Sample selection for Face Identification System

Napa Sae-bae¹ and Utharn Buranasaksee²

¹Computer Science Department, Faculty of Science, Srinakharinwirot University, Bangkok, Thailand ²Department of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Pranakhonsri Ayutthaya, Thailand napasa@g.swu.ac.th, utharn.b@rmutsb.ac.th

Abstract: The face identification problem is the task of identifying incoming users from their face images. The applications of this task include automating user identification at the building entrance. Consequently, many proposed algorithms are proposed for the given task ranging from handcraft computational models to deep learning models.

This paper utilizes the existing effective algorithms and proposes the template selection strategy to enhance its recognition performance when the enrollment of multiple samples for each user is proposed. Besides, the paper investigates the effect of image quality on the recognition performance of the system and the efficacy of the proposed template selection strategy when applied in such a situation. Experiments are performed on the in-house dataset collected from the Thai population -- to evaluate the empirical system performance when the system is deployed in Thailand to automatically identify the user at the building gate -- as well as on the LFW public dataset consisting of 13,000 face images of 5,749 individuals with multiple ethnicities. The results show that, with the proposed sample selection, the identification error on the in-house dataset for the close-set identification decreased from 3.11 to 2.46% at 0.11% and 0.06% FMR. For the open-set identification, the FNMR decreased from 9.36% to 5.40% at 6.20% and 3.23% FAR, respectively. In addition, the experiments on the LFW dataset have also demonstrated the efficacy of the proposed system consisting of the proposed sample selection method and the selection of face recognition modules. That is, the performance improvement is noticeable as compared to the baseline where the sample selection method is not deployed and to the previous work when multiple samples are used for enrollment.

Keywords: face recognition; sample selection; template selection; image quality

1. Introduction

Generally, a face identification system relies on the same generic configuration which contains 3 steps: face detection, feature extraction, and face matching. Consequently, the performance of a face recognition system depends on the effectiveness and compatibility of three algorithms: face detection, feature extraction, and face matching. Specifically, the use of the Multitask Cascaded Convolutional Neural Network (MTCNN) face detection in conjunction with the face encodings as a feature extraction module, and lastly, the cosine distance between each pair of face images is proven effective against unconstrained face recognition (robust to variations in facial pose, orientation, expression, and environmental condition, e.g., lighting) [1]. This combination could recognize users from their face images by using one face image sample per user for constructing their templates. However, a template with a single face often lacks a generalization ability to capture distributions specific to each user (due to pose and face expression variation) resulting in lower performance [2] [3].

In this paper, we propose the sample selection strategy to construct the template from multiple face images that could be used in conjunction with these standard algorithms to enhance the recognition performance of the system. We proposed that these additional samples shall be selected to ensure that more information can be extracted. Consequently, recognition performance can be enhanced from this additional information. One simple method to ensure that an additional sample could provide more information is that it is distinct enough from the existing ones. That is, during enrollment, the system would acquire multiple face sample images

Received: December 22th, 2021. Accepted: June 9th, 2022 DOI: 10.15676/ijeei.2022.14.2.9 with sufficient distinction between the samples from each user, instead of using a single face image. This is to ensure that variation between the face image samples of the same user is adequately captured by the system. In addition, the paper investigates the effect of image quality in terms of image noise, resolution, and sharpness toward the recognition performance of the system and the efficacy of the proposed template selection strategy when applied in such a situation.

The experiments to demonstrate the efficacy of the proposed approach are conducted using the in-house dataset collected in Thailand to evaluate the performance of the system with the target population as there exist performance differences when deploying the same system on different subgroups with differences in gender and ethnicity [4], [5]. In addition, the evaluations to illustrate the effectiveness of the proposed method are also performed on LFW public dataset.

2. Related work

In this section, we discussed related work in face recognition system components as well as user template quality enhancements for face recognition systems.

A. Face recognition system

Typically, there are two types of face recognition applications: face verification and face identification. For face verification, the system verifies whether a person is the one he or she claims to be. The face verification application includes user-device authentication or to be used as an unlocking mechanism for mobile phones, tablets, etc. Besides, it has been used as an access control mechanism for organizations [6]. For face identification, the system identifies the identity of a person based on a reference database. This system has been used by law enforcement agencies to track down criminals [7]. Besides, it has been used in public surveillance systems [8]. This task can be further divided into two subgroups: the closed-set and the open-set identification depending on whether all the input face images to be identified are from the individuals prior known to the system.

This face recognition system generally relies on the same generic configuration which contains 3 steps: face detection, feature extraction, and face matching. That is, given an image or a video frame, the face recognition system first detects a face or multiple faces that appeared in the image. Then, the set of features is extracted from the face image. The distance between the pair of feature sets is then computed. Lastly, this or multiple of these distances are used to derive a decision of verification or identification system [9].

A face detection task is considered one variant of object detection problems. One primary approach for this task is to use the Haar-cascade algorithm proposed in the Viola-Jones object detection framework which contains four computation steps: Integral Image Representation, Haar-like feature extraction and selection, Adaboost training, and cascading classifiers [10]. Haar-cascade algorithm and its handcraft feature companion (e.g., HoG [11], SURF [12], LBP [13], etc.) while computationally efficient (15 fps at 900MHz CPU), is designed for frontal face detection and it is not robust to variations in facial pose, orientation, occlusion, and illumination [14]. Later, many face detectors using a deep-learning model have been developed. These approaches are more robust to those variations as it is trained on many face images with a various facial pose, orientation, occlusion, and illumination. The MTCNN is one well-known example of this category that uses a cascade Convolution Neural Network (CNN) structure with three stages. Details of this face detector will be discussed later as this is the face detector implemented in this work.

