Multimedia Systems (2024) 30:114
https://doi.org/10.1007/s00530-024-01306-y

REGULAR PAPER

Design and implementation of a real-time face recognition system based on artificial intelligence techniques

Chih-Yung Chang[1] · Arpita Samanta Santra[2] · I-Hsiung Chang[3] · Shih-Jung Wu[1] · Diptendu Sinha Roy[4] · Qiaoyun Zhang[1,5]

Received: 29 March 2023 / Accepted: 26 February 2024 / Published online: 5 April 2024

Abstract

This paper mainly discusses the asymmetric face recognition problem where the number of names in a name list and the number of faces in the photo might not be equal, but each face should be automatically labeled with a name. The motivation for this issue is that there had been many meetings in the past. After each meeting, the participant took group photos. The meeting provided only a corresponding name list of participants without one-to-one labels. In the worst case, the group photo might mix with the faces that were not participating in the meeting. Another reason for asymmetric face recognition is that some meeting personnel did not appear in photos because they assisted in taking pictures. This paper proposes an asymmetric face recognition mechanism, called *AFRM* in short. Initially, the proposed *AFRM* adopts the histogram of oriented gradients (HOG) and support vector machine (SVM) to detect and extract all faces from photos. Next, *AFRM* extracts the features from each face using the convolution feature map (*Conv_FF*) and adopts the features to partition the faces into different classes. Then, the *AFRM* applies the statistic-based mechanism to map each name in the name list to each face class. According to this mapping, each face will be associated with one name. To quickly identify a face during the meeting, the *AFRM* applies the *K-nearest neighbors* (*KNN*) to represent the features of each face. During the new meeting, the proposed *AFRM* can extract the feature of one face and then adopts *KNN* to derive the features. Experimental results show that the proposed mechanism achieves more than 97% accuracy without one-to-one name and face labeling.

Keywords Face recognition · Unsupervised learning · Asymmetric training resources · Meeting applications

Communicated by R. Huang.

✉ Chih-Yung Chang
cychang@mail.tku.edu.tw

Arpita Samanta Santra
santraarpita83@gmail.com

I-Hsiung Chang
elite5931.tw@gmail.com

Shih-Jung Wu
wushihjung@mail.tku.edu.tw

Diptendu Sinha Roy
diptendu.sr@nitm.ac.in

Qiaoyun Zhang
zqyun@chzu.edu.cn

[1] Tamkang University, New Taipei, Taiwan

[2] National Tsing Hua University, Hsinchu, Taiwan

[3] National Taiwan Normal University, Taipei, Taiwan

[4] National Institute of Technology, Shillong, India

[5] Chuzhou University, Chuzhou, China

1 Introduction

Facial recognition strategy plays a significant role in identifying human faces for authentication through advanced technology (localization and extraction of face regions) in several applications such as access control systems, check-in and check-out at workplaces, image retrieval, crowd surveillance, and video conference. It applies biometrics patterns to extract facial features and collects a set of unique biometric components from photographs, videos, or any audiovisual element of human faces. Facial recognition is also one of the most extensively demanded identification processes for online identity verification in large-scale utilization of video surveillance, security, commercial areas, and logical access control systems in everyday life. Most of the advantages of facial recognition are in the policing system (law enforcement agencies) to easily distinguish ID fraud without a person's information, preventing social security crime. Face recognition and mapping have been integrated with various

applications such as banking, intelligent devices, attendance systems, and human tracking.

Nowadays, many meetings are organized in companies and organizations with a lot of participants. To better maintain the social relation and exploit cooperation opportunities, a business person needs to remember each face and name of the participant in a new meeting. However, it is challenging and time-consuming to manually label the participants' names with their faces (hardcopy). In recent years, the deep learning model has been widely used to recognize the face and obtain the participant's name. However, almost all deep learning models require one-to-one labeling during their training processes. Though a name list and a set of photos can be easily collected from each meeting, the mapping from each name in the name list to each face in one photo is still difficult. This occurs because the dataset collected from each meeting is asymmetric. The asymmetric dataset can be occurred due to two reasons. First, the participant might be the one who takes photos in the meeting. This results in the problem that the participant's face is disappeared in the photo. Therefore, the name of this participant cannot be mapped to the face of the collected photo. The other reason is that some other persons who are not participants in the meeting but are friends of the participants might appear in the photo. For these persons, there is no name in the name list. The above-mentioned two reasons lead to the dataset to be asymmetric, which increases the difficulties to develop the AI-based algorithm for automatically mapping each name in the name list to the face of the participant.

This article proposes an artificial intelligence-based real-time face recognition approach to alleviate this asymmetrical problem in the automation process. Instead of one-to-one labeling, the proposed approach takes an attendance or participant list and a set of photos captured in different sessions of each meeting as input. The proposed face recognition method processes this input pair, including name lists and a set of photos collected from meetings, and automatically marks each person in the group images during training. The proposed approach can recognize the identity of each person from test images and identify the name of the person who has appeared in the previous meeting and attends the next meeting.

