INDUSTRIAL TECHNOLOGY ADVANCES

The future of biometrics technology: from face recognition to related applications

HITOSHI IMAOKA, (iD) HIROSHI HASHIMOTO, KOICHI TAKAHASHI, AKINORI F. EBIHARA, JIANQUAN LIU, AKIHIRO HAYASAKA, YUSUKE MORISHITA AND KAZUYUKI SAKURAI

Biometric recognition technologies have become more important in the modern society due to their convenience with the recent informatization and the dissemination of network services. Among such technologies, face recognition is one of the most convenient and practical because it enables authentication from a distance without requiring any authentication operations manually. As far as we know, face recognition is susceptible to the changes in the appearance of faces due to aging, the surrounding lighting, and posture. There were a number of technical challenges that need to be resolved. Recently, remarkable progress has been made thanks to the advent of deep learning methods. In this position paper, we provide an overview of face recognition technology and introduce its related applications, including face presentation attack detection, gaze estimation, person re-identification and image data mining. We also discuss the research challenges that still need to be addressed and resolved.

I. INTRODUCTION

Unlike using passwords or physical keys, biometrics technology has great potential to usher in a new world where nobody needs to be conscious of authentication or identification processes. In particular, face recognition technology is evolving very rapidly in terms of its recognition accuracy along with the recent advances in deep machine learning, and has attracted much research attention as a promising technology that can simultaneously offer both convenience and precision. The advantages of face recognition technology are threefold: (1) it enables authentication from a distance, (2) it works simply with a universal apparatus like a smartphone or tablet, no longer requiring any special device, and (3) it assures the convenience of the users by complementing confirmation by humans if it unexpectedly stops working, unlike using fingerprint authentication. At the same time, face recognition technology faces various critical challenges in practical implementation, including discrepancy in the face images of an identical person (squinting or shutting eyes, altering facial expression), changes in the face by aging (baby to elderly), facial resemblance (twins or siblings), and accessories concealing a part of the face (eyeglasses or a mask).

NEC Corporation, Minato-ku, Tokyo, Japan

Corresponding author:
Hitoshi Imaoka
Email: h-imaoka_cb@nec.com

To tackle the aforementioned challenges, as a pioneer, NEC has made a series of contributions to face recognition technology from 1989. We developed a 3D and 2D face recognition system in 1996 and 2000, respectively. In 2004, our face recognition technology was incorporated into an immigration administration system, which has since been deployed in 45 countries.

From the technical perspective, our face recognition technology was evolving by adopting major methods of the time in three different stages: (1) distance comparison among feature points (e.g. eyebrows and nose) in 1990, (2) statistical methods such as Eigenface and FisherFace in the 2000s, and (3) recently used methods such as deep machine learning after the 2010s. At the current stage, our face recognition technology also adopts hand-crafted features or lightweight convolutional neural networks (CNNs) to optimize its processing pipeline due to limited computational resources on a device.

In addition, NEC also actively conducts researches on presentation attack detection (PAD), which aims to distinguish live face samples from spoof artifacts, for securing biometrics authentication. How to develop a robust face PAD on smartphones is one of the most important practical issues. From the viewpoint of ensuring biometrics authentication security, various key technologies have already been developed, including secret authentication that enables matching and identification without decryption of feature values, and a cancellable biometrics technique that changes feature values by utilizing both biological features and a secret key [1–3].

Image　　　　Face Location,　　Aligned Face　　Feature Vector　　Face Identification
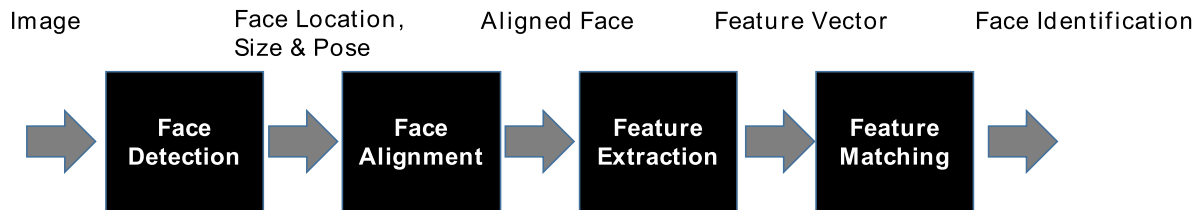　　　　　　　Size & Pose

Fig. 1. Face recognition processing.

We organize the rest of this paper as follows. Section II provides an overview of face recognition technology including face detection, face alignment, face matching as shown in Fig. 1, and reports the recent results of the National Institute for Standards and Technology (NIST) benchmarking regarding face recognition. Section III reports the recent progress of face PAD. Section IV introduces key applications of face recognition including gaze estimation and person re-identification. Following, the use cases in real scenes are introduced in Section V. We conclude this paper and discuss future challenges in Section VI.

II. OVERVIEW OF FACE RECOGNITION TECHNOLOGY

A) Face detection

Face detection technology has two important tasks: determining facial regions in an image against various backgrounds and determining the alignment of each face, such as position, size, and rotation, to obtain a better performance in face-related applications such as face recognition systems. Because this technology is usually utilized in the first step of the applications (Fig. 1), many face detection algorithms have been proposed over the past 20 years. One of the most successful approaches is based on the cascaded structure of the AdaBoost classifiers proposed in 2001 by Viola and Jones [4]. The Viola–Jones algorithm achieved remarkable performance in terms of accuracy and speed for the first time in the history of this technology. The algorithm has also been implemented in various open-source software applications, leading to its extensive use by many researchers in the field of computer vision.

We developed a novel hierarchical scheme combined with face and eye detection in 2005 [5] by using generalized learning vector quantization (GLVQ) [6] as a classifier to improve performance. In the process of face detection, the face position is roughly determined using low-frequency components by searching over multi-scale images. Figure 2 shows the flow of the proposed face detection system. First, multi-scale images are generated from an input image, and then reliability maps are generated by GLVQ. Finally, these maps are merged through interpolation to obtain final results. In the process of eye detection, the positions of both eyes are determined precisely by a coarse-to-fine search using high-frequency components of the image. With this method, we achieved both real-time face detection and precise face alignment, and subsequently have applied this method to many practical applications.

With the development of face-related applications, advanced face detection technology has become increasingly necessary in recent years to detect faces in more difficult situations typified by various head poses, illumination changes, and occlusions such as wearing a surgical mask. The conventional approaches mentioned above are not able to handle such situations due to the limitation of representation capacity of image features and the classifiers they use. Meanwhile, deep learning technology has been applied to generic object detection tasks that focus on common objects found in everyday life. Two major approaches using deep learning are Faster R-CNN [7], the best known two-stage approach, and Single Shot MultiBox Detector [8], the best known for single-stage approach. In general, deep learning-based object detection (including the above methods) consists of two parts: a backbone, which is equivalent to feature extraction, and a detector, which calculates object locations and confidences for each object. In the field of generic object detection using deep learning, many methods that use a deep and large backbone have been proposed, and they have achieved high accuracy. In the cases that algorithms are running on a CPU, not a GPU, various lightweight backbones such as MobileNet [9] and ShuffleNet [10] have been proposed in recent years and are now being applied to generic object detection tasks.