Feature extraction from a face image is one task specific to the face recognition problem. This task has been the main battleground for research in this field. The Eigenfaces algorithm is one of the most well-known feature extraction methods for a face recognition task [15]– [17]. The method computes the eigenvector (so-called eigenface) or the principal components from the distribution of faces in the reference database. Then each face image is decomposed using a subset of these eigenvectors and a set of the eigenvalue is used as a feature vector for that face image. The Fisher faces algorithm is another variation of using the image decomposition

approach [18] – [20]. Instead of using principal component analysis (Eigenfaces) which attempts to minimize the reconstruction error when the components with the smallest eigenvalue are discarded, Fisher's discrimination attempts to derive components that maximize separation between the class. These two are considered holistic or appearance-based approaches. Therefore, the similarity of feature vectors between a pair of face images depends on the similarity of a person's face structure as well as facial appearance. Hence, for the same person variation of feature vector also depends on face orientation and facial expression. Therefore, both the Eigenfaces and Fisherfaces are used in constrained face recognition systems where the users are required to take frontal face images. Another approach for extracting a set of features from a face image is to use CNN. The CNN makes the process of feature extraction for face recognition tasks learnable where it can be trained, and the accuracy is improved as the training data set increases. This approach is also called deep face recognition. The deep face recognition is highly accurate as it yields up to 99% using a large database such as the Deepface by Facebook [21], Facenet by Google [22], VGGFace [23], and VGGFace2 [24], etc. These networks are trained using millions of images. Specifically, VGGFace2 uses the dataset that contains 3.31 million images of 9,131 subjects largely varying in a pose, age, and ethnicity.

Face matching problem is also one derivative of object matching problem where its purpose is to compute the similarity (the opposite of distance function) between two faces [25]. Typically, a pair-wise distance between two face images is defined on the feature vector derived from the two faces. As such, many theoretical grounded distance functions are available to be used. Examples of these distance functions include but are not limited to L-norm distance (L-1 and L-2 in particular), and cosine distance. One widely used L-norm distance function is 2-norm or Euclidean distance. This one is particularly used for PCA-based face features, e.g., Eigenfaces [26]. Manhattan or L-1 distance is another variant of L-norm distance often used for the classification task. In the face recognition system, recognition performances in many tasks between Manhattan or L-1 distance and Euclidean or L-2 distances are only subtle differences [27]. Another relevant distance for the face recognition task is cosine distance. One advantage of the cosine distance function over Euclidean distance is its scale-invariant property since it measures the angular distance between the vectors while discarding the size of the vectors. This is another distance metric often used in a face recognition task, particularly for deep face recognition with Softmax [28]. Lastly, instead of computing a distance on the original feature space, Mahalanobis distance is computed on principal component space [29]. However, to compute this distance in principle component space, the correlation between axes must be learned. As such, the cluster of data points (feature vectors of face images) is needed. The principal component space could be derived individually for each user, or it can be derived globally for all users. The first approach requires that the face template for each user must have an adequate number of enrolled samples. With these many distance functions available, the task of deciding which distance function to be used typically depends on the characteristics of feature vectors retrieved from the feature extraction algorithm.

In this work, the MTCNN face detection, VGG, SENet, and FaceNet feature extraction or face encoding, and cosine distance for face matching are used for a face identification task. This combination has proven effective against unconstrained face recognition (robust to variations in facial pose, orientation, expression, and environmental condition, e.g., lighting) [1].

B. Template quality enhancement

Typically, a face recognition system operates on user templates that are derived from a limited set of samples (e.g., a single image per person or a few samples) [30-32]. As such, the sample quality would have an inevitable impact on the recognition performance of the system. In addition, within user variation may not be captured by the system thereby resulting in decreased recognition performance. The template quality enhancement approaches to mitigate the issues can be categorized into two groups. In the first approach, the system ensures the template quality by validating the quality of samples acquired during the enrollment process [33-

35]. This approach requires the sample quality metric that could derive a quality index of samples that aligns with the recognition performance of a particular template.

The second approach tries to construct a face template that better represents the facial representation variation of each user. In particular, the methods to generate new face patterns or virtual images from a limited set of samples by using the 3D reconstruction method [36], or symmetrical half [37] have been proposed. This approach multiplies a limited set of samples as more photos usually result in more accuracy. Nevertheless, these generated samples may not capture actual facial representation distribution. To resolve the issue, the system could require that each user is enrolled with multiple samples [38]. However, it is possible that those acquired samples are of similar facial angle, pose, and orientation. As such, facial representation variation is not reflected well in those similar samples. To enhance the performance even further, a sample selection approach could be deployed to ensure the quality of the template. Regarding sample selection approaches, in 2010, X. Zhou et al. proposed a method to select samples that are at the boundary of the convex hull to reduce computational complexity without sacrificing recognition accuracy [39]. However, the method focused on rejecting the samples that lay inside the convex hull as they did not help in constructing the decision boundary of the SVM classifier, but it did not mean to ensure the variation of samples. Also, in 2020, F Boutros, et al; proposed that by selecting the most similar probe image from the set of continuously acquired images to the reference images in terms of Ocular Mask Ratio, the performance of the Periocular biometrics recognition system could be enhanced [40]. Again, this method is applied during the recognition process, but it did not apply during the enrollment to ensure that the facial representation variation is adequately captured in the template. In this work, we proposed a sample selection method for a face recognition system to be used during the enrollment process. This proposed method can be viewed as a complementary approach to ensure that enrolled samples of the same subject are enough distinction from each other. In the proposed method, the between-sample distinction is computed from similarity scores, the approach can be applied to any face encoding algorithm. In addition, it could also be applied to other biometric modalities.

3. Face Identification Process

The face identification process is demonstrated in Figure 1. Specifically, to identify the identity of an incoming user, the MTCNN face detector is first performed on a given image to detect a rectangular face area. The detector could return zero or more faces that are found in the image. Then face image is resized to a respective resolution according to the requirement of face encoding schemes: VGG, SENet, FaceNet. The output in this stage is the vector of 2048, 2048, and 512 elements, accordingly. Lastly, this vector is used to compare against all the vectors of the user's faces enrolled in the system to find the best match. That is, during the enrollment, a set of photos are used as the enrollment set of the person. Given an identifying image, the cosine distances between the encoding vector of the identifying image and those of enrolled images are computed. The system than returns the ID of the person that the image in the enrollment has the shortest cosine distance from that identifying image. Details of the three processes are as follows.

A. MTCNN face detection

Multi-task Cascaded Convolutional Network is a face detection module consisting of three cascaded CNN, namely, P-Net, R-Net, and O-Net, respectively. The objective of P-Net or Proposal Network is to obtain the number of candidate face windows. These unlikely windows would then be rejected by R-Net or Refine Network. O-Net or output network then detects used the remaining windows. In this paper, the MTCNN implementation [41] has been used as a face detection module to output face bounding boxes of each input image. Then the input image is then cropped according to the bounding box before proceeding to the VGG Face encoding module [42].

