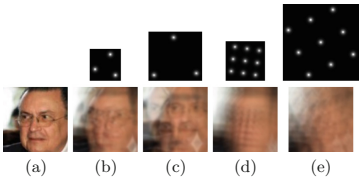

Fig. 3. Examples of simulated images convoluted by a PSF: (a) an unblurred image, and images convoluted by (b) PSFid:3–025, (c) PSFid:3–050, (d) PSFid:9–025, and (e) PSFid:9–050

2.2 PSF of the Proposed Camera

The PSF represents the response of an optical system to a point light source, in terms of the spread of spatial blur. If we know the PSF function as prior information, we can obtain the unblurred image by convolving the blurred image and the inverse PSF. We can measure the PSF before capturing an image.

We prepared four types of coded mask. The specifications of each coded mask are given in Table 1, including the number of pinholes and the distance from the center of the mask to the nearest pinhole. Figure 3 shows the measured PSF and an example of a blurred image, for each coded mask. The overlapping of objects increases with the number of pinholes. In the case where there are three pinholes and the distance from the center of the mask to each pinhole is 0.25 mm, the facial image is blurred, as shown in Fig. 3(b). It is difficult to recognize the individual in the image because parts of the face that exhibit individual characteristics, such as the eyes, nose, and mouth, are blurred. The facial image is extremely blurred when we combine nine pinholes with a distance from the center of the mask of 0.50 mm. It is even difficult to recognize the image as being that of a human face, with these settings.

3 Face Recognition from Images Captured by the Lensless Multi-pinhole Camera

We propose a method of recognizing a face in a blurred image captured by the lensless MPC. As shown in Fig. 1, the proposed method adopts the FFEN and FRN. In the FFEN, we extract effective face features from the blurred image. We can employ a state-of-the-art FRN if we obtain features similar to the features of the unblurred image. The important aspect of the proposed method is that, to preserve privacy, we do not reconstruct the deblurred facial image explicitly. In the proposed method, we train both networks in an end-to-end manner to obtain suitable face features for the FRN. The training of the proposed method comprises three steps. First, we pretrain the FFEN, which extracts face features from a blurred image. Second, we pretrain the FRN using unblurred images. Finally, we train a privacy-aware FRN, which is the connected FFEN and FRN, as one network, in an end-to-end manner. To protect privacy, we use these images only for training.

Many reconstruction methods have been proposed, but they focus on reconstruction of the entire face. This approach fails to reconstruct detail in the facial region. However, facial areas exhibiting individual characteristics are important features from the viewpoint of facial recognition. In contrast, in our approach, the FFEN focuses on the extraction of face features from a blurred image instead of a high-quality facial image.

The FRN extracts features that are effective in verifying the individual and can be easily used with state-of-the-art methods. We employ metric-learning-based methods that achieve high recognition accuracy, such as ArcFace [12], CosFace [46], and SphereFace [25]. After pretraining the FFEN and FRN, we fine-tune both networks using a blurred facial image to extract suitable features.

3.1 Pretraining of the FFEN

We measure the PSF of the lensless MPC before training the FFEN. We initialize the parameter of the network by calculating the inverse PSF, following [36,49]. The blurred image y is obtained by convolving the unblurred image x and PSF k, as expressed in Eq. (1).

$$y = k * x. \tag{1}$$

Here, $*$ is the convolution operation. Equation (1) is replaced by Eq. (2) in the frequency domain. The convolution operation is the product of each element in the frequency domain.

$$\mathcal{F}(y) = \mathcal{F}(k) \times \mathcal{F}(x). \tag{2}$$

Here, $\mathcal{F}(\cdot)$ is the discrete Fourier transform. After converting to the frequency domain, we convert the blurred image y to the unblurred image x by Eq. (3).

$$x = \mathcal{F}^{-1}(1/\mathcal{F}(k)) * y. \tag{3}$$

The function $\mathcal{F}^{-1}(\cdot)$ is the inverse Fourier transform. To prevent division by zero in the frequency domain, the Wiener filter, expressed in Eq. (4), is used.

$$\begin{aligned} x &= \mathcal{F}^{-1}(1/\mathcal{F}(k)\{\frac{|\mathcal{F}(k)|^2}{|\mathcal{F}(k)|^2 + \frac{1}{SNR}}\}) * y \\ &= k^{\dagger} * y. \end{aligned} \tag{4}$$

Here, $k^{\dagger}$ is the pseudo-inverse PSF and the SNR is the signal-to-noise ratio in the pseudo-inverse PSF. If the SNR is large, it is robust to noise.