In the past, researchers usually applied traditional statistical methods for face recognition by manually collecting data. These conventional methods could be time-consuming because they required manual labor at some point during the data collection and analysis. In recent years, artificial intelligence has incrementally become a useful and powerful tool in terms of collecting and analyzing a large amount of open data available online. Being aware of this feature, some researchers started to adopt these AI techniques to determine the names. For example, the machine learning technique was used in Ref. [1], and deep learning technique was also used in Ref. [2]. These research studies had more accurately predicted the target person's name. However, these methods were complex and require various parameters adjustment to achieve the result. In different of these approaches, the proposed *AFRM* aims to build on the methods used by the predecessors and solve the existing problems of the previous methods by training the artificial intelligence models using deep learning techniques.

The studies fell in the real-time artificial intelligence-based face recognition approaches and could be further divided into feature representation, one-to-one labeling, and recognition stages. Although researchers reduced the time and error of the feature representation by transforming the manual into an automatic process using local binary patterns histograms (LBPH), generalized discriminant local median preserving projections (GDLMPP), and DLMPP algorithms, they overlooked the original value, which was not strong enough in those methods to extract all features. The second stage is one-to-one labeling, which required mapping each face to a corresponding name during the training process. Almost all automated face recognition approaches adopted a manual strategy for mapping. However, it is also time-consuming and error-prone. In the last stage, i.e., real-time face recognition, traditional approaches such as convolution neural network (CNN), CNN-recurrent neural network (CNN-RNN), and deep convolutional neural network (DCNN) were widely used. However, most of them needed labeled data, and their capabilities were unsuitable to cope with the problem of asymmetric face recognition.

A big challenge that exists in the asymmetric case is that collecting photos and information (data) separately from the internet or public dataset is difficult to analyze the data in this case. Unlike most previous studies, this paper considers the asymmetric dataset, where one-to-one mapping of participants' faces and names is unavailable. To overcome this challenge, the proposed *AFRM* method adopts HOG with SVM to find the frontal faces in photos. It can completely extract the entire features and handle shape deformation. First, all frontal face features are extracted from all photos of all past meetings using HOG. Second, all faces of all meetings are predicted by labeling HOG features and the training SVM model. Then, all trained face features are fed into the *Conv_FF* layer. In this stage, the *Conv_FF*-based feature representation map is used to represent 128×1 feature vectors, and the name list of the meetings is processed by the name probability inference method to calculate the probability of each name in the meeting. Finally, the approach adopts the *KNN*-classifier, which needed significantly less computational cost without decreasing the recognition accuracy. In a variant of this, the proposed method can automatically label faces to corresponding names.

The proposed *AFRM* successfully tags participants' faces by their names. Assume that the target person is A. In

addition, assume that person A has participated in the new meeting, and A had also participated in different previous meetings. To remind the name of participant A, the *AFRM* is used to find out the forgettable name of that target person, which are the main contributions of the proposed *AFRM* approach. The main contributions of this work are listed as follows:

1. ***Construction of feature vectors***. This paper extracts the face feature vectors using *Conv_FF*, which is qualitative numerical data, that can help to make a decision. It can be used as the presentation of face features for quickly analyzing each face and enhancing the execution process of the model.
2. ***Classification and grouping at the same time***. After extracting the feature vectors, cosine similarities are computed among all face feature vectors to create different categories for different faces. Those face feature vectors that are larger than the threshold value will be collected into the same category. Otherwise, the proposed *AFRM* will create a new category for the new face to improve the classification ability. This implies the classification and grouping of the input face vector will be executed at the same time.
3. ***Probability Distribution.*** First, the probability of each name can be calculated by dividing the total number of participant names in a meeting list. Second, the proposed *AFRM* computed the probability distribution value of each name by adding the total number of the probability of the same name in all meeting lists.
4. ***Mapping and KNN clustering.*** According to the probability value, the name will be tagged with the face category. The probability of the name tags will be accumulated according to the meeting list. All face feature vectors with names as supervised data fed into *KNN* cluster as an input which can improve the efficiency since the *KNN* has the face vectors as inputs and the corresponding name as the output. The *KNN* learns from the inputs and labels and creates a *KNN* model for the classification of an input face vector in the future. When we would like to know the name of a person who attends any meeting in future, then the proposed model will easily find the name of the person.

The remaining parts of this paper are organized as follows. Section 2 discuss the background of existing studies related to the *AFRM* approach. Section 3 focuses on the assumptions and goals of the prospected issue. Section 4 presents the proposed *AFRM* algorithm in detail. The performance improvements of the proposed *AFRM* against the existing studies are proposed in Section 5. Finally, Section 6 concludes this work and gives the future work of this study.

2 Related work

The proposed *AFRM* mainly relates to three major research directions: feature extraction, asymmetric labeling and face recognition to determine an unknown person's name. Therefore, the related studies will be classified into three classes, including feature representation, asymmetric labeling and real-time face recognition.