B. Face encoding

Several effective deep learning network-based face encoding schemes are proven effective on various datasets. This work utilizes the following proposed face encoding for the face identification task.



a). Face recognition system with one enrolled sample per user



b). The proposed sample selection process to enroll a single user with multiple samples Figure 1. System configuration of the proposed face identification system

VGG: Given a cropped face image that is resized to the resolution of 224x224, the VGG face encoding output is a 2048-dimensional vector. This original ResNet50-based VGG face encoding is trained from 3.31 million images of 9131 identities where the network is the size of 50 layers [24], [43]. The implementation used in this work is available at https://github.com/rcmalli/keras-vggface.

SENet Squeeze-and-Excitation Network is the improved version of ResNet50 developed by researchers from the University of Oxford. This version is focused on channel-wise operation instead of spatial-wise operation deployed in the original work. The required size of this encoding scheme is 224x224. The encoding output is a 2048-dimensional vector [44]. The

implementation used in this work is also available at https://github.com/rcmalli/kerasvggface.

FaceNet is an efficient face representation developed by Google in 2015 [45]. The implementation used in this work is trained by the MS-Celeb-1M dataset and is available at **https://github.com/nyoki-mtl/keras-facenet** The required size of this encoding scheme is 160x160 where the encoding output is a 512-dimensional vector.

Cosine distance for face matching

Once the feature vector F_i of each face image I_i is computed, the distance between any pair of these feature vectors F_i and F_j could be calculated using the cosine distance function as shown in Equation 1.

$$distance(F_i, F_j) = 1 - \frac{F_i \cdot F_j}{|F_i| |F_j|}$$
(1)

Once the distance between input and all the enrolled user face images are computed using cosine distance, the best match user would be found by the k-nearest neighbor classifier. That is, the k enrolled samples that are closest to the input are selected as the potential match. Then the system determines the identity of the input based on these n samples either by majority vote or weighted vote. Specifically, the special case of a k-nearest neighbor approach where k = 1 is used as a classification rule. That is, the system chooses the nearest enrolled sample and returns the identity associated with that sample as the output.

The proposed sample selection strategy

The simplest form of a face identification system is to enroll each subject with a single face image. Then the system identifies an identity of an input image, or a probe based on the enrolled sample closest to that input. However, it is known that the identification performance of the system could potentially improve as template size or the number of enrolled samples increases. That is, each subject would be required to enroll with multiple face images instead of using a single face image.

This paper proposed that the system performance could improve even better when the system requires each subject to enroll with multiple images that have enough distinction between each other. One way to achieve this is to require that the enrolled template of each subject E_{subj} composes of the set of N images where distances between all pairs of images are at least D (a pre-defined threshold) as shown in Equation 2.

$$E_{subj} = \{I_i, j = 1, \dots, N | distance(I_i, I_j) \ge D \text{ for all } j\}$$

$$\tag{2}$$



Figure 2. Illustration of the proposed template selection strategy benefit

The algorithm to perform the enrollment process using the proposed sample selection algorithm is presented in Algorithm 1. That is, the first sample will be enrolled. For the latter one, the distance between this and the ones already included will be computed. It will be enrolled only if all the distances are at least D. Otherwise, the sample is discarded and a new sample is acquired. The process is finished when N samples are enrolled. Specifically, the enrollment process is finished once the adequate diversity of the samples is captured. The proposed approach is simple yet robust to outliers as the outlier sample would not have an impact on the decision to enroll other samples. Figure 2. demonstrated that the distance between the closest enrolled sample and a genuine input sample could be reduced when the proposed template selection is implemented. This mechanism is to ensure adequate variation between the face image samples of the same user thereby enhancing the diversity of enrolled samples.

Algorithm 1: Sample selection algorithm for enrollment Input: an enrolled sample size N, threshold for sample selection D Output: Enrolled sample set E Initialize: sample input index $i = 0, E = \{\}$ While $|\mathbf{E}| < N$ $I_i \leftarrow$ take a new input image sample from a user If i = 0 $E \leftarrow E \cup \{I_i\}$ Else *isValid* = True For i = 0 to i - 1If $Distance(I_i, I_i) < D$ isValid = FalseBreak If isValid $E \leftarrow E \cup \{I_i\}$ $i \leftarrow i + 1$ Return E

The proposed enrolled sample selection strategy is then compared against the baseline approach which enrolls the first 'N' genuine samples regardless of the pairwise distances between the pairs of enrolled samples (the algorithm to perform the enrollment process using the first 'N' samples is described in Algorithm 2). The result is reported in the Experimental results section using the evaluation protocols and metrics defined in the following section.

5. Methods

In this section, we declare the scope of our experimental settings.

A. Dataset

The experiments were performed on two datasets. One was an in-house dataset, and another

was LFW public dataset. The face images in the in-house dataset were collected from 2 sessions

Algorithm 2: First N sample algorithm for enrollment
Input: an enrolled sample size <i>s</i>
Output: Enrolled sample set E
Initialize: sample input index $i = 0, E = \{\}$
While $ \mathbf{E} < s$
$s_i \leftarrow$ take a new input sample from a user
$E \leftarrow E \cup \{s_i\}$
$i \leftarrow i + 1$
Return E
20 days apart). In total, 35 subjects participated in the first session where the number of image

(20 days apart). In total, 35 subjects participated in the first session where the number of images taken for each user was 6-29. Of these 35 subjects, 15 participated in the second session where the number of images taken for each user was 39-90. There were 1814 images in total. The variation in terms of image samples for each subject was the result of his or her time constraint and pose limitation as the data collection supervisor was instructed to collect images of as many poses as possible.

The second dataset used in our experiment was the Labelled Faces in the Wild (LFW) [46]. This is a public face database collected from websites. Specifically, this dataset contains faces with a high degree of variations in facial orientation and expression as well as environmental factors. This is one of the widely used benchmark datasets for face recognition tasks. It is

composed of more than 13,000 face images of 5,749 individuals. The number of samples for each subject is ranging from 1 to 552.

B. Evaluation protocol

Typically, applications of face identification systems can be divided into the following two settings.

Closed-set identification: In this setting, all users are known to the system. In other words, the system could try to find the best match between the template and the challenge image and there is no need to reject any sample.

Open-set identification: In this setting, not all the users are known to the system. Therefore, the system needs to decide whether to reject or recognize him or her as one of the enrolled subjects.

C. Identification performance metric

The three main performance metrics that were used to evaluate biometric identification systems were the following.

FNMR (False Non-Match Rate): The rate at which the system has incorrectly unrecognized a genuine attempt from an authorized user.

FMR (False Match Rate): The rate at which the system has incorrectly recognized an imposter attempt as an authorized user.