Since the face detection task has already been utilized in many real-life situations, e.g. searching a query face against tens of millions of faces in an image database or analyzing faces from thousands of IP cameras processed on hundreds of servers, a faster algorithm is required to gain competitive business advantage in this field. Therefore, we have developed an original, real-time face detection algorithm using a ResNet-based backbone with a single-shot detector network. In our technology, we trained our model with face images and human body images from our own-captured internal database to output the positions of the faces and human bodies captured in input images. We also applied this technology to person re-identification described in Section IV.B. Due to differences in individual posture and clothing, human bodies can present near-infinite appearance variations, making them much more complex than faces. To deal with this complexity, a massive number of images of the human body – showing an enormous variety of individuals engaged in activities such as walking and running – is input into the system as training data. This makes it possible to ensure the detection of faces and human bodies in various settings, which is shown in Fig. 3. In our algorithm, we carefully tuned the parameters of the backbone network with a focus on CPU processing. As a result,
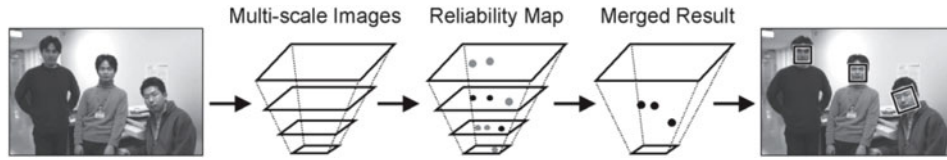
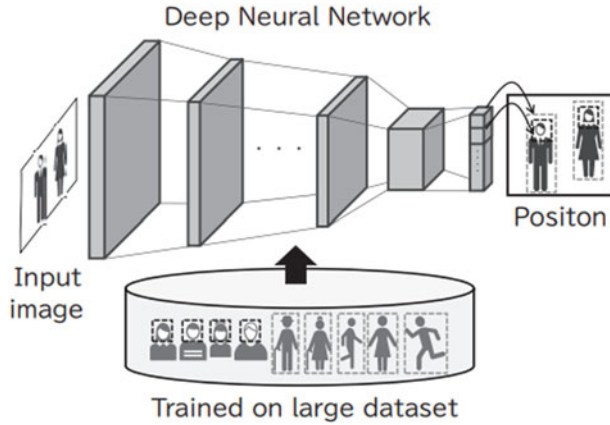**Fig. 2.** Processing flow of the face detection proposed in 2005 [5].

Fig. 3. Processing flow of the face/human body detection technology based on a deep neural network.

Fig. 4. Face alignment for detecting the feature points of facial parts.

the proposed network is able to achieve the throughput of 25 fps on average on a single core of Core i7 when a 2K image is down-sampled by a factor of 2.25 before passing to the network. The network detects faces of size 50×50 pixels in an original 2K image, which is a sufficient condition for practical use. In other words, the proposed network is able to process a 2K image almost in real-time, with high detection accuracy for not only usual faces but also faces in the wild.

B) Face alignment

Figure 4 shows an example of face alignment for detecting the feature points of facial parts, such as the eyes, nose, and mouth. To achieve accurate face recognition, it is crucial to align the position and shape of facial parts precisely, as the face recognition accuracy is affected by facial pose and facial expressions. A robust face alignment algorithm is required, especially for wild face recognition, where there is no limitation on the photographing conditions. Low computational cost is another important concern from the viewpoint of practical application in real environments. Recent face-matching algorithms require a time-consuming large-scale CNN. We therefore aim to reduce the computational cost of face alignment compared with face-matching.

Recent face alignment algorithms are roughly divided into two types: handcrafted feature-based methods and deep learning-based methods.

Regarding handcrafted feature-based methods, cascaded regression models were commonly utilized in the 2010s. Cascaded regression models have multiple stages of handcrafted feature extraction and linear regression. Designing
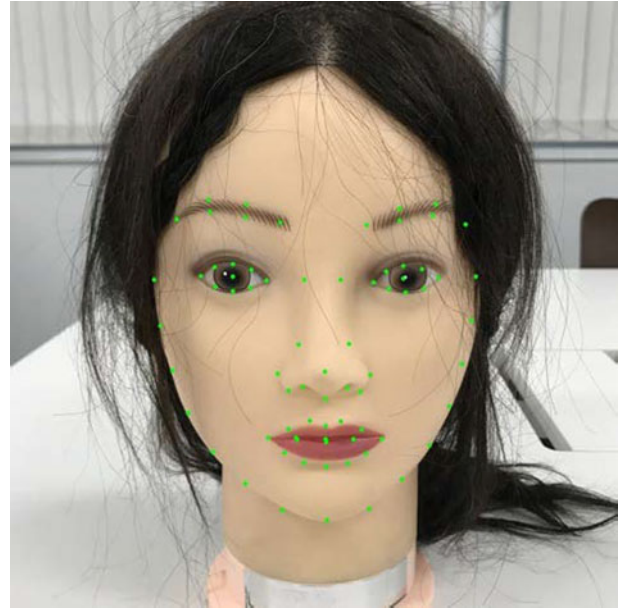
effective handcrafted features is the major issue facing cascaded regression models for fast and accurate face alignment. For example, the size of the feature descriptor affects the face alignment speed and accuracy. The coarse-to-fine strategy, where large descriptors are used in the early stage and small descriptors in the latter, can improve the accuracy, but it also decreases the speed. Thanks to the adoption of histogram of oriented gradients, fast feature extraction of a large descriptor can be achieved by means of the integral images of each gradient [11]. As a result, cascaded regression models have achieved accurate face alignment over 1000 fps on ordinary CPUs.

Deep learning-based methods have been utilized extensively since the late 2010s. In contrast to cascaded regression models, where handcrafted features need to be designed manually, deep learning models automatically learn the effective feature representation for a face alignment task. However, the computational cost of deep learning models is much higher than that of the cascaded regression models. Thus, low computational deep learning models are expected to utilize for lowering the cost. On the other hand, deep learning models are effective for accuracy improvement in difficult situations, such as large occlusions and head poses, thanks to their higher representation performance than the cascaded regression models.

We have two choices for face alignment algorithms: fast cascaded regression models and robust deep learning models. It is important to select an appropriate face alignment algorithm depending on practical situations.

C) Face matching

Face matching technology extracts a feature vector from a face image and identifies whether the person in the image is a pre-registered person. The query image and the registered image are not always shot under the same conditions. Variations of posture or illumination, as well as facial expressions and aging, constitute important factors in matching performance degradation.

In order to solve the problem of pose variation, we developed a face normalization technology using the obtained facial feature points. The face normalization technology corrects the posture to frontal face as well as the position and size of a face image by utilizing a 3D shape model of an average frontal face. For the facial expressions and aging that are difficult to model, we apply a discriminative multi-feature fusion method [12] to extract features that are useful for person identification from a large amount of face image data to reduce the performance degradation. With this method, various features such as edge direction and local textures are extracted from the face image, and the feature vectors are then projected to the feature space that remains unaffected by variation and is effective for person identification. Then, the query image is compared with the registered images on the basis of the angle between the vectors in the feature space. In this way, by utilizing the two different methods, we can achieve high-accuracy face matching that is able to cope with diverse variation factors.

Recently, we have achieved more accurate face matching by using a deep learning-based technology. The normalized face image created by our face normalization method is input to a CNN to extract the optimal features (Fig. 5), necessary to accurately identify an individual. We use a ResNet-based architecture combined with a novel loss function and our original deep metric learning method [13, 14]. This metric learning method is designed to simultaneously minimize intra-class distance and maximize inter-class distance. This makes the system less susceptible to recognition problems caused by partial occlusion, aging, wearing a mask, etc. The CNN trained with the aforementioned methods shows more robust individual identification performance despite changes in appearance.

D) Benchmarking results

In the field of face recognition, especially, differences in evaluation data often lead to a completely different evaluation of recognition accuracy. The Face Recognition Vendor Test (FRVT) conducted by NIST has promoted practical applications of face recognition technology by providing a fair and reliable comparative evaluation of face recognition algorithms. To ensure the fairness and reliability of the test, the NIST defines evaluation prerequisites in a very
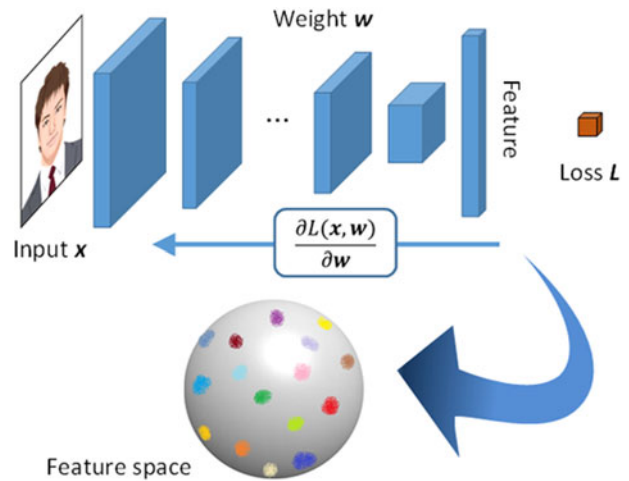

Fig. 5. Feature extraction with a CNN for the face matching.

strict manner and uses a common database that is well-suited for practical application. We participated in such tests ever since we joined the Multiple Biometric Grand Challenge (MBGC) in 2009 [15], and had significant recognition accuracy in many evaluation indices in the Multiple Biometric Evaluation (MBE) [16], FRVT2013 [17], Face in Video Evaluation (FIVE) [18] 2015, and FRVT2018 [19]. In the FRVT2018, specifically, our algorithm achieved the highest accuracy with a false-negative-identification-rate of 0.5% at a false-positive-identification-rate of 0.3% when registering 12 million people. Furthermore, our algorithm showed high robustness in matching the images of a subject taken over 10 years ago and performed the extreme high-speed face matching of 7 ms when registering 1.6 million people.