2.1 Feature representation

The feature representation is a fundamental operation since most face classifications need the feature representation as their inputs. In Ref. [3], the author combined the LBPH [4] with Eigenface [5] algorithms to extract the face by taking photos and self-registered names. A database was used to store the unique features of faces and their corresponding names. During the inference phase, the face features were extracted from the input image and compared with the database features. The self-registered name of the most similar face features was the outcome for the input face. Although the method can reidentify the faces, the computational cost is higher. The method can only recognize the faces closer to the camera and can only recognize the frontal faces. To solve the problem of optimal projection matrix for face recognition, Wan et al. [6] proposed a generalized discriminant local median preserving projection (GDLMPP) algorithm based on DLMPP, which could transform the original sample to a low-dimensional eigenspace within non-singular class scatter matrix. This approach could improve the recognition rate. However, tuning the appropriate parameters for this method was very difficult. Rose et al. [7] mainly focused on facial recognition, whether eyes are open or closed, wearing glasses or not, either frontal or non-frontal faces under constrained and unconstrained conditions, applying SVMs using LBP and HOG features and convolutional neural networks. The HOG feature extractor model did not get satisfactory accuracy results. However, using HOG with the SVM algorithm achieved high accuracy results. In variants of these approaches, inspired by the research work *AFRM* approach adopts a combination of HOG and SVM methods to extract each face from photos, which is suited for the face detection module. These identified faces are passed through our *Conv_Fm* layer to extract rich features.

2.2 Asymmetric labeling

Asymmetric labeling indicates that the data inside a training dataset is not labeled with their classes and the numbers of data and labels are not equal. In asymmetric conditions,

features are learned with unlabeled input data in the sense of lack of equality or equivalence between inputs and expected value. Most of the existing multi-label methods disregard asymmetry label correlation. To overcome this problem, Bao et al. [8] presented asymmetry label correlation (ACML) as a label adjacency matrix by measuring cosine similarity with label and correlation matrix for multi-label learning. Zhang et al. [9] addressed an asymmetric active querying strategy by appointing different probabilities that achieved higher F-scores with imbalanced streaming data under query budget data. However, this method overlooked imbalance issues in the optimization process and tended to query more majority data due to the recommended parameter settings. To solve those imbalance problems, *AFRM* proposes a statistic-based asymmetric labeling approach.

2.3 Real-time face recognition

Face reidentification aims to map a specific feature set to an identity (name) from images according to targets. In recent trends, increasing demand for real-time reidentification without high computational time to annotated faces has become popular [10]. In the literature, existing face reidentification approaches could be classified as unsupervised domain adaptation methods [11–14], which required annotated trained datasets and unannotated target datasets to calculate model accessibility. The other type was fully unsupervised methods, which required only an unannotated dataset [15–18]. The annotation process in training datasets needed immense manual labor. However, the unannotated dataset did not require much time but suffered from an acceptable accuracy. In variants of these approaches, the proposed *AFRM* method adopts an automated approach to map the trained datasets.

The fully unsupervised face reidentification approaches rely on pseudolabels to train the network. The HCT [19] adopted hierarchical clustering to produce pseudo labels and train the CNN for learning the features. The study [20] generated multiple labels to samples and proposed a new loss function for multi-label training. Cheng et al. [21] proposed a two-layer convolutional neural network (CNN) to learn the high-level features in variations of illumination and expression of facial images. Although CNN can improve the superior performance of the Sparse Representation Classifier (SRC) in the image classification area, the huge number of trainable parameters makes it difficult to train when a small dataset was used. In Ref. [22], Zangeneh devised a coupled mappings method composed of two parts of DCNN on a low-resolution dataset. In this approach, the author matched high-resolution (HR) and low-resolution (LR) images to find a nonlinear transformation in the common space between LR and HR-resolution images. To overcome the face recognition (face-PAD) problem, Nguyen et al. [23] demonstrated the CNN-RNN network and multi-level local binary pattern (MLBP) to extract handcrafted image features and learn the discrimination features, respectively. However, in the early stages, feature representations are not good enough to produce high-quality pseudo labels and contaminate the network training process. Therefore, it is necessary to design a cluster refinement method to improve the clustering quality before feeding the pseudo labels to train the network. In Ref. [11], Fu measured specific portions of the facial, namely eyes, nose, and mouth, for the solution of FR by deep-learning-based IQA CNN model. To improve face recognition quality significantly, Fu et al. [12] appraised the prediction performance of handcrafted features by image quality assessment (IQA) and Face-Image Quality Assessment (FIQA). Mainly, IQA methods made decisions while FIQA methods focused on the center part of the face. However, the asymmetric face recognition investigated in this paper is an unsupervised application and the mechanisms proposed in Refs. [11, 12] cannot be applied.

3 Assumptions and problem formulation

This section presents the system model and problem formulation of the investigated issue. Most well-known face identification approaches required symmetric labeling during training. It is almost impossible to label all the faces at different meetings. This paper aims to present an efficient algorithm called *AFRM*, which automatically labels the name of each participant in meetings conducted over the last year. Let $M = \{m_i | 1 \le i \le |M|\}$ represent the set of meetings and $L = \{L_i | 1 \le i \le |L|\}$ represent the set of name lists where a name list L_i is existed for the i-th meeting m_i. Therefore, two types of data, i.e., name list and sets of images, are collected in each meeting. Let $m_i = (P_i, L_i)$, where $P_i = \{p_{i,j} | 1 \le i \le m_i, 1 \le j \le |P_i|\}$ represents the set of images, where $p_{i,j}$ represents the j-th photo captured in i-th meeting. Let $L_i = \{name_{i,l} | 1 \le l \le |L_i|\}$ denote the names in the name list L_i where $name_{i,l}$ represents the l-th participant name in i-th meeting.