The pseudo-inverse PSF can be resolved into $k^{\dagger} = USV^T$ through singular value decomposition (SVD). The elements of the j^{th} rows of U and V are u_j and v_j, respectively, and the j^{th} singular value is s_j. In Eq. (4), SVD replaces the convolution of the two-dimensional pseudo-inverse PSF with the product of the convolution of the one-dimensional vectors u_j and v_j and the scalar s_j, as in Eq. (5).

$$x = \sum_j s_j \cdot u_j * (v_j^T * y). \tag{5}$$

Conversion from the blurred image to the unblurred image using the pseudo-inverse PSF can be considered to be the adoption of a convolutional neural network taking s_j, u_j, and v_j^T as the convolutional kernels of three layers. These three layers have neither an activation function nor normalization, such as batch normalization. We use the outlier rejection subnetwork in addition to the last three layers, following [36,49].

The FFEN module in Fig. 1 shows the network architecture. The first and second layers have $K \times 1$ and $1 \times K$ kernels, respectively. Both layers have K channels. The third layer has a 1×1 kernel with K channels. The initial values of the kernels are the K eigenvectors and eigenvalues selected from the larger eigenvalue. The kernel sizes of the fourth, fifth, and sixth layers are 15×15, 1×1, and 7×7, respectively. The fourth, fifth, and sixth layers have 64, 128, and 128 channels, respectively. In optimization, we use the L_1 loss for FFEN.

$$\begin{aligned} loss_{FFEN} &= \frac{1}{N} \sum_{n=1}^{N} |x_n - z_n| \\ z &= DF(y) \end{aligned} \tag{6}$$

Here, N is the number of pixels, x is the unblurred image, y is the blurred image, and $DF(y)$ is the face feature of y.

3.2 Pretraining of the FRN

The FFEN outputs features that are effective for face recognition. These features are then input to the FRN. We first perform pretraining of the FRN with unblurred facial images. The FRN network is based on ArcFace [12], CosFace [46], and SphereFace [25], which are state-of-the-art methods. ArcFace obtains effective features using cosine distance. The loss function for pretraining in ArcFace is given by Eq. (7).

$$loss_{arcface} = -\frac{1}{M} \sum_{i=1}^{M} log \frac{e^{s(cos(\theta_{y_i} + m)))}}{e^{s(cos(\theta_{y_i} + m))} + \sum_{i=1, j \neq y_i}^{n} e^{scos(\theta_j)}}. \tag{7}$$

Here, M is the number of data items, s is the scale parameter for cosine similarity, and m is the margin with other classes.

To recognize whether two facial images are of the same person, we extract features of the faces using the trained FRN. The two facial images are of the same person if the cosine distance between the extracted features is greater than or equal to a threshold.

3.3 Fine-Tuning of the Privacy-Aware FRN

The FFEN and FRN are trained independently. The output from the FFEN comprises face features, and the input to the FRN is the face features. We perform fine-tuning to adapt to the input and output of the two networks in

an end-to-end manner, using the loss function given by Eq. (7). The proposed method is less affected by blur because of the combination of these networks.

For our networks, particularly the FFEN, the feature extraction accuracy of the entire face region is not essential. By fine-tuning both networks, it is possible to extract the feature only the region in which it is effective to extract features for recognizing the individual. Even if the subject wears eyeglasses, and there are few samples of faces wearing eyeglasses in the training data, the network can extract features in other important regions. When we do not pretrain the FFEN, it is necessary to extract features of the entire facial image that represent individual characteristics. However, it is difficult to extract them because the network cannot extract a feature of a small region that represents individual characteristics. The proposed method improves the accuracy of face recognition by training a FFEN that extracts feature maps representing individual characteristics.

4 Experiments

4.1 Details of Implementation

The parameters of each layer of the FFEN are shown in Fig. 1. An activation function is not arranged in the first three layers. In the second three layers, Leaky ReLU, with a gradient of 0.02, is arranged as an activation function. The mini-batch size is 1, the learning rate is 0.0001, and the number of iterations is 50 epochs in the FFEN. We use 58,346 images, randomly sampled from MS1MV2 [12] and LFW [18]. To validate the performance of feature extraction, we use 58,346 images, randomly sampled from MS1MV2 and LFW images, that are not used in training.