FAR (False alarm Rate): The rate at which the system has incorrectly accepted an imposter attempt. This rate is only applied to the open-set identification.

Note that, the number of samples taken from each user could be different. Thus, if these rates were computed from all samples basis, the users with a greater number of samples would have a larger influence on reported identification performance. Therefore, to avoid such bias, the following steps were taken to compute the rate of the system First, all the rates were computed individually for each user. Secondly, the average rate from all the users was then reported. Besides, FAR was applied to the open-set identification task only. This is because, in the closed-set identification, the system would first decide whether to reject a sample as an intruder according to a pre-defined threshold. If the sample is not rejected, the system then recognizes the sample as one of the enrolled or authorized users. Therefore, for open-set identification, all these rates are threshold-dependent.

6. Results

In this section, the experimental results were reported when the proposed method was applied to the collected dataset as well as the LFW public dataset.

First, the performance of the proposed method on the collected dataset was evaluated. That is, the identification performance of the system when the samples were chosen using the proposed method as compared to the first N samples was reported. Then, identification performance of the system when the image quality was disturbed by environmental factors was reported. In addition, the efficacy of the proposed sample selection method in mitigating the degradation of the system performance due to those image quality defects was demonstrated. In this case, only the test images are degraded as in a realistic setting the quality of enrolled image samples can be controlled but not the test images. Next, the benefit of the proposed sample selection method when it is used in conjunction with other face encoding schemes was illustrated. Note that, in this dataset, 15 subjects who participated in both sessions are selected as enrolled users. The samples of these subjects from the first session were used as training samples (for enrollment) and the samples from all the rest 18 subjects were used as imposter samples. Note that, in close-set identification, test samples are only the samples drawn from enrolled users.

Lastly, the performance evaluation of the proposed method on the LFW dataset and the comparison with other existed works were conducted.

A. Identification system performance

The performance of the close-set and open-set identification system when the first 'N' samples of each subject were used as enrolled samples along with the performance result of the system when the proposed sample selection method was used to select 3 enrolled samples were reported in Table 1 (the distance threshold D in this experiment was set at 0.18). The best performance for the first 'N' samples was when each subject was enrolled with a maximum of five images whereas there are subtle performance differences when subjects were enrolled with the first 1 or 3 samples. However, the performance enhances when the proposed sample selection was applied as compared to the system when the first 'N' samples were used in all metrics and both open-set and closed-set identification even with the smaller number of enrolled samples (3 selective samples versus the first 5 samples). Specifically, for the system with VGG face encoding, the best identification rate was found at 2.46% FNMR at 0.10% FMR and 5.40% FNMR at 3.23%FAR for the closed-set and the open-set identification task, respectively, when 3 samples were selected using the proposed method where the reported performance was computed from 10 users with 3 enrolled samples each since the rest 5 users had only 1 or 2 enrolled samples.

Note that, as mentioned earlier in section 4.3 identification performance metric, there is no rejection decision in closed-set identification, therefore, no threshold is set in the system. However, for open-set identification, not all the incoming samples shall be recognized (some users are unknown to the system). As a result, the threshold is set to reject the samples from those unknown users. The result for open-set identification was reported where the threshold wsas selected at the EER (Equal Error Rate) between FAR and FRR.

	Closed-se	et	Open-se	et		
# enrolled samples	FMR	FNMR	FMR	FNMR	FAR	
First 'N' sample enrollmen	nt					
1	0.11	3.11	0	9.38	6.41	
3	0.11	3.11	0	9.36	6.20	
5	0.10	2.97	0.04	7.02	4.79	
Template selection strategies (3 samples enrollment) (threshold $= 0.18$)						
3	0.10	2.46	0	5.40	3.23	

Table 1. Identification performance for N samples enrollment on VGG face encoding

B. Effect of environmental factors

The system could perform generally well when input images were of good quality [47]. However, it is known the identification performance of the system could be degraded when images are of low quality due to the increase in the distance between the two images of the same identity [48]. Specifically, two important factors that could affect image quality when it is deployed as an automated user identification system are the low resolution and noise disturbance. These are caused by user-camera distance and environmental factors respectively. In addition, the sharpness of the image could be varied due to the image focus issue. The experiments in this subsection were performed to investigate the effect of image quality towards recognition performance of the system and the efficacy of the proposed sample selection method when applied in such a situation to mitigate the performance degradation caused by such image quality. Note that, in this experiment, only the test images were degraded as in a realistic setting the quality of enrolled image samples can be controlled but not the test images.

In the first part of this experiment, each original test image was first corrupted by gaussian noise at sigma = 0 (noise-free image), 5, 10, and 20 and was then resized to 32x32, 64x64, 128x128, and 256x256 resolution respectively before it proceeded to the VGG encoding module. Examples of these modified images with different noise levels and resolutions are illustrated in Figure 2. Note that the size of 224x224 was the required size of the VGG face encoding. The performance results for the closed-set identification system and the open-set identification were reported in Table 2-3. It was seen that FNMR for the system that deployed the proposed sample

selection was lower than that used the first three samples for enrollment in all tests except when images of 32x32 resolution were corrupted with noise sigma at 10 and when images of 64x64 resolution were corrupted with noise sigma at 20. The test images in these two tests were considered as very low-quality ones (as can be visualized from Figure 3). Specifically, for the tests when noise level less than 10 and all image resolution greater than 64, the FNMR reduced by 23 % on average. For the open-set identification which was a more challenging task as the input image could be from an unknown individual, it was observed that the recognition performance improved greatly in all tests when the proposed sample selection was deployed instead of using the first three samples. Specifically, the FNMR was reduced by 41.12% on average. For noise-free images with 256x256 resolution, the performance of the system improves from 10.96% FNMR at 6.93% FAR to 4.55% FNMR at 3.28% FAR.

In terms of image quality, performance was noticeably degraded when the noise level is at sigma 20 or the face image was at the lowest resolution (32x32). Particularly, in these two settings, the recognition performance for the close-set identification did not improve when deploying the proposed method. That is, for the images with strong noise signals, the distance between each pair of face images reflects not only the facial variation but also noise correlation.