III. RECENT PROGRESS OF FACE PRESENTATION ATTACK DETECTION

Although it has clear advantages over the conventional authentication systems, face authentication has a major drawback common to other forms of biometric authentication: a nonzero probability of false rejection and false acceptance. While false rejection is less problematic, since a genuine user can usually make a second attempt to be authorized, false acceptance entails a higher security risk. When a false acceptance occurs, the system may actually be under attack by a malicious imposter attempting to break in. Acquiring facial images via social networks is now easier than ever, allowing attackers to execute various attacks using printed photos or recorded video. The demand for technologies for face PAD is thus rising in an effort to ensure the security of sites deploying face recognition systems.

A) Face presentation attack databases

Face presentation attacks can be subdivided into two major categories: 2D attacks and 3D attacks (Fig. 6). The former includes print attacks and video-replay attacks, while the
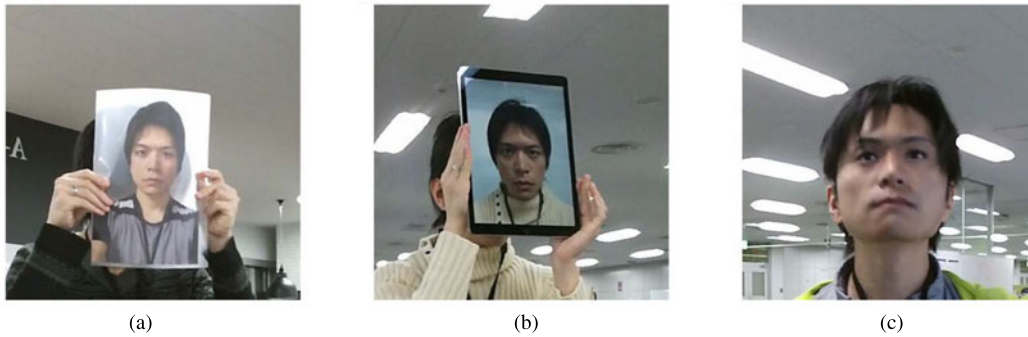
Fig. 6. Example of presentation attack types. (a) 2D print attack. (b) 2D replay attack. (c) 3D spoofing mask attack.

latter includes 3D spoofing mask attacks. Several publicly available databases simulate these attacks. To name a few, the NUAA [20] and Print-Attack [21] databases simulate print attacks. The Replay-Attack [22], CASIA Face Anti-Spoofing [23], MSU Mobile Face Spoofing [24], and Spoofing in the Wild (SiW, [25]) databases contain replay attacks in addition to photo attacks. The 3D Mask Attack Database [26] and HKBU-Mask Attack with Real World Variations [27] simulate 3D mask attacks. Example countermeasures to each attack type are summarized below.

B) Countermeasures to 2D attacks

The 2D attacks, including print and replay attacks, have prominent features in common: the characteristic surface texture and flatness. To use the texture as a key feature, PAD algorithms that utilize a local binary pattern [28, 29] or Gaussian filtering [30, 31] have been proposed. To detect flatness, stereo vision [32] and depth measurement from defocusing [33] are used to detect spoofing attacks.

Infrared imaging can be used to counter replay attacks, as the display emits light only at visible wavelengths (i.e. a face does not appear in an infrared picture taken of a display whereas it does appear in an image of an actual person [34]). Another replay-attack-specific surface property is the moiré pattern [35].

C) Countermeasures to 3D mask attacks

The recent 3D reconstruction and printing technologies have given malicious users the ability to produce realistic spoofing masks [36]. One example countermeasure against such a 3D attack is multispectral imaging. Steiner *et al.* [37] reported the effectiveness of short-wave infrared imaging for detecting masks. Another approach is remote photoplethysmography, which calculates pulse rhythms from periodic changes in face color [38].

D) End-to-end deep neural networks

The advent of deep learning has enabled researchers to construct an end-to-end classifier without having to design an explicit descriptor. Research on face PAD is no exception; that is, deep neural network-based countermeasures have

been found for not only photo attacks but also replay and 3D mask attacks [39–41].

E) Flash-based PAD algorithm for mobile devices

Face recognition systems are being used at places as diverse as airports and office entrances and as the login systems of edge devices. Each site has its own hardware availability; i.e. it may have access to a server that can perform computationally expensive calculations, or it may be equipped with infrared imaging devices. On the other hand, it may only have access to a low-performance CPU. It is thus natural that the suitable face PAD algorithm will differ according to the hardware availability. The advent of deep-learning technologies has enabled high-precision image processing that competes with human abilities at the expense of high computational cost. On the other hand, there is still a need for an efficient PAD algorithm that works with minimal computational resources. Specifically, countermeasures for 2D attacks including photo and display attacks are important because they are more likely to occur than 3D attacks due to their low production cost. In order to prevent the 2D attacks, we recently proposed an efficient face PAD algorithm that requires minimal hardware and only a small database, making it suitable for resource-constrained devices such as mobile phones [42].

Utilizing one monocular visible light camera, our proposed algorithm takes two facial photos, one taken with a flash and the other without a flash. The proposed feature descriptor is constructed by leveraging two types of reflection: (i) specular reflections from the iris region that have a specific intensity distribution depending on liveness, and (ii) diffuse reflections from the entire face region that represents the 3D structure of a subject's face. The descriptor is then classified into either live or spoof face class, using Support Vector Machine (SVM, [43, 44]).

Our tests of the proposed algorithm on three public databases and one in-house database showed that it achieved statistically significantly better accuracy than an end-to-end deep neural network classifier (Fig. 7(a), Table 1). Moreover, the execution speed of the proposed algorithm was approximately six times faster than that of
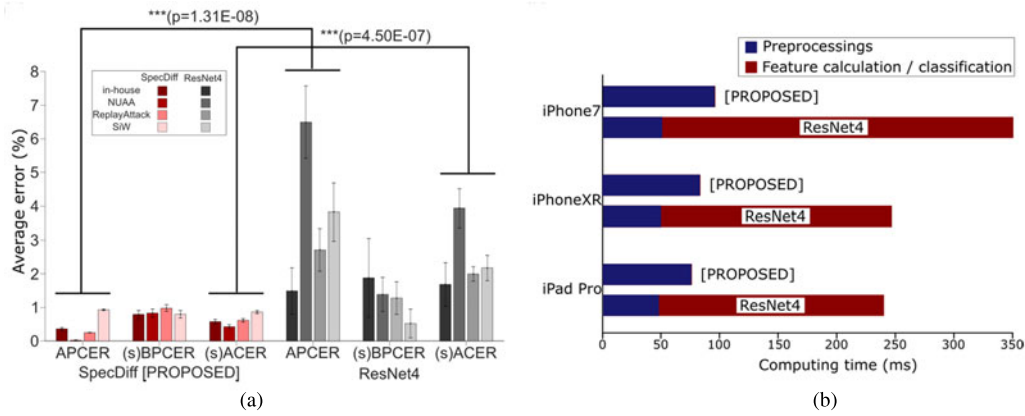
Fig. 7. Speed and accuracy test results, adapted from [42]. (a) Two-way ANOVA comparing SpecDiff and ResNet4. (s)BPCER and (s)ACER indicate that sBPCER and sACER are used for evaluating public databases, while genuine BPCER and ACER are used for evaluating the in-house database. Resulting *p*-values show statistical significance in APCER and (s)ACER. (b) Summary of execution speeds. The proposed SpecDiff descriptor classified with the SVM RBF kernel is compared with ResNet4. Execution speeds on iPhone7, iPhone XR, and iPad Pro are measured.