Furthermore, let $p_{i,j} = \{f_{i,j}^k | 1 \le k \le |p_{i,j}|\}$ represent the set of faces consisting of the k-th face of the j-th photo in the i-th meeting. Any face $f_{i,j}^k$ belongs to F ($f_{i,j}^k \in F$) and name $name_{i,l}$ belongs to L ($name_{i,l} \in L$). Let F_i denote the set of all frontal faces extracted from the meeting m_i. Let F be the set of all faces that appeared in the pictures which were collected in M meetings. That is,

$$F_i = \bigcup_{p_{i,j} \in P_i} \bigcup_{f_{i,j}^k \in p_{i,j}} f_{i,j}^k \tag{1}$$

$$F = \bigcup_{i=1}^{|M|} F_i$$

Let L denote all names which appeared in the name lists of the previous M meetings. That is,

$$L = \bigcup_{i=1}^{|M|} \bigcup_{l=1}^{|L_i|} name_{i,l} \quad (2)$$

The following gives an example to help understand the investigated problem. Assume that there are a set of pictures and name lists collected from the previous 100 meetings. That is, $M = \{m_1, m_2, \ldots m_{100}\}$, where $m_1, m_2, \ldots, m_{100}$ are the labels of 100 meetings. As shown in Fig. 1, there are 10 pictures collected from the meetingm_5. Therefore, the set of pictures in meeting m_5 can be presented as $P_5 = \{p_{5,1}, p_{5,2}, \ldots p_{5,10}\}$. In particular, there are nine faces in the first picture $p_{5,1}$ ofP_5. These faces can be further represented as $p_{5,1} = \{f_{5,1}^1, f_{5,1}^2, \ldots f_{5,1}^9\}$. In addition to the pictures collected in the meetings, there are a set of name lists collected for the 100 meetings. As shown in Fig. 1, the name list L_5 is collected in the meeting m_5. The meeting list L_5 has 7 names. They are actual participants in the meetingm_5. That is, $L_5 = \{LiSi, Chien-IWeng, LuoZhiyun, ZhangSan, ShenghuiZhao, LeeMinWoo, Chao-TChang\}$. Among these nine faces of $p_{5,1} = \{f_{5,1}^1, f_{5,1}^2, f_{5,1}^3, f_{5,1}^4, f_{5,1}^5, f_{5,1}^6, f_{5,1}^7, f_{5,1}^8, f_{5,1}^9\}$, three faces $\{f_{5,1}^3, f_{5,1}^7, f_{5,1}^9\}$ are participants' friends. In addition, Chien-I Weng is the one who takes pictures for the group photo. Hence, he is absent from the photo $p_{5,1}$, but he is an actual participant because his name is on the name listL_5. As the number of extracted faces of $p_{5,1}$ and the number of names present in the name list L_5 are not the same, and they are not appropriately mapped. This condition is treated as an asymmetric case. Table 1 describes an example of actual and predicted names of all faces $f_{i,j}^k$ in meeting m_5 for easily understanding the prob-

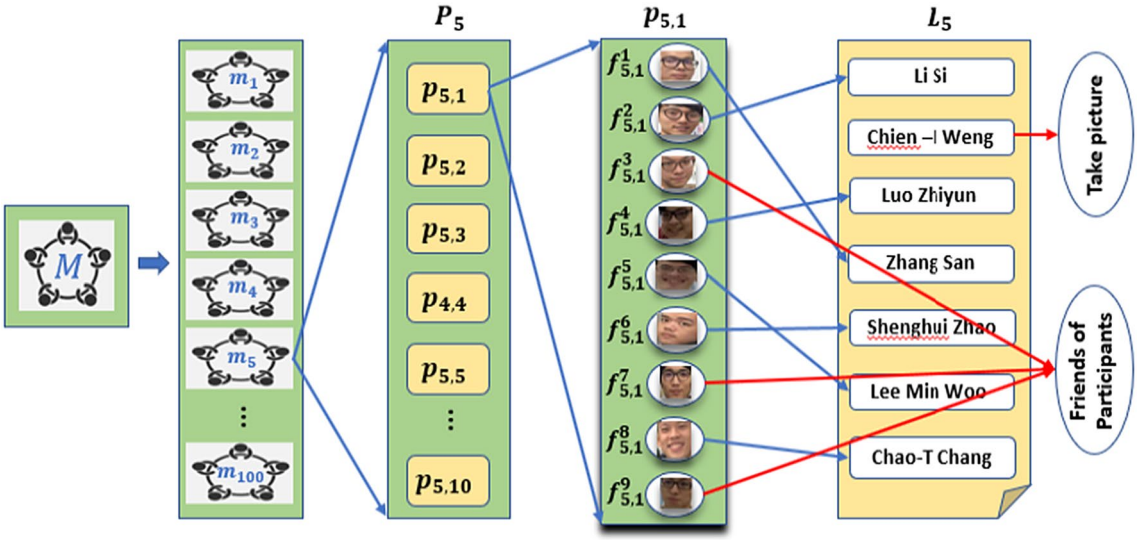


Fig. 1 Conceptual diagram of asymmetric case

Table 1 An example of the actual and predicted names of the faces

Face number ($f_{i,j}^k$) in photo $p_{5,1}$	Names ($name_{i,l}$) in Name List L_5	Actual name ($name_{i,l}$)	Predicted name ($name_{i,l}'$)
$f_{5,1}^1$	Li Si	Zhang San	Zhang San
$f_{5,1}^2$	Chien-I Weng	Li Si	Li Si
$f_{5,1}^3$	Luo Zhiyun	Friend of Li Si	Chao-T Chang
$f_{5,1}^4$	Zhang San	Luo Zhiyun	Luo Zhiyun
$f_{5,1}^5$	Shenghul Zhao	Lee Min Woo	Lee Min Woo
$f_{5,1}^6$	Lee Min Woo	Shenghui Zhao	Chao-T Chang
$f_{5,1}^7$	Chao-T Chang	Friend of Zhang San	NULL
$f_{5,1}^8$	–	Chao-T Chang	Chao-T Chang
$f_{5,1}^9$	–	Friend of Li Si	NULL