Noise-free Images at 32x32, 64x64, 128x128, and 256x256 resolution



Noisy Images with sigma = 5 at 32x32, 64x64, 128x128, and 256x256 resolution



Noisy Images with sigma = 10 at 32x32, 64x64, 128x128, and 256x256 resolution



Noisy Images with sigma = 20 at 32x32, 64x64, 128x128, and 256x256 resolution Figure 3. Illustration of face images with different noise levels and resolutions

Subsequently, the system could select the set of face images with low noise correlation instead of the set with highly discrete facial attributes. Similarly, for the image with low resolution, image details that are particularly useful for the recognition part are discarded as it located in the high-resolution part. As a result, these details cannot be utilized when the system selects the set of face images. Therefore, for the closed-set identification where the performance room is limited (the system already achieved greater than 94% accuracy), the advantage of the proposed method is diminished. This inferred that the camera shall be installed in good lighting conditions to avoid getting face images with strong noise signals. In addition, the system requires the face image (from the MTCNN module) to have at least 64x64 resolution to get reasonable identification performance (the face image resolution is directly related to the distance between the camera and the person).

Lastly, the effect of image focus towards identification performance and the efficacy of the proposed sample selection strategy when applied in such a situation was investigated. In this experiment, a focal blur of the original image was modeled using a Gaussian blur kernel on the face images with the size of 3x3, 5x5, and 7x7, respectively, before resizing it to the required size (224x224) of the VGG face encoding module. Specifically, only the test images were affected by these blurring artifacts as in a realistic setting the quality of enrolled image samples can be controlled but not the test images. Examples of these modified images with different kernel sizes were illustrated in Figure 4.



Original (b) 3X3 (c) 5X5 (d) 7X7 Figure 4. Illustration of face images with different Gaussian blurring kernel sizes

	contapt	cu og u Ouu	Soluli olul Ke	linei		
Kernel size	3x3		5x5		7x7	
Closed-set identification	on					
Enrollment	FNMR	FMR	FNMR	FMR	FNMR	FMR
First 3 samples	3.10	0.09	4.07	0.18	6.04	0.46
Propose method	2.01	0.12	3.42	0.19	5.50	0.48
Open-set identification						
Enrollment	FNMR	FAR	FNMR	FAR	FNMR	FAR
First 3 samples	11.83	7.98	17.29	13.77	26.96	27.11
Propose method	6.49	4.74	11.47	9.70	22.12	22.00

Table 4. Performance for 3 samples enrollment when test samples are corrupted by a Gaussian blur kernel

The performance results were reported in Table 4. It was seen that the closed set identification performance for the system that enrolled the first three images per subject dropped from 3.10 % to 6.04 %FNMR as the kernel size increased from 3X3 to 7X7 (compared to 2.46% FNMR when face images were without blurring artifact). The performance was enhanced from 3.10 to 2.01 %FNMR for the kernel size of 3X3 when the proposed sample selection was applied. For the kernel size of 5x5 and 7x7, a similar improvement was also observed. For the open-set identification task, when the proposed sample selection was applied, the performance was enhanced from 11.83 % to 6.49%FNMR at the kernel size of 3x3 and it deteriorated to 22.00% FNMR (compared to 27.11% when the sample selection was not applied) when the kernel size

increased to 7x7. This result demonstrated the benefit of the proposed sample selection method when images were disturbed by all three image distortion factors: image resolution, noise, and blurring artifacts.

C. Performance on other face encoding approaches

This experiment ensured the benefit of the proposed method when it was used in conjunction with other deep learning face encoding schemes with different network configurations, were trained on a different number of samples and produced a different number of output elements. Specifically, the SENet50 and FaceNet are the other two face encoding schemes that were used in this experiment in addition to VGG-16 (ResNet50). This experiment was also performed on the in-house dataset where the training samples were from the first session and test samples were from the second session. The experiment for the proposed sample selection method was conducted at the threshold where 10 subjects were with 3 enrolled samples (the other five subjects may have only 1 or 2 samples with enough distinction between samples.). According to the result reported in Table 5, VGG and SENet50 in this experiment were the best-performing face encoding schemes whereas the worst performing one was FaceNet. Specifically, the FNMR of VGG-face encoding for the close-set identification decreased from 3.11 to 2.46% at 0.11% and 0.06% FMR, respectively, when the enrolled samples were selected using the proposed method. For the open-set identification, the FNMR decreased from 9.36% to 5.40% at 6.20% and 3.23% FAR, respectively. Nevertheless, the performance improvement was observed in all three face encoding schemes deployed in this work except for the open-set identification task with FaceNet encoding.

Ident. task	Sample selection	Performance	VGG-16	SENet50	FaceNet
Closed-set	The first three samples	FNMR	3.11	2.88	16.79
		FMR	0.11	0.14	1.51
	The proposed method	FNMR	2.46	2.70	13.98
		FMR	0.06	0.10	1.42
Open-set	The first three samples	FNMR	9.36	7.73	29.75
		FMR	0	0	0.45
		FAR	6.20	4.79	18.93
	The proposed method	FNMR	5.40	4.93	31.47
		FMR	0	0	0.41
		FAR	3.23	3.28	20.28

 Table 5. Performance of the proposed method in conjunction with face encoding schemes on the in-house dataset

D. Evaluation on the public dataset

In this experiment, the evaluation of the proposed sample selection was performed on the public dataset, Labelled Faces in the Wild (LFW) using 158 subjects who have more than 10 samples per each (with a total of 4,324 images) to evaluate the efficacy of the proposed sample selection method. Specifically, the subjects with at least 16 samples were identified as known identities (85 subjects). In the open-set identification, the samples of remaining subjects were used as imposter samples or samples from subjects that have not enrolled in the system. Of these 16 samples, 10 samples were used as a training set and the rest are used as a genuine test set. Enrolled samples were then selected from the training set according to the selection strategies. As samples in this dataset were unconstrained face images downloaded from the internet, they were not sorted in chronological order. It was not possible to assign samples taken before as a training set and the ones taken later as a test set. Therefore, a random selection process was performed for samples assignment, and the performance statistics were reported where the random process was repeated 100 times. Note that the reported FNMR and FMR were the average rates of the subjects with 3 samples enrollment for the proposed sample selection

method. The distribution of FNMR for the close-set and the open-set identification task of VGG encoding were depicted in Figure 5. In addition, the performance results for closed and open set



a). Close-set identification
 b). Open-set identification
 Figure 5. Estimation of the probability of FNMR of VGG encoding for the first three samples approach and the proposed sample selection approach at threshold 0.34

Identification tasks were reported in Table 6 along with the benchmark performance when all 10 samples were used in enrollment dataset. Note that, using the first N samples for enrolment is a commonly used process in biometric recognition systems [49,50]. Specifically, the higher N typically results in the higher recognition performance as additional samples provide clues on intra-user variation. Particularly, the proposed sample selection method is designed to complement this intra-user variation information on the regular enrolment process. Specifically, the proposed sample selection could enhance the identification performance from 8.62% to 7.53% FNMR for open-set identification at 3.94 % FAR in VGG face encoding and from 2.77% to 2.48% FNMR for closed-set identification. The improvement could also be observed in SENet50 and FaceNet encodings. However, these two encodings performed worse than the VGG encoding in this case. Note that, the average number of subjects with 3 samples enrollment for the proposed sample selection method for VGG, SENet50, and FaceNet were 42.15, 40.76, and 39.36 with the threshold set at 0.34, 0.28, and 0.34 respectively.