Table 1. Mean validation errors of selected algorithms.

Descriptor/classifier	In-house		NUAA [20]		Replay-attack [22]		SiW [25]	
	APCER	BPCER	APCER	sBPCER	APCER	sBPCER	APCER	sBPCER
LBP – SVM RBF kernel [45]	1.73	10.53	21.55	16.49	2.11	22.07	3.36	21.95
ResNet4 [19]	1.49	1.88	6.50	1.39	2.71	1.28	3.83	0.52
SpecDiff – SVM RBF kernel [42]	**0.36**	**0.79**	**0.021**	**0.83**	**0.25**	**0.98**	**0.93**	**0.79**

For the experimental details, see [42].
Bold significance (ANOVA) followed by Tukey-Kramer multicomparison test to show that our proposed method achieved statistically significantly better accuracy than the ResNet4.
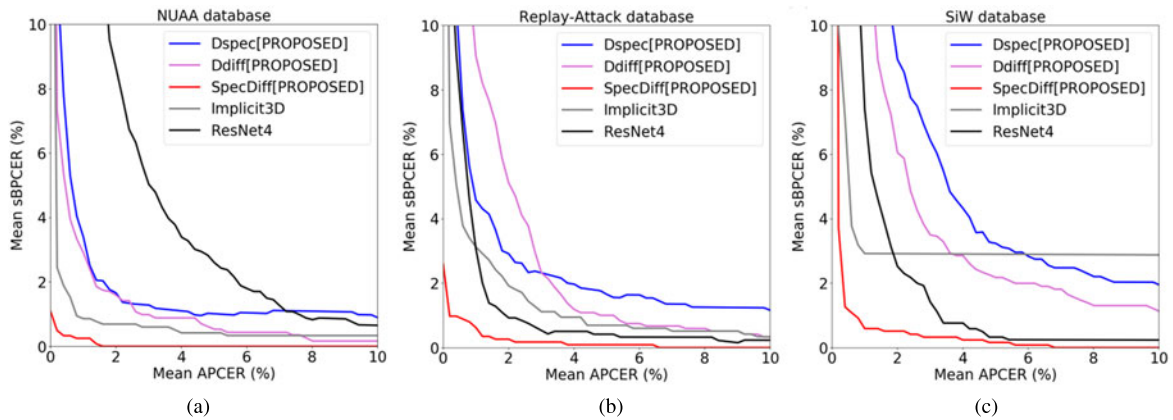
Fig. 8. Average detection error tradeoff (DET) curves across 10-fold cross-validation trials, adapted from [42]. Implicit3D is a different flash-based algorithm [46] included for comparison. (1) NUAA database. (2) Replay-Attack database. (3) SiW database.

the deep neural networks (Fig. 7(b)). The evaluation metrics for these tests were attack presentation classification error rate, bona fide presentation classification error rate (BPCER), and average classification error rate (ACER), following ISO/IEC 30107-3. Note that one problem in evaluation is that our proposed algorithm needs pairs with photos with and without flash. However, we cannot access the live subjects appearing in the public databases. Thus, in order to obtain metrics equivalent to BPCER and ACER, we isolated a part of the live faces of an in-house database from the training dataset and used them as a substitute for live faces of the public databases. Hereafter we refer to these simulated metrics as simulated BPCER (sBPCER) and simulated ACER (sACER) (see Fig. 8)

PAD systems are now an essential part of face authentication systems for secure deployment. To ensure the highest possible accuracy, we consider the following two approaches to be effective for strengthening PAD systems. First, multiple PAD algorithms should be combined. Each PAD algorithm has its own drawbacks, so relying on just one entails security risks. Second, multiple modalities should be used together. Most of the spoofing attacks these days have a similar appearance to live faces in the visible-light domain, thus using only a visible-light-based algorithm may increase risks. For example, the flash-based PAD algorithm mentioned above cannot detect 3D mask attacks. Combining the flash-based algorithm with an infrared-based PAD algorithm will ensure robustness against various spoofing

as well as various environmental conditions (e.g. adverse lighting).

IV. APLLICATIONS OF FACE RECOGNITION

In this section, we introduce the recent progress of key applications adopting face recognition technology with its advantages, including gaze estimation and person re-identification.

A) Gaze estimation

Gaze estimation is one of the amazing applications that can help to capture the users' interests or intents by their eyes. We developed a remote gaze estimation technology (Fig. 9) that enables real-time detection of the direction an individual is looking at, even when using existing cameras from remote locations.

Conventional technologies estimate an individual's line-of-sight using specialized devices equipped with infrared lights and advanced cameras that detect light as it is reflected from an individual's eye. In contrast, our technology uses face alignment, one of the key components of our face recognition, to identify characteristics in and around the eye (e.g. the pupil and the corners of the eye) accurately from images taken by ordinary cameras, including web, surveillance, tablet, and smartphone cameras, without specialized equipment. After face alignment, image features are extracted by a ResNet-based neural network and then an individual's line-of-sight is estimated based on the extracted features.

Since a deep neural network-based gaze estimation method [47] has been proposed so far, we have been seeking a lightweight network to realize real-time processing. We proposed a new formalism of knowledge distillation for regression problems [48]. In this formalism, we made a twofold contribution: (1) a novel teacher outlier rejection loss, which rejects outliers in training samples using teacher model predictions, and (2) a multi-task network. The multi-task network estimates both noise-contaminated training labels and the teacher model's output, the latter of which is expected to modify the noise labels following the memorization effects. Our experiments in [48] showed that the mean absolute error (MAE) of the proposed method using our knowledge distillation was 1.6 in degree on MPIIGaze [47]. In addition, its standard deviation was 0.2. It indicates that the proposed method enables high-accuracy detection of an individual's line-of-sight, with an error of 2.5 degrees or less in most cases. Meanwhile, the MAE of their method [47] was 5.4 in degree for leave-one-person-out evaluation protocol, and 2.5 in degree for person-specific evaluation protocol, respectively. We could not provide a fair comparison with the previous work because we did not follow their evaluation protocol but we split MPIIGaze database into training and test set randomly in [48]. However, it is worth noting that the difficulty of our randomly-split protocol lies between the two evaluation protocols used in [47] and it suggests that our proposed method achieved significantly better accuracy than the previous work. Also, in this paper, we could not provide a comparison between our RGB image-based technology and conventional technologies using IR images because there is no public database with RGB and IR images captured on the same condition.

Furthermore, in our remote gaze estimation technology, by adopting our facial alignment method described in Section II.B, the response to low-resolution images and changes in brightness is enhanced to detect an individual's line-of-sight even when they are separated from the camera by as much as 10 m, as shown in Fig. 9, because our facial alignment method is highly robust even in such case. This enhancement makes our gaze estimation technology well suited for the real-world application of automatically detecting products that draw the attention of shoppers in retail stores. Applying the strengths of this technology, we can analyze the line-of-sight of pedestrians and help to optimize the placement of important announcements on public streets. The technology can also contribute to the safety and the security of our communities by monitoring the eye behavior of suspicious individuals. This is in addition to its business potential, where the technology can help retailers learn more about which products are attracting the most attention from visiting customers.

B) Person re-identification

Subsequently, person re-identification should be counted as another key application of face recognition, which recognizes (or retrieves) the same people from the images taken by non-overlapping cameras. Similar to face-matching processing, person re-identification is utilized to determine whether an individual is the same person in the gallery or not. The difference is that person re-identification uses whole-body images as the basis for identification, rather than just using face images. In this case, an image of an individual's entire body is input to a feature extractor, extracted feature vectors in the gallery images are then compared with those of the individual and finally similarity scores are computed. The similarity score is then used to determine whether the subject is the person in the gallery.