We employ SphereFace [25], CosFace [46], and ArcFace [12] as the FRN. The backbone network is ResNet50. We use the MS1MV2 dataset for training. The number of images is 5,822,653 and the number of IDs is 85,741. We use LFW [18], CPLFW [3], and CALFW [7] for the evaluation data; there are 12,000 images in each dataset. Each image is normalized in terms of orientation and cropped to 112×112 pixels. The mini-batch size is 256 and there are four epochs. The initial learning rate for pretraining is 0.1, and the learning rate is multiplied by 0.1 in epoch 3. The learning rate, momentum, and weight decay in fine-tuning are determined by adopting Bayesian optimization [1].

Public face recognition datasets do not include both unblurred and blurred images. Therefore, we first simulate the blurred images using the PSF of this camera, as shown in the leftmost four columns of Fig. 4. Blurring of PSFid:3–025 can be seen for the eyes, nose, and mouth, and it is difficult to recognize the individual. In the case of PSFid:3–050, the distance of each pinhole from the center of the mask is large, and it is possible to identify the individual, but the positional deviation is large and feature extraction is therefore difficult. It is generally possible to identify the shape of the facial contours in the blurred image of PSFid:9–025, but it is difficult to identify facial parts, because of the blur. For the blurred image of PSFid:9–050, it is difficult to identify the contours of the face and face parts. In order to prevent personal identification, our method blurs

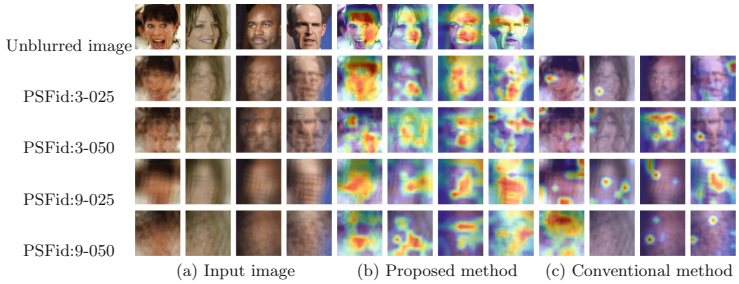

Fig. 4. Examples of unblurred and blurred images and attention maps of each image. (a) Left four columns: examples of unblurred images and blurred images for each PSFid. (b) Center four columns: attention maps of unblurred images and proposed method for each PSFid. (c) Right four columns: attention maps of conventional method for each PSFid

the face by overlaying the image. When the number of pinholes is small or the distance from the center of the mask is large, the area where the image overlaps decreases. In this case, it is difficult to protect privacy, so in order to make it difficult to identify individuals, it is desirable to make pinholes where face images overlap in many areas. We used blurred images captured by this apparatus as real images in the experiments reported below, in which we evaluated a simulated image against a real image captured by the lensless MPC.

4.2 Face Recognition Results of the Privacy-Aware FRN

In this section we present the results of the pretraining of the FFEN and FRN, and the results of the fine-tuning of the privacy-aware FRN for each PSF. Table 2 shows the face recognition results of each PSF for LFW, CPLFW, and CALFW. In Table 2, the first column shows the dataset, the second column shows the PSFid, and the third and subsequent columns show evaluation results using different FRN algorithms. (A), (B), (C), and (D) show the result of training with a blurred image, the result of training without pretraining the FRN, the result without fine-tuning, and the result of the proposed method, respectively. (A) is a conventional result. Each row shows the result for different coded masks, and the other rows show the results of SphereFace, CosFace, and ArcFace trained with unblurred images. The first value of each PSFid is the number of pinholes, and the second value is the distance of the pinholes from the mask center.

When there are three pinholes, the performance is similar to that when unblurred images are used in training. Even for nine pinholes, the performance of the proposed method is superior to that without pretraining or fine-tuning. This result shows that both pretraining and fine-tuning, which are the training steps of the proposed method, are effective. The recognition rate of CPLFW and CALFW is lower than that of LFW. This is not limited to this study, but it has been reported that this trend is similar in [12]. CPLFW performs face verification of pair images with different face pose, and CALFW performs face

Table 2. Comparison of face verification results (%)