It is important to highlight that, the threshold in the above experiments was set to ensure that the template contains samples with adequate variation between them. The higher threshold could result in a longer enrollment process as it would be harder to find the samples that satisfy the condition. On the other hand, the smaller threshold set could result in selected samples being too similar, and thus less information is provided by additional samples. As a result, the performance improvement is not optimized. In addition, for the range of defined threshold, there is a trade-off between verification and first rank identification performance. The focus of this work is to investigate the efficacy of the proposed method in different environmental scenarios. Therefore, the threshold is set empirically. The question of how to optimize this threshold and whether the proposed method performs well on other compositions of face recognition modules is warranted further investigation.

E. Performance comparison with the existing system

The performance comparison of different face identification systems for the closed-set and the open-set identification were reported in Table 7 and Table 8, respectively. For the closed-set identification, the baseline face identification algorithm used in this paper which consisted of the MTCNN face detector and the VGG face encoder achieved impressive accuracy of 93.24% (the average class accuracy between all subjects). By adding two more samples to the template the accuracy increased to 97.30%. And with the control of discrepancy guarantee between enrolled

samples, the accuracy reached 97.70%. This performance was higher than previous work even when compared to the RPL(L2) system [32] where 2 to 3 samples were used for each user in enrollment. With the open-set identification task, accuracy for the baseline face identification algorithm dropped to 91.41%. Note that, the accuracy of this work is reported in terms of the average between detection identification rate and the opposite of false alarm rate where rejection threshold was set at which false acceptance rate and false alarm rate are equal. The same system reached 93.66% accuracy as compared to 88% accuracy of the OSSR+CRC system [36] when the first 3 samples were used for enrollment. Moreover, the accuracy reached 94.28% when applying the proposed sample selection. Note that the performance improvement seen in this experiment was lower than in the previous experiment as the variation between samples in this dataset was already high. Nevertheless, a noticeable performance improvement was observed for open-set identification. Lastly, it is important to note that, how the dataset is used in identification protocol should be considered when comparing the performance result of different systems. (In this work, we do the best we can to compare our results with the work that has the closest identification protocol to ours.

off the LT w dataset					
Ident. task	Sample selection	Performance	VGG	SENet50	FaceNet
Close-set	The first three samples	FNMR(%)	2.77	2.43	3.56
	The proposed method		2.48	2.21	3.45
	All 10 samples benchmark		1.88	1.57	2.23
	The first three samples	FMR(%)	0.02	0.02	0.04
	The proposed method		0.05	0.03	0.09
	All 10 samples benchmark		0.02	0.02	0.03
Open-set	The first three samples	FNMR(%)	8.62	9.47	13.69
-	The proposed method		7.53	7.96	12.52
	All 10 samples benchmark		5.44	6.38	8.82
	The first three samples	FMR(%)	0.02	0.02	0.03
	The proposed method		0.03	0.03	0.06
	All 10 samples benchmark		0.01	0.01	0.02
	The first three samples	FAR(%)	3.94	4.22	5.75
	The proposed method		3.95	4.11	6.01
	All 10 samples benchmark		2.89	3.26	4.28

 Table 6. Performance of the proposed method in conjunction with each face encoding scheme on the LFW dataset

Table 7. Closed-set identification performance comparison on LFW dataset

Method, Year	Description of data usage	Acc.
SSFR [30], 2020	Images from the first 50 subjects out of 158 subjects, of which	38.01%
	contained more than 10 samples per subject, were used in this	
	experiment. One random sample per subject is used as the	
	gallery set and the remaining images were used as the probe	
	set.	
TDL [31], 2018	Images from the selective 50 subjects out of 158 subjects, of	74.00%
	which contained more than 10 samples per subject, were used	
	in this experiment. One random sample per subject is used as	
	the gallery set and the remaining images were used as the	
	probe set.	
RPL(L2) [32],	Images from artificially selected 478 subjects, of which have	91.74%
2017	3 or 4 images per person, were used in this experiment. For	
	each subject, 1 randomly selected image was selected for the	
	gallery, 1 for testing. The rest were used for training.	

Method, Year	Description of data usage	Acc.
MTCNN+ VGG,	Images from 158 subjects of which contained more than 10	93.24%
(1 sample	samples, were used in this experiment. For each subject, the	
training)	first image was used for the gallery. The rest were used for	
	testing.	
MTCNN+ VGG	Images from 158 subjects of which contained more than 10	97.30%
(3 sample	samples, were used in this experiment. For each subject, the	
training)	first 3 images were used for the gallery. The rest were used	
	for testing.	
MTCNN+VGG	Images from 85 subjects of which contained more than 16	97 .70 %
with the	samples, were used in this experiment. For each subject, the 3	
proposed sample	images selected from 10 images in the training set were used	
selection	for the gallery. The rest 6 images were used for testing. The	
	random training selection process was repeated 100 times	