One of the common approaches for person re-identification is to design robust features [49–52]. For example, Liao *et al.* [49] developed the Local Maximal Occurrence (LOMO) features that used the maximum amongst the local histogram bins to handle viewpoint variations. Another technique for solving the person re-identification task is to learn a discriminative metric [49, 53–56]. For example, Li *et al.* [54] proposed the Locally Adaptive Decision Function (LADF) wherein they learn a metric as well as a rule for thresholding. Recently, artificial neural networks have shown excellent performance in many computer vision tasks. In person re-identification task, neural networks also have performed well [57–59]. Xiao *et al.* in [59] learn better deep features by using images from multiple datasets
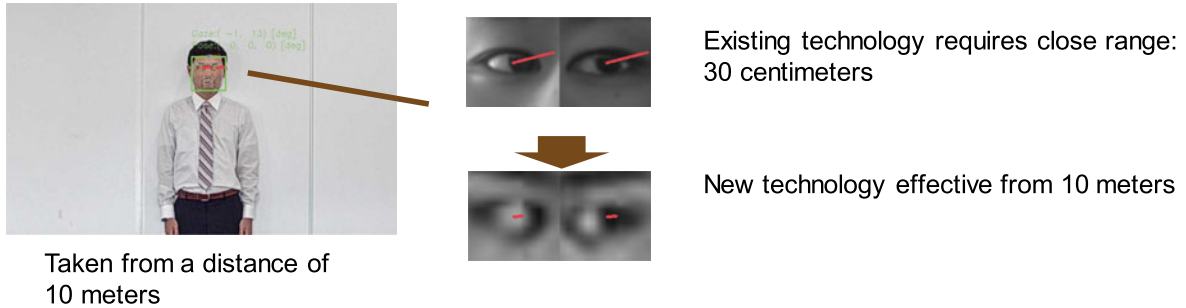
Fig. 9. Outline of remote gaze estimation technology.

Fig. 10. Examples of non-mate image pairs with similar backgrounds. Images are selected from VIPeR [60] dataset.

(domains) and use a new dropout to fine-tune the CNN to a particular dataset.

Similar to most recognition tasks, the accuracy of person re-identification is affected by background, viewpoint, illumination, scale variation, and occlusion. Typically, we focused on the background variation problem. The person re-identification task involves extracting features from the images of people and using a discriminant metric to match the features. The feature extraction process has to be robust enough to deal with background variations. As shown in Fig. 10, the backgrounds in the pair of images are very similar. This often leads to false-positive matching results in non-mate images.

We addressed this problem by using saliency maps in a deterministic dropout scheme to help a CNN learn robust features. We defined a saliency map as the probability of a pixel belonging to the foreground (person) or background. Similar to Simonyan *et al.* [61] computing class saliency maps by back-propagation of a multi-class CNN, we computed the saliency maps of a binary output by CNN. Given the label y of an input image x and a binary classifier $f(x)$, we would like to find an image x_o, such that the score $f(x_o)$ is maximized. According to [61], we can approximate the classifier $f(\cdot)$ by its Taylor expansion as

$$f(x) \approx w^T x + b \qquad (1)$$

where b is a bias term and the weights w is given by equation (2) below.

$$w = \frac{\partial f(x_o)}{\partial x} \qquad (2)$$

From equation (1), it is clear that the contribution of the pixels in x is given by w. By using a ConvNet trained for the binary classification problem (human or not), we can back propagate to the input and get w according to equation (2).

These maps highlight the parts of the image that contribute highly to the score or label of that image and need to be smoothed to contain other portions of the person as well. We combined the saliency map with a CNN using a deterministic dropout technique to enhance the performance. The workflow of the technique, shown in Fig. 11, is divided into two stages. At the first stage, for each input color image, a saliency map is computed. This map is the same size as the input image except it has only one channel. For clarity, the saliency map in Fig. 11 is shown by adding pseudo colors. At the second stage, the color image and its saliency map are input to a CNN that uses the deterministic dropout and outputs the ID of the input image.

To evaluate the effectiveness of our technique, we conducted experiments on a publicly available dataset and measured the contribution of the different components. We mainly compared the performance of three systems:

(1) CNN1 with only color image as input and without the deterministic dropout,
(2) CNN1 with four-channel image as input, i.e. RGB with saliency map, and no deterministic dropout, and
(3) Full CNN with color images as input to CNN1, saliency maps as input to Multi Layered Perceptron, and using the deterministic dropout.

We used the Cumulative Matching Characteristic (CMC) accuracies as the evaluation metric. The performance results are summarized in Table 2, where the first three rows list the various components of our method. We can see that the performance of CNN1 (row 1) was improved when a four-channel image was input instead of a three-channel color image (row 2). This indicates that the saliency information is useful for the re-identification task. Moreover, additional performance gains can be made by using this information in a principled manner with our method (row 3). This suggests that we can make the features learned by a CNN more robust by using the saliency information.
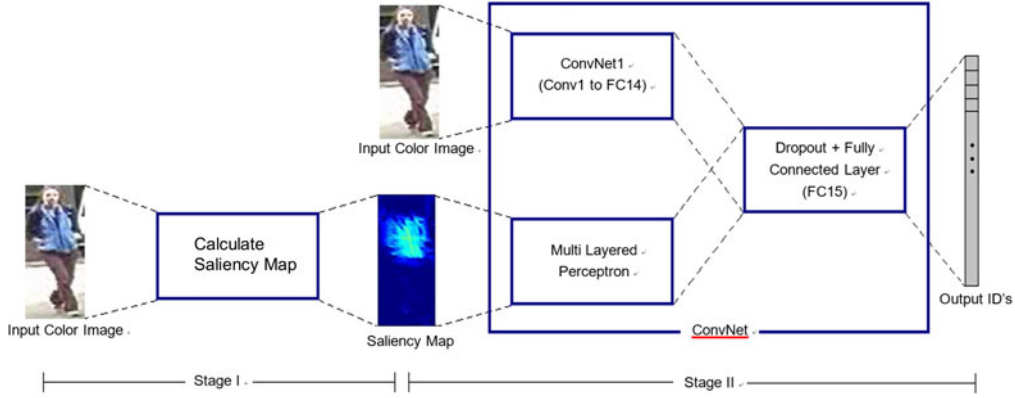
Fig. 11. Workflow of dropout technique. In the first stage, a color image is input and the saliency map is calculated. This map is used in the next stage along with the original image to learn robust features by a CNN. The output is shown as a vector of IDs, but the CNN codes learned in the penultimate layer can be used as the extracted features.

Table 2. CMC accuracies on VIPeR [60].

Method	Rank 1	Rank 10	Rank 20
RGB	37.1	77.6	89.1
RGB + SM	43.8	82.5	92.7
Ours	**49.2**	**89.3**	**96.4**
[62]	45.9	88.9	95.8
[59]	38.6	–	–

"RGB" means only color image is input and "RGB + SM" means a four-channel image is input. "Ours" means color image and its saliency map are input and the deterministic dropout is used.
Bold significance accuracy of our proposed method which includes all of our functions.

V. USE CASES IN REAL SCENES

Thanks to the aforementioned advantages of our facial recognition technology, we are able to apply the technology directly for retrieval or mining applications on large-scale surveillance videos rather than utilizing conventional person-tracking techniques. To demonstrate the feasibility of such a straightforward approach, we introduce three real industrial applications in this section.

First, it is a well-known difficulty that person tracking across multiple cameras often fails to monitor a large area without the camera view overlapping. Since conventional tracking techniques require continuous frames in which the same person appears, it becomes too difficult to recover the tracking of the same person when two frames are from two different cameras without any view overlapping or when images are captured from different angles.