		Basis network: SphereFace				Basis network: CosFace				Basis network: ArcFace			
Dataset	PSFid	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)
LFW	3–025	98.6	97.2	96.5	**99.2**	99.1	98.8	98.2	**99.4**	98.4	99.1	98.2	**99.4**
	3–050	97.7	98.6	93.6	**99.4**	98.5	98.0	95.2	**99.2**	97.8	99.0	94.3	**99.4**
	9–025	97.4	98.4	90.8	**99.0**	97.8	97.1	92.6	**98.8**	97.0	95.4	92.0	**99.1**
	9–050	93.7	92.8	84.1	**97.7**	94.6	91.1	84.9	**97.3**	90.3	92.9	84.4	**97.7**
	SphereFace [25]	99.3				99.3				99.3			
	CosFace [46]	99.5				99.5				99.5			
	ArcFace [12]	99.5				99.5				99.5			
CPLFW	3–025	86.4	80.2	82.1	**87.1**	85.6	85.2	84.1	**88.0**	82.0	86.3	82.8	**89.0**
	3–050	82.4	85.0	77.4	**88.4**	83.8	77.9	79.2	**83.6**	80.1	85.4	77.8	**88.0**
	9–025	81.1	82.9	74.1	**86.1**	82.1	74.2	74.8	**82.4**	78.9	77.8	73.8	**85.3**
	9–050	74.2	72.8	65.9	**76.4**	73.3	71.2	66.3	**75.5**	67.4	73.4	66.1	**78.5**
	SphereFace [25]	87.7				87.7				87.7			
	CosFace [46]	87.6				87.6				87.6			
	ArcFace [12]	88.3				88.3				88.3			
CALFW	3–025	91.6	84.5	86.5	**93.4**	92.7	92.2	91.5	**94.5**	91.0	93.6	90.9	**95.0**
	3–050	90.1	92.6	82.8	**94.3**	91.3	90.6	87.0	**93.0**	90.4	93.1	85.9	**94.3**
	9–025	87.4	90.5	78.1	**92.7**	88.8	87.5	82.2	**91.7**	87.4	79.9	81.0	**93.3**
	9–050	81.9	75.4	71.6	**91.1**	83.2	72.7	74.3	**90.4**	76.3	75.5	73.1	**90.8**
	SphereFace [25]	94.3				94.3				94.3			
	CosFace [46]	94.9				94.9				94.9			
	ArcFace [12]	95.0				95.0				95.0			

verification of pair images of different ages. Therefore, CPLFW and CALFW are more difficult images than LFW.

4.3 Analysis Using the Area of Focus of Features and Extracted Features

We visualize whether a fine-tuning model extracts face features, using Grad-Cam [38]. ArcFace obtains similarity based on the cosine distance between feature vectors. The visualization is performed using a one-hot vector that has a value of 1 for the most similar person.

The leftmost four columns (a) of Fig. 4 show the unblurred image and the blurred image for each PSFid, the center four columns (b) show the attention maps of the proposed method for each PSFid, and the rightmost four columns (c) show the attention maps of the blurred image (conventional method) for each PSFid. When the face has little blur, such as in the case of PSFid:3–025, the attention maps of both the unblurred image and the proposed method are similar in position and strength within the area of the face. For other PSFs, the position of the attention map is slightly different, but parts of the face such as the eyes, nose, and mouth respond strongly. In the case of the conventional method, the attention maps are largely outside the face. It is therefore difficult to obtain effective features for face recognition from the blurred image.

(a) 3-025 (b) 9-025

Fig. 5. Examples of captured images. Real captured images are more blurred than simulated images

Table 3. Face verification results for captured images (%)

		Basis network: SphereFace				Basis network: CosFace				Basis network: ArcFace			
Dataset	PSFid	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)
LFW	3–025	92.4	88.5	80.7	**93.8**	92.9	88.5	84.1	**96.3**	92.7	73.0	82.4	**94.2**
	9–025	86.7	80.9	73.4	**90.4**	86.6	81.3	75.0	**89.8**	87.1	66.7	75.5	**88.9**
	SphereFace [25]	97.9				97.9				97.9			
	CosFace [46]	99.1				99.1				99.1			
	ArcFace [12]	96.7				96.7				96.7			
CPLFW	3–025	68.5	64.7	60.5	**71.1**	70.7	64.6	62.8	**75.0**	69.4	59.1	61.2	**72.0**
	9–025	64.0	60.5	56.9	**67.3**	63.6	60.4	59.2	**75.9**	64.9	55.6	58.0	**66.5**
	SphereFace [25]	75.7				75.7				75.7			
	CosFace [46]	81.9				81.9				81.9			
	ArcFace [12]	60.9				60.9				60.9			
CALFW	3-025	78.1	72.4	68.1	**93.7**	80.9	73.6	70.8	**86.6**	77.8	58.4	68.5	**79.9**
	9-025	71.0	64.7	61.0	**76.4**	71.8	65.3	62.7	**75.9**	68.6	55.5	62.0	**71.0**
	SphereFace [25]	89.4				89.4				89.4			
	CosFace [46]	92.9				92.9				92.9			
	ArcFace [12]	57.7				57.7				57.7			