Table 8. Open-set identification performance comparison on LFW dataset

Method, Year	Description of data usage	Acc.
OSSR+CRC [36], 2017	50% of subjects are included in a gallery with one image per subject.	83.05%
COTRESNET50 [51], 2020	All subjects are divided into 3 subsets. 602 subjects, of which contained more than 10 samples, were assigned to the known subset. For these subjects, the first 3 images were used as the known gallery set and the remaining images were used as the probe set. 1,070 subjects which contained 2 to 3 samples, were assigned to the known-unknown subset. For these subjects, the first image was used as the unknown gallery set and the remaining images were used as the probe set. The rest 4,096 subjects with a single image were used for the transmission.	88%
MTCNN+ VGG	Images from 158 subjects of which contained more than 10 samples, were used. The first 50 subjects were used as probe and gallery, and the images of the remaining 108 subjects were used as a generic set. For each of the 50 subjects, the first image was used as the gallery set and the remaining images were used as the probe set.	91.47%
MTCNN+ VGG	Images from 158 subjects of which contained more than 10 samples, were used. The first 50 subjects were used as probe and gallery, and the images of the remaining 108 subjects were used as a generic set. For each of the 50 subjects, the first 3 images per identity were selected for the gallery. The rest were used for testing.	91.31 %
MTCNN+ VGG with the proposed sample selection	Images from 158 subjects of which contained more than 10 samples, were used in this experiment. The first 50 subjects were used as probe and gallery, and the images of the remaining 108 subjects were used as a generic set. For each of the 50 subjects, the selective 3 images were selected for the gallery. The rest were used for testing.	93.66%
MTCNN+ VGG with the proposed sample selection	Images from 158 subjects of which contained more than 10 samples, were used in this experiment. 85 subjects with more than 16 samples were used as probe and gallery, and the images of the remaining subjects were used as a generic set. For each subject, the 3 images selected from 10 images in the training set were used for the gallery. The rest 6 images were used for testing. The random training selection process was repeated 100 times.	94.28%

7. Discussion and future work

This paper proposed the template selection strategy to enhance the recognition performance of the system when multiple samples are enrolled for each user. The results implied that noise disturbance and user-device distance could make a great impact to face recognition performance. These two are very common in the application of an automated user identification system at the building gate. Therefore, the mechanism to cope with these factors is important to mitigate the performance degradation issue. In this paper, the performance enhancement for such a situation when the system incorporates the proposed template selection strategy had been demonstrated. The result suggested that the proposed mechanism could be particularly useful for one session enrollment as the variety of samples in the template could be ensured.

Besides, the results suggested that the camera shall be installed in a good lighting condition to avoid getting images with a high level of noise and that the system requires the face image (from the MTCNN module) to have at least 64x64 resolution to get reasonable identification performance. In addition, as the system is not robust to out-of-focus face images, image deblurring techniques could be applied to enhance the sharpness of images thereby improving the identification performance. Further, other deep learning face encoding such as Google's FaceNet, Facebook's DeepFace, CMU's OpenFace, could also benefit from the proposed method as demonstrated in the experiment.

While previous work on the sample selection approach is to ensure the quality of sample due to signal degradation [40], the proposed sample selection method is designed to complement intra-user variation information on the regular enrollment process (enrolling the first N samples). These two approaches could be combined to construct the template with good quality samples as well as provide good coverage of intra-user variation. In addition, the more advanced clustering analysis like the Silhouette technique for sample selection would be an interesting study on a dataset (possibly other biometric modality) that contains a larger number of samples per user, e.g., EEG signal. This is one area for future work.

One limitation of this work is the size of the dataset from the target population used in this experiment. Evaluating the recognition performance of the system on a much larger and more realistic dataset is necessary to ensure the efficacy of the proposed sample selection mechanism. Lastly, it is also possible to incorporate a gait recognition module to enhance the focus application's identification performance.

Funding: This research was funded by the RMUTSB research funding program.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Research Ethics Committee of Rajamangala University of Technology Suvarnabhumi (IRB-RUS-2564-028) on 2 November 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the subjects to publish this paper.

Data Availability Statement: Labelled Faces in the Wild (LFW) is available at http://viswww.cs.umass.edu/lfw/. Source code for performing evaluation is available at https://gitlab.com/benapa/sample-selection-for-face-identification-system

8. References

- B. Maze et al., "IARPA janus benchmark-C: Face dataset and protocol," in Proceedings -2018 International Conference on Biometrics, ICB 2018, Jul. 2018, pp. 158–165, doi: 10.1109/ICB2018.2018.00033.
- [2]. Adjabi, Insaf, et al. "Multi-Block Color-Binarized Statistical Images for Single-Sample Face Recognition." Sensors 21.3 (2021): 728.
- [3]. Bodini, Matteo, et al. "Single sample face recognition by sparse recovery of deep-learned lda features." *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, Cham, 2018.
- [4]. J. P. Robinson, G. Livitz, Y. Henon, C. Qin, Y. Fu, and S. Timoner, "Face Recognition: Too Bias, or Not Too Bias?," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2020-June, pp. 1–10, Feb. 2020, Accessed: Mar. 08, 2021. [Online]. Available: http://arxiv.org/abs/2002.06483.
- [5]. J. G. Cavazos, P. Jonathon Phillips, C. D. Castillo, and A. J. O'Toole, "Accuracy comparison across face recognition algorithms: Where are we on measuring race bias?," arXiv. arXiv, Dec. 16, 2019, doi: 10.1109/tbiom.2020.3027269.

- [6]. "Face recognition terminal for access control | Face authentication for attendance system | FindFace." https://findface.pro/en/cases/access-management/ (accessed Mar. 08, 2021).
- [7]. "The use of facial recognition to fight crime: Japan case Geospatial World." https://www.geospatialworld.net/blogs/the-use-of-facial-recognition-to-fight-crime-japancase/ (accessed Mar. 08, 2021).
- [8]. "How China uses facial recognition to control human behavior CNET." https://www.cnet.com/news/in-china-facial-recognition-public-shaming-and-control-gohand-in-hand/ (accessed Mar. 08, 2021).
- [9]. N. Hazim Barnouti, S. Sameer Mahmood Al-Dabbagh, and W. Esam Matti, "Face Recognition: A Literature Review," *International Journal of Applied Information Systems*, vol. 11, no. 4, pp. 21–31, Sep. 2016, doi: 10.5120/ijais2016451597.
- [10]. P. Viola and M. J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.
- [11]. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 2005, vol. I, pp. 886–893, doi: 10.1109/CVPR.2005.177.
- [12]. H. Bay, A. Ess, T. Tuytelaars, and L. van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, Jun. 2008, doi: 10.1016/j.cviu.2007.09.014.
- [13]. T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 3021, pp. 469–481, 2004, doi: 10.1007/978-3-540-24670-1 36.
- [14]. Y. Zhou, D. Liu, and T. Huang, "Survey of face detection on low-quality images," in Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Jun. 2018, pp. 769–773, doi: 10.1109/FG.2018.00121.
- [15]. M. Kirby and L. Sirovich, "Application of the Karhunen-Loéve Procedure for the Characterization of Human Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, 1990, doi: 10.1109/34.41390.
- [16]. M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586– 591, doi: 10.1109/CVPR.1991.139758.
- [17]. M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991, doi: 10.1162/jocn.1991.3.1.71.
- [18]. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis* and Machine Intelligence, vol. 19, no. 7, pp. 711–720, 1997, doi: 10.1109/34.598228.
- [19]. K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *Journal of the Optical Society of America A*, vol. 14, no. 8, p. 1724, Aug. 1997, doi: 10.1364/josaa.14.001724.
- [20]. W. Zhao, W. Zhao, R. Chellappa, and P. J. Phillips, "Subspace Linear Discriminant Analysis for Face Recognition," 1999, Accessed: Mar. 08, 2021. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.7.6280.
- [21]. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to humanlevel performance in face verification," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Sep. 2014, pp. 1701–1708, doi: 10.1109/CVPR.2014.220.
- [22]. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 815–823, Mar. 2015, doi: 10.1109/CVPR.2015.7298682.