lem formulation of asymmetric cases. The actual and predicted names are denoted by $name_{i,l}$ and $name'_{i,l}$, respectively. Four cases have occurred. First, the predicted name are the same as the actual names of faces ($f^1_{5,1}$,$f^2_{5,1}$,$f^4_{5,1}$,$f^5_{5,1}$, and $f^8_{5,1}$). Second, the predicted name is not the same as the actual name of the face $f^6_{5,1}$. Third, the predicted name is wrong for the face $f^3_{5,1}$ because he is a friend of Li Si, whose name is not available on the meeting listL_5. Finally, the predicted names become NULL values for the faces $f^7_{5,1}$, and $f^9_{5,1}$ because they are the friend of Zhang San and Li Si, who are not participants in the meetingm_5. The confusion matrix is used to analyze the actual and predicted name classes. To create a confusion matrix, the rows of binary values of actual name classes ($\lambda^{k,q}_{i,j}$) are entered along with the columns of binary values of predicted name classes ($\hat{\lambda}^{k,q}_{i,j}$) into a table, where $i, j, k,$ and q be meeting, photo, face, and class number, respectively. The column values are the predicted values of the proposed *AFRM* model. Let ***COUNT (i, j)*** be a function which returns the number in row i and column j in Table 2. Table 2 describes a confusion matrix which shows the value of predicted name classes ($\hat{\lambda}^{k,q}_{i,j}$) from the proposed model vs the value of actual name classes ($\lambda^{k,q}_{i,j}$) from the dataset. When one class name ($\lambda^{k,q}_{i,j}$) is considered for predictive analysis, that class name becomes the positive class, and the rest of the classes will be negative classes. Four name identification functions occur in the confusion matrix table. It takes one face as an input and predicts the class name of that face as an output.

Let $\lambda^{k,q}_{i,j}$ denote the positive class name of the face $f^k_{i,j}$. That is, the hypothesis is "Is the class name $\lambda^{k,q}_{i,j}$?" Let $\hat{\lambda}^{k,q}_{i,j}$ denote the predicted class name for the face $f^k_{i,j}$. Let $\delta^{k,q}_{i,j}$ be a Boolean variable which represents whether or not the prediction of $\hat{\lambda}^{k,q}_{i,j}$ is correct. That is,

$$\delta^{k,q}_{i,j} = \begin{cases} 1, \text{ if } \hat{\lambda}^{k,q}_{i,j} \text{ is correct} \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

Let true positive, denoted by $TP^{k,q}_{i,j}$, be a Boolean variable which represents the true positive of the prediction $\hat{\lambda}^{k,q}_{i,j}$. That is,

$$TP^{k,q}_{i,j} = \begin{cases} 1, \ \delta^{k,q}_{i,j} \text{ and } (\hat{\lambda}^{k,q}_{i,j} = \lambda^{k,q}_{i,j}) \\ 0, \text{ otherwise} \end{cases} \tag{4}$$

where $i, j, k,$ and q be meeting, photo, face, and class number, respectively. The predicted answer for the hypothesis is YES, and the prediction is correct. Similarly, let true negative, denoted by $TN^{k,q}_{i,j}$, be a Boolean variable which represents the true negative of the prediction $\hat{\lambda}^{k,q}_{i,j}$. That is,

$$TN^{k,q}_{i,j} = \begin{cases} 1, \ \delta^{k,q}_{i,j} \text{ and } (\hat{\lambda}^{k,q}_{i,j} \neq \lambda^{k,q}_{i,j}) \\ 0, \text{ otherwise} \end{cases} \tag{5}$$

where $i, j, k,$ and q be meeting, photo, face, and class number, respectively. It indicates that the predicted answer is NO, and the prediction is correct.

Let false positive, denoted by $FP^{k,q}_{i,j}$, be a Boolean variable which represents the false positive of the prediction $\hat{\lambda}^{k,q}_{i,j}$. That is,