4.4 Experiments Using Real Images

We compared the accuracy of face recognition for a real image, using blurred images captured by the lensless MPC for PSFid:3-025 and 9-025. The unblurred image was displayed on the monitor in a dark room and considered as a captured image with real blur. As a result, a pair, comprising an unblurred image and a real blurred image, was obtained. To train the FFEN, we used 53,143 images randomly sampled from MS1MV2 and LFW. The captured images are presented in Fig. 5. The real image used in the experiment was more blurred than the simulated image. To train the FRN, we used 147,464 images sampled from MS1MV2. The image size is 112×112 as well as simulation.

Comparison results are shown in Table 3. The proposed method achieved higher accuracy than the conventional method. Although the blurred image was extremely blurred and there was little training data, the setting for both PSFid:3–025 and PSFid:9–025 had higher accuracy than the other settings. In an experiment using real images, pretraining and fine-tuning were effective, as in an experiment using simulated data. Because the proposed method achieved high accuracy even with real images, we conclude that it achieves face recognition that can protect privacy.

4.5 Evaluation of Privacy Protection Performance of Proposed System

We evaluate the privacy protection performance of blurred images. As noted in Sect. 2.1, PSF does not leak. Therefore, we evaluated the privacy protection performance using CycleGAN [52], which is a generative model, and SelfDeblur [34], which is one of the state-of-the-art methods for blind deconvolution.

For training CycleGAN, we require unblurred and blurred images. Unblurred images were randomly selected from LFW. Two types of blurred images were used: The blurred images selected from LFW did not overlap with the unblurred images from LFW. The number of training images in each set (unblurred images from LFW and blurred images) was 5000. We used images that were not used for training, as the evaluation images. The number of training iteration is 10000. In each PSFid, losses of a generator are shown in Fig. 6. The vertical axis is the loss, and the horizontal axis is the number of the iteration. From this figure, losses converged in approximately 7000 iterations, so in this experiment, sufficient training has been done by 10000 iterations. SelfDeblur estimates the unblurred image and the PSF, given a single image.

Figure 7(a) shows an unblurred image, Fig. 7(b) shows the reconstruction result of the simulated image, and Fig. 7(c) shows the reconstruction result using the captured image. For each PSFid, the figure shows the blurred image, the results of CycleGAN, and the results of SelfDeblur. The image generated by CycleGAN is a sharp image. When the distance between pinholes is small, such as PSFid:3–025 and PSFid:9–025, the contour shape is similar to an unblurred image, but the face parts of the generated image are different from those of the unblurred image. In contrast, when the distance from the center of PSF is large, such as PSFid:3–050 and PSFid:9–050, the unblurred image and the generated image differ greatly in the shape of the contours, in addition to the face parts. Therefore, it is difficult to recognize the blurred image and the generated image as the same person. In general, the more training, the higher the accuracy. However, in GAN, the distribution of training data is trained. In this experiment,