- [23]. O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," British Machine Vision Association, 2015. Accessed: Mar. 08, 2021. [Online].
- [24]. Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Jun. 2018, pp. 67–74, doi: 10.1109/FG.2018.00020.
- [25]. M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," Dec. 2002, pp. 566–568, doi: 10.1109/icpr.1994.576361.
- [26]. V. Perlibakas, "Distance measures for PCA-based face recognition," doi: 10.1016/j.patrec.2004.01.011.
- [27]. L. Greche, M. Jazouli, N. Es-Sbai, A. Majda, and A. Zarghili, "Comparison between Euclidean and Manhattan distance measure for facial expressions classification," May 2017, doi: 10.1109/WITS.2017.7934618.
- [28]. H. Wang et al., "CosFace: Large Margin Cosine Loss for Deep Face Recognition," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 5265–5274, Jan. 2018, Accessed: Mar. 08, 2021. [Online]. Available: http://arxiv.org/abs/1801.09414.
- [29]. R. de Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The Mahalanobis distance," Chemometrics and Intelligent Laboratory Systems, vol. 50, no. 1, pp. 1–18, Jan. 2000, doi: 10.1016/S0169-7439(99)00047-7.
- [30]. Adjabi, Insaf, et al. "Multi-Block Color-Binarized Statistical Images for Single-Sample Face Recognition." Sensors 21.3 (2021): 728.
- [31]. Zeng, Junying, et al. "Deep convolutional neural network used in single sample per person face recognition." *Computational intelligence and neuroscience* 2018 (2018).
- [32]. You, Fucheng, Yue Cao, and Chenwei Zhang. "Deep Domain Adaptation with a Few Samples for Face Identification." 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR). IEEE, 2017.
- [33]. Grother, Patrick, and Elham Tabassi. "Performance of biometric quality measures." *IEEE transactions on pattern analysis and machine intelligence* 29.4 (2007): 531-543.
- [34]. Gao, Xiufeng, et al. "Standardization of face image sample quality." International Conference on Biometrics. Springer, Berlin, Heidelberg, 2007.
- [35]. Ahmed, Sulayman, et al. "Optimum feature selection with particle swarm optimization to face recognition system using Gabor wavelet transform and deep learning." BioMed Research International 2021 (2021).
- [36]. Moeini, Ali, et al. "Open-set face recognition across look-alike faces in real-world scenarios." Image and Vision Computing 57 (2017): 1-14.
- [37]. Allagwail, S., Gedik, O. S., & Rahebi, J. (2019). Face recognition with symmetrical face training samples based on local binary patterns and the Gabor filter. Symmetry, 11(2), 157.
- [38]. Lu, Jiwen, Yap-Peng Tan, and Gang Wang. "Discriminative multimanifold analysis for face recognition from a single training sample per person." *IEEE transactions on pattern analysis and machine intelligence* 35.1 (2012): 39-51.
- [39]. Zhou, Xiaofei, et al. "Kernel subclass convex hull sample selection method for SVM on face recognition." Neurocomputing 73.10-12 (2010): 2234-2246.
- [40]. Boutros, Fadi, et al. "Periocular biometrics in head-mounted displays: A sample selection approach for better recognition." 2020 8th International Workshop on Biometrics and Forensics (IWBF). IEEE, 2020.
- [41]. "GitHub ipazc/mtcnn: MTCNN face detection implementation for TensorFlow, as a PIP package." https://github.com/ipazc/mtcnn (accessed Mar. 08, 2021).
- [42]. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.
- [43]. "GitHub rcmalli/keras-vggface: VGGFace implementation with Keras Framework." https://github.com/rcmalli/keras-vggface (accessed Mar. 08, 2021).

- [44]. J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu, "Squeeze-and-Excitation Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 8, pp. 2011-2023, 1 Aug. 2020,
- [45]. Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- [46]. "LFW Face Database : Main." http://vis-www.cs.umass.edu/lfw/ (accessed Mar. 08, 2021).
- [47]. Mehdipour Ghazi, Mostafa, and Hazim Kemal Ekenel. "A comprehensive analysis of deep learning based representation for face recognition." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2016.
- [48]. Sellahewa, Harin, and Sabah A. Jassim. "Image-quality-based adaptive face recognition." *IEEE Transactions on Instrumentation and measurement* 59.4 (2010): 805-813.
- [49]. Dos Santos, Cassio Elias, and William Robson Schwartz. "Extending face identification to open-set face recognition." 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images. IEEE, 2014.
- [50]. Moctezuma, Luis Alfredo, and Marta Molinas. "Towards a minimal EEG channel array for a biometric system using resting-state and a genetic algorithm for channel selection." Scientific RepoRtS 10.1 (2020):
- [51]. Vareto, Rafael Henrique, and William Robson Schwartz. "Unconstrained Face Identification using Ensembles trained on Clustered Data." 2020 IEEE International Joint Conference on Biometrics (IJCB). IEEE.



Napa Sae-Bae received her PhD in Computer Science from Polytechnic School of Engineering, New York University in 2014, whose advisors are Prof. Nasir Memon and Prof. Katherine Isbister. She currently holds a faculty position at computer science department, faculty of science, Srinakharinwirot University, Thailand. Her research interests lie in the area of biometric, authentication, consumer security, pattern recognition, signal processing, and image processing, e-mail: napasa@g.swu.ac.th



Utharn Buranasaksee is a faculty member of Faculty of Science and Technology at Rajamangala University of Suvarnabhumi. He received a bachelor's degree in Computer Application from Christ college Bangalore University and a master's degree in Information Technology and a doctoral's degree in Computer Science at King Mongkut's University of Technology Thonburi. His current fields are database indexing, algorithm optimization, artificial intelligent, and blockchain.