Table 2 Confusion matrix for multiclass of names

Table Name Q	Participant names		Predicted class ($\hat{\lambda}^{k,q}_{i,j}$)								
			Zhang San	Li Si	Friend of Li Si	Luo Zhiyun	Lee Min Woo	Shenghui Zhao	Friend of Zhang San	Chao-T Chang	Friend of Li Si
			1	2	3	4	5	6	7	8	9
Actual class ($\lambda^{k,q}_{i,j}$)	Zhang San	1	15	0	0	0	0	1	0	2	0
	Li Si	2	1	21	0	3	0	0	0	0	0
	Friend of Li Si	3	0	0	12 (NULL)	0	1	0	0	0	0
	Luo Zhiyun	4	0	0	0	16	1	1	0	1	0
	Lee Min Woo	5	0	0	0	0	11	0	0	1	0
	Shenghui Zhao	6	3	0	0	0	0	17	0	0	0
	Friend of Zhang San	7	2	1	0	0	2	0	10 (NULL)	0	0
	Chao-T Chang	8	1	0	0	1	2	1	0	12	0
	Friend of Li Si	9	1	2	0	0	3	0	0	0	14 (NULL)

$$FP_{i,j}^{k,q} = \begin{cases} 1, \ (1 - \delta_{i,j}^{k,q}) \text{ and } (\widehat{\lambda}_{i,j}^{k,q} = \lambda_{i,j}^{k,q}) \\ 0, \ \text{otherwise} \end{cases} \tag{6}$$

where $i, j, k,$ and q be meeting, photo, face, and class number, respectively. It indicates that the predicted answer is YES, but it is incorrect.

Similarly, let false negative, denoted by $FN_{i,j}^{k,q}$, be a Boolean variable which represents the false negative of the prediction $\widehat{\lambda}_{i,j}^{k,q}$. That is,

$$FN_{i,j}^{k,q} = \begin{cases} 1, \ (1 - \delta_{i,j}^{k,q}) \text{ and } (\widehat{\lambda}_{i,j}^{k,q} \neq \lambda_{i,j}^{k,q}) \\ 0, \ \text{otherwise} \end{cases} \tag{7}$$

where $i, j, k,$ and q be meeting, photo, face, and class number, respectively.

Let N be all classes in M meetings. The TP value for N classes is expressed by Eq. (8):

$$\begin{aligned} TP_{i,j}^{k} &= \sum_{1 \le q \le N} \lambda_{i,j}^{k,q} \\ TP_{i,j} &= \sum_{1 \le k \le p_{i,j}} TP_{i,j}^{k} \\ TP_{i} &= \sum_{1 \le j \le P_i} TP_{i,j} \\ TP &= \sum_{1 \le i \le M} TP_{i} \end{aligned} \tag{8}$$

Similarly, the N, FP, and FN values are calculated for each of the N classes in M meetings. Those equations are expressed by

$$TN = \sum_{1 \le i \le M} TN_i = \sum_{1 \le i \le M} \sum_{1 \le j \le P_i} \sum_{1 \le k \le p_{i,j}} \sum_{1 \le q \le N} \lambda_{i,j}^{k,q} \tag{9}$$

$$FP = \sum_{1 \le i \le M} FP_i = \sum_{1 \le i \le M} \sum_{1 \le j \le P_i} \sum_{1 \le k \le p_{i,j}} \sum_{1 \le q \le N} \lambda_{i,j}^{k,q} \tag{10}$$

$$FN = \sum_{1 \le i \le M} FN_i = \sum_{1 \le i \le M} \sum_{1 \le j \le P_i} \sum_{1 \le k \le p_{i,j}} \sum_{1 \le q \le N} \lambda_{i,j}^{k,q} \tag{11}$$

The *Precision* is defined as positive classes, which refers to the quality of being exact. Let $\wp$ denote the *Precision* of the prediction of all classes for all input faces F ($f_{i,j}^{k} \in F$). Precision ($\wp$) is mentioned as the number of true positives (TP) over the number of true positives plus the number of false positives (FP). The values of *Precision* ($\wp$) can be further derived by applying Exp. (12):

$$\text{Precision } (\wp) = \frac{TP}{TP + FP} \tag{12}$$

The well-known *Recall* also measures the proportion of actual positives that are predicted correctly. Let $\mathcal{R}$ denote the *Recall* of the prediction of all classes for all input faces $f_{i,j}^{k} \in F$. Recall ($\mathcal{R}$) is defined as the number of true positives (TP) over the number of true positives plus the number of false negatives (FN). The values of *Recall* ($\mathcal{R}$) can be further derived by applying Exp. (13):

$$Recall(\mathcal{R}) = \frac{TP}{TP + FN} \tag{13}$$

The F_1-score is the harmonic mean of precision and recall. Expression (14) gives the calculation of F_1:

$$F_1 = \frac{(2 * \wp * \mathcal{R})}{\wp + \mathcal{R}} \tag{14}$$

Objective:

This paper aims to determine the name for each face $f_{i,j}^{k}$. The objective function of this paper can be expressed by Exp. (15):

Objective:

$$\underset{\widehat{\lambda}_{i,j}^{k,q} \in \boldsymbol{L}_i}{\boldsymbol{max}} \ (\boldsymbol{F}_1) \tag{15}$$

4 The proposed *AFRM* model

The problem of asymmetric face recognition refers to a situation where the number of labels does not match the number of faces to be recognized. This issue often arises in scenarios where a company provides a list of attendees for a meeting and takes many photos during the meeting. However, it is possible that some attendees may not have their faces captured in the photos because someone else is responsible for taking pictures. As a result, the number of names on the attendee list may exceed the number of faces in the photos. Another possibility is that individuals who did not attend the meeting may have their faces appear in the photos, making it challenging to match the names on the attendee list to the faces in the photos. As a result, the number of names on the attendee list may be greater or fewer than the number of faces in the photos.