To overcome this difficulty, we completely abandon the conventional tracking techniques and instead perform person retrieval across multiple cameras by utilizing only facial recognition to achieve the person tracking. The key idea is as follows. We first simply apply pairwise matching between every two facial features extracted from multiple camera videos. Then the same person extracted from different and non-continuous frames can be easily connected into a tracking sequence, which has the same effect as person tracking. However, the computation of such pairwise facial matching is costly, as we have the complexity of $O(N^2)$, where

N is the number of facial features. Suppose we have only 10K facial features, although we need 100 million times of pairwise matching. It is obvious that such computation is very inefficient no matter how fast our facial matching might be.

To tackle this problem, we designed a novel indexing method called Luigi [63, 64] to dynamically self-organize a large amount of feature data into a hierarchical tree structure based on the similarity scores between any two facial features. We take a generic approach to build the Luigi tree structure on-the-fly and form similar face groups along the tree traversal path near the leaf levels, as shown in Fig. 12. With this novel approach, the original $O(N^2)$ computational complexity can be maximally reduced to $O(N\log N)$ in ideal cases, which enables person retrieval without conventional tracking in a practical, efficient way.

We adopted the Luigi index in our development of an automated system (AntiLoiter [63] and its visualization VisLoiter [65]) to discover loitering people from long-time multiple surveillance videos. A screenshot of the VisLoiter system is shown in Fig. 13, which depicts the visualized discovery results of loiterers who most frequently appeared in multiple cameras. This loiterer discovery system has been transformed into a real product, named NeoFace image data mining (IDM) [66], for surveillance purposes in the public safety domain.

Second, although this automated system could discover frequent loiterer candidates, it was still far from making a clear decision to detect real loiterers. Therefore, we extended the AntiLoiter system to analyze the characteristics of the appearance patterns of potential loiterer candidates [67]. We developed a novel analytical model by utilizing the mathematical entropy to capture the change of movement, durations, and re-appearances of loiterer candidates, which enables us to understand the common characteristics of behavior patterns regarding true loiterers. As shown in Fig. 14, potential loitering people are likely to appear in similar graph patterns of entropy changes (he_j), such as blue, red, and green curves. For this purpose, we extended VisLoiter [65] to a system called VisLoiter+ [68], shown in Fig. 15, to enhance the visualization of loiterer discovery results.
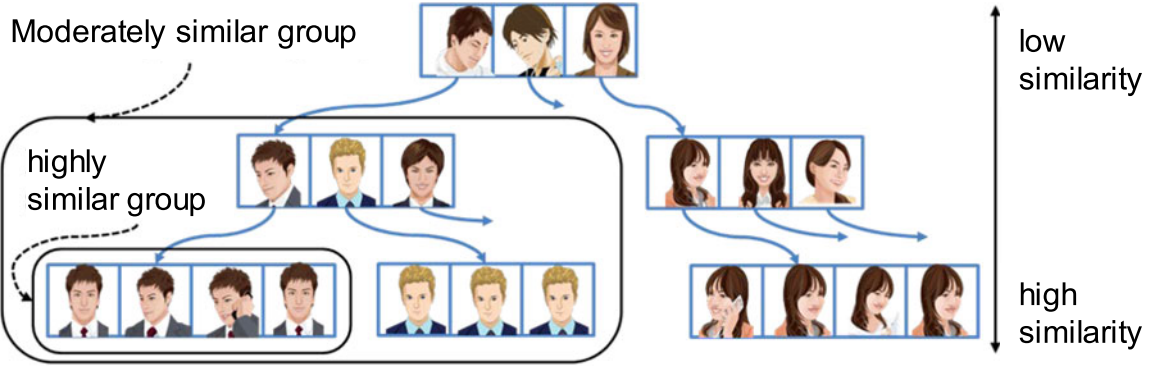
Fig. 12. An example of face grouping by Luigi index [64].

Fig. 13. Visualization results [65] of loiterer candidates discovered by AntiLoiter [63].
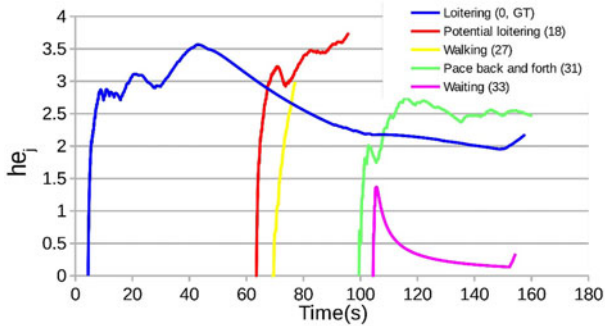

Fig. 14. Examples of behavior patterns regarding potential loiterers.

Third, to demonstrate the feasibility of person retrieval utilizing only face recognition, we developed a novel method to discover "stalker patterns" [69] based on the similar person re-identification and frequent loiterer discovery. Figure 16 shows an example of a stalking scenario [69] in which the same man (marked in a green box) keeps following the same woman (marked in a red box) across different surveillance cameras. The key idea of our approach is to retrieve the same person across videos from different cameras and then to mine the frequently co-appearing patterns of man–woman pairs from the videos in an efficient way,
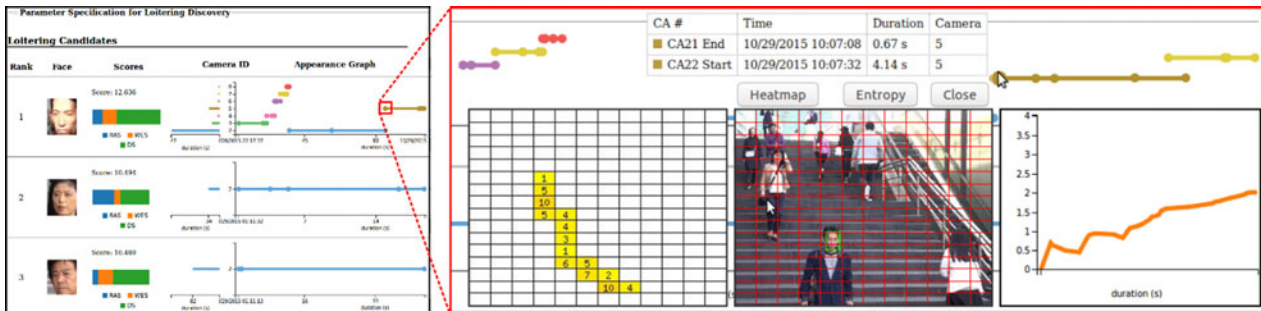

Fig. 15. System VisLoiter+ [68] implemented with the models proposed in [67].

Fig. 16. An example of stalking scenario [69].

such as adopting Luigi to index large numbers of feature data.

In addition to the aforementioned three use cases in real scenes of person retrieval utilizing only our face recognition technology, there are many potentially interesting applications related to group re-identification and group activity analysis. It would prove fruitful for research communities to explore the potential of adopting only face recognition as much as possible for real-world applications in a straightforward yet challenging new direction.

VI. CONCLUSION AND FUTURE CHALLENGES

In this paper, we have provided an overview of face recognition technology, the PAD technology required for practical utilization of face recognition, gaze estimation and person re-identification as an aspect of application technologies, and IDM by time series analysis.

To guide future directions, we suggest the research challenges that still need to be addressed for more practical applications as follows.

(1) Face recognition algorithms suitable for changes to facial appearance throughout life from baby to old age. Lifetime invariance is a crucially important factor for face recognition. In particular, the question of how long face images registered for newborns or children will work is interesting from the viewpoint of technical challenges and limitations.
(2) Countermeasures for shields against face recognition. Although extreme high matching accuracy in normal situation has already been envisaged, improvements are needed for situations where the people to be authenticated are wearing a face mask and/or sunglasses, or when their faces are totally covered by a scarf or beard.
(3) Gaining higher matching accuracy for authenticating twins, siblings, or relatives. This is still a technical challenge [19]. The face similarity of non-mate-pair twins is higher than that of the same person at different ages.

We also suggest to develop a practical, more robust technology for detecting face spoofing and protecting against various kinds of attacks, come up with techniques for template protection of feature values, and develop a protection technology to combat cyber-attacks by artificial intelligence, as well as to combine them with existing face recognition technologies. With the development of such technologies, face recognition will seep into society more widely.