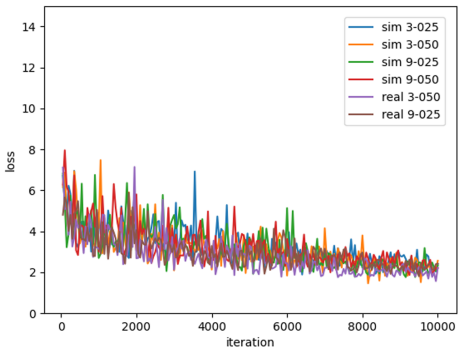

Fig. 6. Training losses of CycleGAN

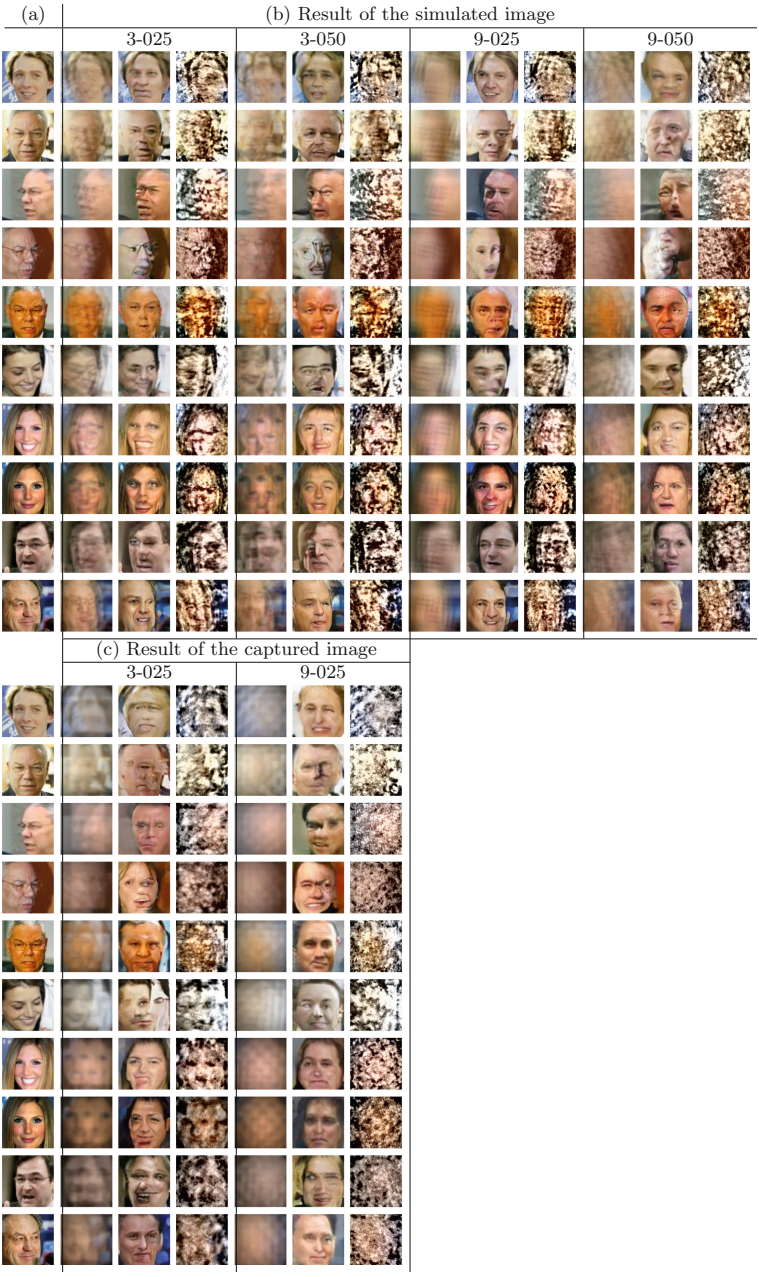

Fig. 7. Example of privacy protection performance. (a) is an unblurred image, (b) is a result of the simulated image, and (c) is a result of the captured image. For each PSFid, the leftmost image is the blurred image, the center image is that generated by CycleGAN [52], and the rightmost image is that reconstructed by SelfDeblur [34]

the feature distribution for expressing the face is trained rather than the character which represents the individual. Therefore, it is possible to generate face images without blur, but since individuality is lost, the face recognition accuracy does not necessarily increase even if the number of images is increased in this experiment.

The deconvolution image created by SelfDeblur from the simulated image can approximately distinguish the face area from the background. However, the artifact is so large that the subject cannot be identified. This tendency is the same for all PSFids, but increasing the number of pinholes causes the face shape to collapse more, making deconvolution difficult. In the result of deconvolution using a captured image, it is difficult even to visually recognize the position of the face area. From these results, it was confirmed that it is difficult to identify the person in image generation and image reconstruction when the PSF is unknown, and the proposed system is effective for privacy protection.

5 Conclusion

We have proposed a privacy-aware face recognition method that solves the dilemma of simultaneously realizing good privacy protection and face recognition accuracy. To be successful at both, we constructed an acquisition system based on a lensless MPC that captures extremely blurred face images. The MPC has several pinholes and captures a blurred image. From this blurred image, we extract face features that are similar to those of an unblurred image using a FFEN. The FFEN is trained with initial parameters calculated using the inverse PSF. An FRN based on ArcFace recognizes a person using the face features. These networks are fine-tuned, in an end-to-end manner, after each is pretrained.

We are concerned that privacy may not be protected in the event that a hacker steals the captured image. If the PSF is unknown, it is difficult to reconstruct the image only from the blurred image; however, if the PSF is known, image reconstruction can be performed relatively easily. However, because the PSF is different for each camera, a hacker would need to measure the PSF, in addition to stealing the captured image. Therefore, in a real environment, it is unlikely that a hacker could recover a blurred image. By experiments using image reconstruction when the PSF is unknown, we showed that it is difficult to reconstruct a blurred image without PSF into an unblurred image.