REFERENCES

[1] Bringer, J.; Chabanne, H.; Patey, A.: Privacy-preserving biometric identification using secure multiparty computation: an overview and recent trends. *IEEE Signal Process. Mag.*, **30** (2) (2013), 42–52.

[2] Araki, T.; Furukawa, J.; Lindell, Y.; Nof, A.; Ohara, K.: High-throughput semi-honest secure three-party computation with an honest majority, in *Proc. of the 2016 ACM SIGSAC Conf. on Computer and Communications Security*, 2016, 805–817.

[3] Ratha, N.K.; Connell, J.H.; Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. *IBM Syst. J.*, **40** (3) (2001), 614–634.

[4] Viola, P.; Jones, M.: Robust real-time object detection. *Int. J. Comput. Vis. (IJCV)*, **57** (2) (2004), 137–154.

[5] Sato, A.; Imaoka, H.; Hosoi, T.: Advances in face detection and recognition technologies. *NEC J. Adv. Technol.*, **2** (1) (2005), 28–34.

[6] Sato, A.; Yamada, K.: Generalized learning vector quantization, in *Advances in Neural Information Processing Systems, 8*, MIT Press, 1996, 423–429.

[7] Ren, S.; He, K.; Girshick, R.; Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks, in *Advances in Neural Information Processing Systems, 28*, MIT Press, 2015, 91–99.

[8] Liu, W. *et al.*: SSD: single shot multibox detector, in *Proc. of the 14th European Conf. on Computer Vision (ECCV)*, 2016.

[9] Howard, A.G. *et al.* : MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.

[10] Zhang, X.; Zhou, X.; Lin, M.; Sun, J.: ShuffleNet: an extremely efficient convolutional neural network for mobile devices. arXiv:1707.01083, 2017.

[11] Takahashi, K.; Mitsukura, Y.: Adaptively designed shape regression model for facial point detection. *J. Inst. Image Inf. Telev. Eng.*, **69** (3) (2015), J126–J132.

[12] Imaoka, H.; Hayasaka, A.; Morishita, Y.; Sato, A.; Hiroaki, T.: NEC's face recognition technology and its applications. *NEC Tech. J.*, **5** (3) (2010), 28–33.

[13] Sakurai, K.; Hashimoto, H.; Morishita, Y.; Hayasaka, A.; Imaoka, H.: How face recognition technology and person re-identification technology can help make our world safer and more secure. *NEC Tech. J.*, **13** (2) (2018), 69–73.

[14] He, K.; Zhang, X.; Ren, S.; Sun, J.: Identity mappings in deep residual networks. arXiv:1603.05027 [cs] 2016.

[15] NIST: Multiple Biometric Grand Challenge, MBGC, 2008.

[16] Grother, P.J.; Quinn, G.W.; Phillips, P.J.: MBE 2010: Report on the Evaluation of 2D Still-Image Face Recognition Algorithms, National Institute of Standards and Technology, NISTIR 7709, 2010.

[17] Grother, P.J.; Ngan, M.L.: Face recognition vendor test (FRVT) performance of face identification algorithms, National Institute of Standards and Technology, NISTIR 8009, 2014.

[18] Grother, P.J.; Ngan, M.L.; Quinn, G.W.: Face in video evaluation (FIVE) face recognition of non-cooperative subjects, National Institute of Standards and Technology, NISTIR 8173, 2014.

[19] Grother, P.; Grother, P.; Ngan, M.; Hanaoka, K.: Face recognition vendor test (FRVT) part 2: identification, US Department of Commerce, National Institute of Standards and Technology, NISTIR 8271, 2019.

[20] Tan, X.; Li, Y.; Liu, J.; Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model, in *European Conf. on Computer Vision.* Springer, Berlin, Heidelberg, 2010, 504–517.

[21] Anjos, A.; Marcel, S: Counter-measures to photo attacks in face recognition: a public database and a baseline, in *2011 Int. Joint Conf. on Biometrics (IJCB), IEEE*, 2011, 1–7.

[22] Chingovska, I.; Anjos, A.; Marcel, S: On the effectiveness of local binary patterns in face anti-spoofing, in *2012 BIOSIG-Proc. of the Int. Conf. of Biometrics Special Interest Group (BIOSIG), IEEE*, 2012, 1–7.

[23] Zhang, Z.; Yan, J.; Liu, S.; Lei, Z.; Yi, D.; Li, S.Z.: A face antispoofing database with diverse attacks, in *2012 5th IAPR Int. Conf. on Biometrics (ICB)*, 2012, 26–31.

[24] Wen, D.; Han, H.; Jain, A.K.: Face spoof detection with image distortion analysis. *IEEE Trans. Inf. Forensics Secur.*, **10** (4) (2015), 746–761.

[25] Liu, Y.; Jourabloo, A.; Liu, X.: Learning deep models for face antispoofing: binary or auxiliary supervision, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, 2018, 389–398.

[26] Erdogmus, N.; Marcel, S.: Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect, in *2013 IEEE Sixth Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, 1–6.

[27] Liu, S.; Yang, B.; Yuen, P.C.; Zhao, G.: A 3D mask face anti-spoofing database with real world variations, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016, 100–106.

[28] de Freitas Pereira, T.; Anjos, A.; De Martino, J.M.; Marcel, S.: LBP–TOP based countermeasure against face spoofing attacks, in *Asian Conf. on Computer Vision*, Springer, Berlin, Heidelberg, 2012, 121–132.

[29] Määttä, J.; Hadid, A.; Pietikäinen, M.: Face spoofing detection from single images using texture and local shape analysis. *IET Biometrics*, **1** (1) (2012), 3–10.

[30] Kollreider, K.; Fronthaler, H.; Bigun, J.: Verifying liveness by multiple experts in face biometrics, in *2008 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, IEEE*, 2008, 1–6.

[31] Peixoto, B.; Michelassi, C.; Rocha, A.: Face liveness detection under bad illumination conditions, in *18th IEEE Int. Conf. on Image Processing*, 2011, 3557–3560.

[32] Singh, A.K.; Joshi, P.; Nandi, G.C.: Face liveness detection through face structure analysis. *Int. J. Appl. Pattern Recognit.*, **1** (4) (2014), 338–360.

[33] Kim, S.; Yu, S.; Kim, K.; Ban, Y.; Lee, S.: Face liveness detection using variable focusing, in *2013 Int. Conf. on Biometrics (ICB), IEEE*, 2013, 1–6.

[34] Song, L.; Liu, C.: Face liveness detection based on joint analysis of RGB and near-infrared image of faces. *Electron. Imaging*, **2018** (10) (2018), 373-1–373-6.

[35] Garcia, D.C.; de Queiroz, R.L.: Face-spoofing 2D-detection based on moiré-pattern analysis. *IEEE Trans. Inf. Forensics Secur.*, **10** (4) (2015), 778–786.

[36] Liu, S.Q.; Yuen, P.C.; Li, X.; Zhao, G.: *Recent Progress on Face Presentation Attack Detection of 3D Mask Attacks*, in Handbook of Biometric Anti-Spoofing, Springer, Cham, 2019, 229–246.

[37] Steiner, H.; Kolb, A.; Jung, N.: Reliable face anti-spoofing using multispectral SWIR Imaging, in *2016 Int. Conf. on Biometrics (ICB), IEEE*, 2016, 1–8.

[38] Liu, S.; Yuen, P.C.; Zhang, S.; Zhao, G.: 3D mask face anti-spoofing with remote photoplethysmography, in *European Conf. on Computer Vision*. Springer, Cham, 2016, 85–100.

[39] Yang, J.; Lei, Z.; Li, S.Z.: Learn convolutional neural network for face anti-spoofing. arXiv, vol. abs/1408.5601, 2014.

[40] Menotti, D. *et al.*Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Trans. Inf. Forensics Secur.*, **10** (4) (2015), 864–879.

[41] Nagpal, C.; Dubey, S.R.: A performance evaluation of convolutional neural networks for face anti spoofing. arXiv, vol. abs/1805.04176, 2018.