We experimented with four types of coded masks, but these are not always optimal for privacy protection. In future studies, we intend to design a pattern that is optimal for both recognition and privacy protection, by treating the coded mask pattern as a training parameter. And, The loss of face recognition is back-propagated to FFEN by fine-tuning, but it does not specify explicitly whether to train the effective region for face recognition. We consider effective use of combining with attention and facial feature inspection.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2623–2631 (2019)
2. Asif, M.S., Ayremlou, A., Sankaranarayanan, A., Veeraraghavan, A., Baraniuk, R.G.: Flatcam: thin, lensless cameras using coded aperture and computation. IEEE Trans. Comput. Imag. **3**(3), 384–397 (2016)
3. Best-Rowden, L., Bisht, S., Klontz, J.C., Jain, A.K.: Unconstrained face recognition: Establishing baseline human performance via crowdsourcing. In: IEEE International Joint Conference on Biometrics, pp. 1–8. IEEE (2014)
4. Browarek, S.: High resolution, Low cost, Privacy preserving Human motion tracking System via passive thermal sensing. Ph.D. thesis, Massachusetts Institute of Technology (2010)
5. Canh, T.N., Nagahara, H.: Deep compressive sensing for visual privacy protection in flatcam imaging. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 3978–3986. IEEE (2019)
6. Cannon, T., Fenimore, E.: Tomographical imaging using uniformly redundant arrays. Appl. Opt. **18**(7), 1052–1057 (1979)
7. Chen, B.C., Chen, C.S., Hsu, W.H.: Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. IEEE Trans. Multimedia **17**(6), 804–815 (2015)
8. Chen, R., Mihaylova, L., Zhu, H., Bouaynaya, N.C.: A deep learning framework for joint image restoration and recognition. In: Circuits, Systems, and Signal Processing, pp. 1–20 (2019)
9. Chrysos, G.G., Zafeiriou, S.: Deep face deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 69–78 (2017)
10. Cossalter, M., Tagliasacchi, M., Valenzise, G.: Privacy-enabled object tracking in video sequences using compressive sensing. In: 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 436–441. IEEE (2009)
11. Dai, J., Wu, J., Saghafi, B., Konrad, J., Ishwar, P.: Towards privacy-preserving activity recognition using extremely low temporal and spatial resolution cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 68–76 (2015)
12. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4690–4699 (2019)
13. Fernandes, F.E., Yang, G., Do, H.M., Sheng, W.: Detection of privacy-sensitive situations for social robots in smart homes. In: 2016 IEEE International Conference on Automation Science and Engineering (CASE), pp. 727–732. IEEE (2016)
14. Gade, R., Moeslund, T.B.: Thermal cameras and applications: a survey. Mach. Vis. Appl. **25**(1), 245–262 (2014)
15. Gallego, G., et al.: Event-based vision: A survey. arXiv preprint arXiv:1904.08405 (2019)
16. Gupta, K., Bhowmick, B., Majumdar, A.: Motion blur removal via coupled autoencoder. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 480–484. IEEE (2017)