[42] Ebihara, A.F.; Sakurai, K.; Imaoka, H.: Specular- and diffuse-reflection-based face spoofing detection for mobile devices, in *2020 Int. Joint Conf. on Biometrics (IJCB)*, 2020.

[43] Vapnik, V.; Lerner, A.: Pattern recognition using generalized portrait method. *Autom. Remote Control*, **24** (1963), 774–780.

[44] Chang, C.-C.; Lin, C.-J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, **2** (2011), 27:1–27:27, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[45] Chan, P.P.K. *et al.*: Face liveness detection using a flash against 2D spoofing attack. *IEEE Trans. Inform. Forensic Secur.*, **13** (2) (2018), 521–534.

[46] Di Martino, J.M.; Qiu, Q.; Nagenalli, T.; Sapiro, G.: Liveness detection using implicit 3D features. arXiv: 1804.06702 [cs] 2018.

[47] Zhang, X.; Sugano, Y.; Fritz, M.; Bulling, A.: MPIIGaze: real-world dataset and deep appearance-based gaze estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, **41** (1) (2017), 162–175.

[48] Takamoto, M.; Morishita, Y.; Imaoka, H.: An efficient method of training small models for regression problems with knowledge distillation, in *The 3rd IEEE Int. Conf. on Multimedia Information Processing and Retrieval (MIPR2020)*, 2020.

[49] Liao, S.; Hu, Y.; Zhu, X.; Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2015, 2197–2206.

[50] Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M.: Person re-identification by symmetry-driven accumulation of local features, in *2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, IEEE*, 2010, 2360–2367.

[51] Ma, B.; Su, Y.; Jurie, F.: Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image. Vis. Comput.*, **32** (6–7) (2014), 379–390.

[52] Matsukawa, T.; Okabe, T.; Suzuki, E.; Sato, Y.: Hierarchical Gaussian descriptor for person re-identification, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, 1363–1372.

[53] Prosser, B.J.; Zheng, W.S.; Gong, S.; Xiang, T.; Mary, Q.: Person re-identification by support vector ranking. *BMVC*, **2** (5) (2010), 6.

[54] Li, Z.; Chang, S.; Liang, F.; Huang, T.S.; Cao, L.; Smith, J.R.: Learning locally-adaptive decision functions for person verification, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2013, 3610–3617.

[55] Pedagadi, S; Orwell, J.; Velastin, S.; Boghossian, B.: Local fisher discriminant analysis for pedestrian re-identification, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2013, 3318–3325.

[56] Hirzer, M.; Roth, P.M.; Köstinger, M.; Bischof, H.: Relaxed pairwise learned metric for person re-identification, in *European Conf. on Computer Vision*, Springer, Berlin, Heidelberg, 2012, 780–793.

[57] Li, W.; Zhao, R.; Xiao, T.; Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2014, 152–159.

[58] McLaughlin, N.; Del Rincon, J.M.; Miller, P.: Recurrent convolutional network for video-based person re-identification, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, 1325–1334.

[59] Xiao, T.; Li, H.; Ouyang, W.; Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, 1249–1258.

[60] Gray, D.; Brennan, S.; Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking, in *Proc. of the IEEE Int. Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 3, (5), 2007, 1–7.

[61] Simonyan, K.; Vedaldi, A.; Zisserman, A.: Deep inside convolutional networks: visualising image classification models and saliency maps. arXiv:1312.6034, 2013.

[62] Paisitkriangkrai, S.; Shen, C.; Van Den Hengel, A.: Learning to rank in person re-identification with metric ensembles, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2015, 1846–1855.

[63] Liu, J.; Nishimura, S.; Araki, T.: AntiLoiter: a loitering discovery system for longtime videos across multiple surveillance cameras. *in Proc. of the 24th ACM Int. Conf. on Multimedia*, 2016, 675–679.

[64] Liu, J.; Nishimura, S.; Araki, T.; Nakamura, Y.: A loitering discovery system using efficient similarity search based on similarity hierarchy. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, **100A** (2) (2017), 367–375.

[65] Liu, J.; Nishimura, S.; Araki, T.: VisLoiter: a system to visualize loiterers discovered from surveillance videos. *SIGGRAPH Posters*, **47** (2016), 1–2.

[66] NEC Press Release: NEC launches AI software that searches video for specific individuals. 2016-10-31. https://www.nec.com/en/press/201610/global_20161031_04.html

[67] Sandifort, M.L.; Liu, J.; Nishimura, S.; Hürst, W: An entropy model for loiterer retrieval across multiple surveillance cameras, in *Proc. of the 2018 ACM Int. Conf. on Multimedia Retrieval (ICMR)*, 2018, 309–317.

[68] Sandifort, M.L.; Liu, J.; Nishimura, S.; Hürst, W: VisLoiter+ An entropy model-based loiterer retrieval system with user-friendly interfaces, in *Proc. of the 2018 ACM Int. Conf. on Multimedia Retrieval (ICMR)*, 2018, 505–508.

[69] Liu, J.; Yung, D.; Nishimura, S.; Araki, T.: Stalker retrieval on surveillance videos using spatio-temporal coappearance, in Proc. of the *2019 IEEE Int. Conf. on Multimedia Information Processing and Retrieval (MIPR)*, 2019, 127–134.

Hitoshi Imaoka received the Doctoral degree of Engineering in applied physics from Osaka University in 1997. He has been working at NEC Corporation since 1997 and now serves as NEC Fellow supervising technical issues within the company.

His current research interests include development, research and industrialization of face recognition, biometrics, and medical image processing.

Hiroshi Hashimoto received the B.S. degree from Tokyo Metropolitan University in 2011 and received the Ph.D. degree in physics from Tohoku University in 2016. He now works as a researcher at Biometrics Research Laboratories of NEC Corporation. His research interests include deep learning, computer vision, and biometric authentication.

Koichi Takahashi received the M.S. degree from Tokyo University of Agriculture and Technology in 2012, and the Ph.D. degree from Keio University in 2015. He now works as a researcher at Biometrics Research Laboratories of NEC Corporation. His main research interests include face recognition and its applications.

Akinori F. Ebihara received the B.S. degree in biophysics and biochemistry from the University of Tokyo, Japan, in 2008 and received the Ph.D. degree in biological science from the Rockefeller University, US, in 2015. Currently, he is an Assistant Manager at Biometrics Research Laboratories of NEC Corporation. His research interests include bio-inspired machine learning, sequential probability ratio test, face recognition, and presentation attack detection.

Jianquan Liu received the M.E. and Ph.D. degrees from the University of Tsukuba, Japan, in 2009 and 2012, respectively. He was a Development Engineer in Tencent Inc. from 2005 to 2006, and was a visiting research assistant at the Chinese University of Hong Kong in 2010. He joined NEC Corporation in 2012, and is currently a Principal Researcher at Biometrics Research Laboratories of NEC Corporation. He is also an Adjunct Assistant Professor at Hosei University, Japan. His research interests include multimedia databases, data mining, information retrieval, cloud computing, and social network analysis. Currently, he is/was serving as an Associate Editor of *IEEE MultiMedia* and *Journal of Information Processing*, the General Co-chair of IEEE MIPR 2021, and the PC Co-chair for a series of IEEE conferences including ICME 2020, BigMM 2019, ISM 2018, ICSC 2017, etc. He is a member of IEEE, ACM, IPSJ, APSIPA, and the Database Society of Japan (DBSJ).

Akihiro Hayasaka received the B.S. degree from the Tohoku University, Japan, in 2004 and received the Ph.D. degree in information science in 2009. He now works as a researcher at Biometrics Research Laboratories of NEC Corporation. His main research interests include face recognition and peripheral technologies thereof.

Yusuke Morishita received the M.E. degree from the University of Tsukuba in 2008. Currently, he is a Principal Researcher at Biometrics Research Laboratories of NEC Corporation. His research interests include face detection, pedestrian detection, and gaze estimation.

Kazuyuki Sakurai received the M.E. degree from the University of Tokyo in 1995. Currently, he is a Senior Engineer at Biometrics Research Laboratories of NEC Corporation. His research interests include image recognition, face recognition, person re-identification, and presentation attack detection.