17. Hiura, S., Matsuyama, T.: Depth measurement by the multi-focus camera. In: Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231), pp. 953–959. IEEE (1998)
18. Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database forstudying face recognition in unconstrained environments (2008)
19. Inagaki, Y., Kobayashi, Y., Takahashi, K., Fujii, T., Nagahara, H.: Learning to capture light fields through a coded aperture camera. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 418–434 (2018)
20. Jiao, S., Feng, J., Gao, Y., Lei, T., Yuan, X.: Visual cryptography in single-pixel imaging. arXiv preprint arXiv:1911.05033 (2019)
21. Jin, M., Hirsch, M., Favaro, P.: Learning face deblurring fast and wide. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 745–753 (2018)
22. Khan, S.S., Adarsh, V., Boominathan, V., Tan, J., Veeraraghavan, A., Mitra, K.: Towards photorealistic reconstruction of highly multiplexed lensless images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 7860–7869 (2019)
23. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional camera with a coded aperture. ACM Trans. Graph. (TOG) **26**(3), 70 (2007)
24. Liang, C.K., Lin, T.H., Wong, B.Y., Liu, C., Chen, H.H.: Programmable aperture photography: multiplexed light field acquisition. ACM Trans. Graph. (TOG) **27**, 55 (2008)
25. Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L.: Sphereface: Deep hypersphere embedding for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 212–220 (2017)
26. Nagahara, H., Zhou, C., Watanabe, T., Ishiguro, H., Nayar, S.K.: Programmable aperture camera Using LCoS. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6316, pp. 337–350. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15567-3_25
27. Nguyen Canh, T., Nagahara, H.: Deep compressive sensing for visual privacy protection in flatcam imaging. In: Proceedings of the IEEE International Conference on Computer Vision Workshops (2019)
28. Nikonorov, A.V., Petrov, M., Bibikov, S.A., Kutikova, V.V., Morozov, A., Kazanskii, N.L.: Image restoration in diffractive optical systems using deep learning and deconvolution. Comput. Opt. **41**(6), 875–887 (2017)
29. Nodari, A., Vanetti, M., Gallo, I.: Digital privacy: replacing pedestrians from google street view images. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), pp. 2889–2893. IEEE (2012)
30. Padilla-López, J.R., Chaaraoui, A.A., Flórez-Revuelta, F.: Visual privacy protection methods: a survey. Exp. Syst. Appl. **42**(9), 4177–4195 (2015)
31. Pittaluga, F., Koppal, S.J.: Privacy preserving optics for miniature vision sensors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 314–324 (2015)
32. Pittaluga, F., Koppal, S.J.: Pre-capture privacy for small vision sensors. IEEE Trans. Pattern Anal. Mach. Intell. **39**(11), 2215–2226 (2016)
33. Raskar, R.: Less is more: coded computational photography. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) ACCV 2007. LNCS, vol. 4843, pp. 1–12. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76386-4_1

34. Ren, D., Zhang, K., Wang, Q., Hu, Q., Zuo, W.: Neural blind deconvolution using deep priors. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
35. Ren, D., Zuo, W., Zhang, D., Xu, J., Zhang, L.: Partial deconvolution with inaccurate blur kernel. IEEE Trans. Image Process. **27**(1), 511–524 (2017)
36. Ren, W., et al.: Deep non-blind deconvolution via generalized low-rank approximation. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 31, pp. 297–307. Curran Associates, Inc. (2018). http://papers.nips.cc/paper/7313-deep-non-blind-deconvolution-via-generalized-low-rank-approximation.pdf
37. Schuler, C.J., Christopher Burger, H., Harmeling, S., Scholkopf, B.: A machine learning approach for non-blind image deconvolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1067–1074 (2013)
38. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)
39. Shen, Z., Lai, W.S., Xu, T., Kautz, J., Yang, M.H.: Deep semantic face deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8260–8269 (2018)
40. Sinha, A., Lee, J., Li, S., Barbastathis, G.: Lensless computational imaging through deep learning. Optica **4**(9), 1117–1125 (2017)
41. Sloane, N.J., Harwitt, M.: Hadamard transform optics (1979)
42. Son, H., Lee, S.: Fast non-blind deconvolution via regularized residual networks with long/short skip-connections. In: 2017 IEEE International Conference on Computational Photography (ICCP), pp. 1–10. IEEE (2017)
43. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: a persistent memory network for image restoration. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4539–4547 (2017)
44. Thorpe, C., Li, F., Li, Z., Yu, Z., Saunders, D., Yu, J.: A coprime blur scheme for data security in video surveillance. IEEE Trans. Pattern Anal. Mach. Intell. **35**(12), 3066–3072 (2013)
45. Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., Tumblin, J.: Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. ACM Trans. Graph. (TOG). **26**, 69 (2007)
46. Wang, H., et al.: Cosface: large margin cosine loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5265–5274 (2018)
47. Wang, R., Tao, D.: Training very deep CNNs for general non-blind deconvolution. IEEE Trans. Image Process. **27**(6), 2897–2910 (2018)
48. Wang, Z.W., Vineet, V., Pittaluga, F., Sinha, S.N., Cossairt, O., Bing Kang, S.: Privacy-preserving action recognition using coded aperture videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2019)
49. Xu, L., Ren, J.S., Liu, C., Jia, J.: Deep convolutional neural network for image deconvolution. In: Advances in Neural Information Processing Systems, pp. 1790–1798 (2014)

50. Zhang, K., Xue, W., Zhang, L.: Non-blind image deconvolution using deep dual-pathway rectifier neural network. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2602–2606. IEEE (2017)
51. Zhang, L., Zuo, W.: Image restoration: from sparse and low-rank priors to deep priors [lecture notes]. IEEE Signal Process. Mag. **34**(5), 172–179 (2017)
52. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)