The asymmetric face recognition problem faces many challenges. Performing face recognition with a given list of meeting attendees and a collection of facial photos is a challenging task. Even when the number of names on the list matches the number of faces in the photos, it remains a difficult problem because there is a lack of one-to-one correspondence between individual faces and names, making supervised learning impractical. Therefore, traditional supervised learning classification techniques cannot be employed

to address this issue. What adds further complexity is when the number of names exceeds or falls short of the number of faces in the photos, making it even more challenging to identify each individual face in the photos.

This paper proposed a novel approach called asymmetric face recognition mechanism (*AFRM*), which is a prediction model for finding out the exact name for the face photo of a target person. To solve these challenges, the proposed *AFRM* model is primarily divided into four phases: feature extraction (FE), cosine similarity (CS) and similar face grouping, probability distribution, as well as name-to-face mapping and *KNN* for learning as shown in Fig. 2. The FE phase encompasses the extraction of features from all faces in every meeting, efficiently diminishing computational complexity by converting the data from a high-dimensional space to a lower-dimensional one. Then, the cosine similarity (CS) and similar face grouping phase extracts the face vectors from faces and applies the cosine similarity to compare feature vectors, aiming to recognize the similar faces and facilitate effective facial grouping. Since the facial data lacks labels, the proposed *AFRM* adopts the similar face grouping phase, which allows similar faces to be categorized into the same class, providing a foundation for subsequent recognition. This approach is suitable for unsupervised or privacy-sensitive scenarios, while direct classification is ideal for fine-grained recognition and identity verification tasks where the module needs to be labeled training data and needs specific identity assignment. Next, the probability distribution is calculated for each name for each given face. It can effectively link faces with names, addressing the issue of varying participant counts in different meetings, and calculating probabilities for each name further enhancing the accuracy of matching names with faces. Finally, the *KNN* is applied to classify the participant's face with the corresponding name. With the addition of new facial photos, the system can continuously learn and reinforce itself, thereby enhancing its recognition capabilities. The details of the four phases are given below.

4.1 Feature extraction (FE) phase

One of the crucial roles played in this phase is extracting face features. When the input data are large and complex, it may require large memory and overfit the training sample. To tackle this problem, this phase aims to extract the features from each face which involves transforming data from a high-dimensional space to a low-dimensional space of the original data. The extracted features from each face are called a feature vector. It can build a set of variables with sufficient characteristics to improve the accuracy of the face recognition result. The following presents the details of the feature extraction applied in the proposed *AFRM*.

4.1.1 SVM classifier with HOG features

Consider each meeting $m_i \in M$, $i = 1, 2, \ldots M$. Let $P_i = \{p_{i,1}, p_{i,2}, \ldots p_{i,|P_i|}\}$ represent the set of images in meetings m_i. Let $F_i = \{f_{i,1}, f_{i,2}, \ldots f_{i,|F_i|}\}$ be the face set of P_i, where $f_{i,1}$ is the set of faces in the picture $p_{i,1}$. In this phase, the set of all photos P_i have been taken as input, and the features of all frontal faces F_i are extracted by applying the feature descriptor extraction algorithm, called histogram of oriented gradients (HOG). The descriptor calculates the gradients of images which are the combination of magnitude (μ) and angle (θ). To calculate the gradient of images, the horizontal g_x and vertical g_y components are determined using Eqs. (16) and (17), respectively, for each pixel value of each image. Those are

$$g_x(r, c) = p_{i,1}(r, c + 1) - p_{i,1}(r, c - 1) \tag{16}$$

$$g_y(r, c) = p_{i,1}(r - 1, c) - p_{i,1}(r + 1, c) \tag{17}$$

where r and c refer to rows and columns of each pixel of the image ($p_{i,1}$), respectively. After calculating g_x and g_y the

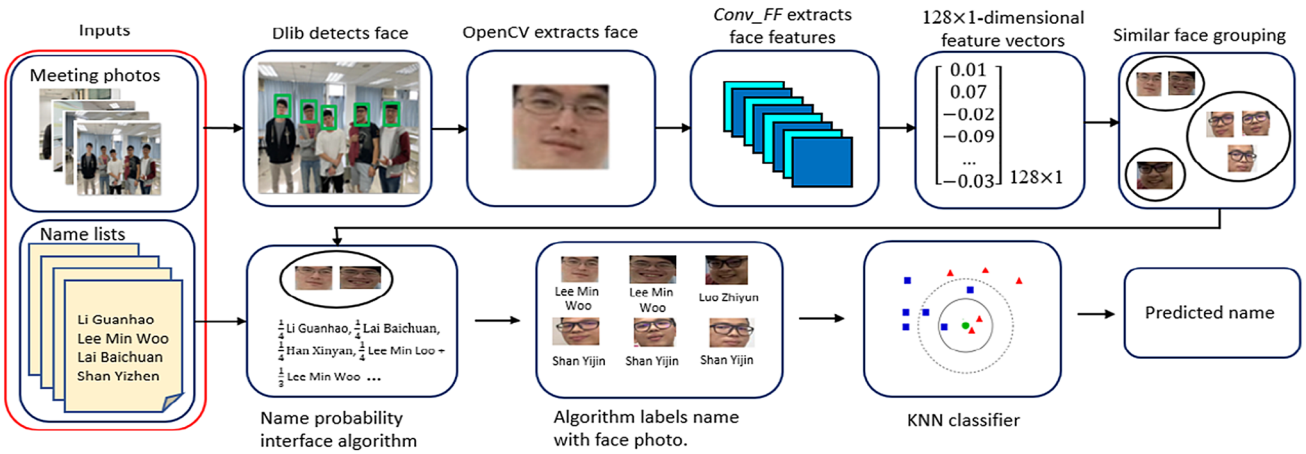


Fig. 2 Face recognition system architecture design

descriptor calculates gradient using magnitude (μ) and angle (θ) by Eqs. (18) and (19), respectively:

$$\mu = \sqrt{g_x^2 + g_y^2} \tag{18}$$

$$\theta = tan^{-1}\left(\frac{g_x}{g_y}\right) \tag{19}$$

The features of each face will be extracted as a vector. Then, these vectors are used in the SVM model to determine a matching score for the input face vector with each of the labels. The SVM returns the label with the maximum score, representing the closest match's confidence within the trained face data. An SVM model can be considered a point space wherein multiple classes are isolated using hyperplanes. The SVM model is trained using a number of feature vectors for multiple faces and stored image features. All extracted face images F of all meetings M may be presented from different viewing angles. For real-time participant identification, the proposed *AFRM* approach only considers frontal faces in the photo $p_{i,j} = \{f_{i,j}^1, f_{i,j}^2, \ldots f_{i,j}^{|p_{i,j}|}\}$, where $f_{i,j}^k$ represents the extracted faces. The number of faces $f_{i,j}^k$ are extracted from each photo $p_{i,j}$ in a meeting m_i, where i, j, and k be the meeting number, photo number, and face number, respectively.

Sometimes participants may not be presented in photos in different situations (photos) in meetings. That is, the organizers, volunteers, or participants' friends may be presented in some photos. As a result, the number of participants' names in the name list and the number of faces in the photos are not equal. Recall that F_i denotes the set of all frontal faces extracted from the meeting m_i and let L_i denote the corresponding name list of the meeting m_i. The asymmetric Face Recognition problem refers to the situation where the number of labels does not match the number of faces to be recognized. Let Φ represent a function that captures the relationship between name lists L_i and extracted faces F_i in different scenarios. That is,

$$\Phi = \left\{ \begin{array}{ll} |L_i| > |F_i|, & \text{if the number of name lists exceeds the number of extracted faces} \\ |L_i| < |F_i|, & \text{if the number of extracted faces exceeds the number of name lists} \\ |L_i| = |F_i|, & \text{otherwise.} \end{array} \right\}$$

In the first two situations, the frontal faces and names cannot be mapped one-to-one in the meeting, indicating the presence of an asymmetric scenario. Traditional face recognition problems often assume that each face is labeled with a corresponding name. Therefore, applying supervised learning classification techniques, models can be trained to extract and learn features of each face during the training process. Once the model is trained, it can use the input face features to determine which name the face belongs to.

However, in the context of asymmetric face recognition, even when the number of names matches the number of faces in attendance at a meeting, traditional supervised learning techniques used in typical face recognition problems are not applicable. The primary challenge in asymmetric face recognition is the lack of one-to-one training data mapping faces to names. This mismatch between the number of categories and labels is the first obstacle. An even more formidable challenge arises when the number of names does not correspond to the number of faces, creating a situation where there may be more or fewer categories compared to labels. Therefore, asymmetric face recognition problems are more challenging than traditional face recognition problems.

4.1.2 Extract feature vector by *Conv_FF*

The illumination and background conditions on different images affect the feature extraction. Hence, the same face in different environments will be more challenging face-matching tasks. The convolutional layers in the CNN model extract high-dimensional (128-dimensional) invariant features for each detected face. The set of photos will be the input. The training model aims to extract the feature vector of each face. In the following, a three-layer training model, called *Conv_FF*, is defined as *Conv_FF=(I-layer, H-layer, O-layer, X, Y)*. Let *Conv_FF=(I-layer, H-layer, O-layer, X, Y)* denote the face recognition feature map extracted by convolutional layers. *I-layer*, *H-layer*, and *O-layer* are the input layer, hidden layer, and output layer, respectively. The X is the training dataset, and Y is the set of feature vectors of faces. To learn face characteristics, all faces $X = (f_{i,j}^1, f_{i,j}^2 \ldots, f_{i,j}^{|F_i|})$ should be taken as the inputs of *Conv_FF* and the set of vectors $V_i = (v_{i,j}^1, v_{i,j}^2 \ldots, v_{i,j}^{|F_i|})$ where $v_{i,j}^k$ represents the face vector of the face $f_{i,j}^k$ in the j-th photo of i-th meeting, $|F_i|$ is the number of feature vectors of all faces in the meeting m_i and V_i is the set of the output labels of X. Herein, it is noticed that $f_{i,j}^k$ and $f_{i,j}^{\hat{k}}$ might belong to the same face. This indicates that the two face vectors $v_{i,j}^k$ and $v_{i,j}^{\hat{k}}$ should be similar. The goal of the *Conv_FF* is to extract the feature vector $v_{i,j}^k$ with 128×1 columns

for each face in the meeting m_i, where 128 represents each feature vector length. Therefore, if feature vectors $v_{i,j}^{k}$ and $v_{i,j}^{\hat{k}}$ are similar, the corresponding faces $f_{i,j}^{k}$ and $f_{i,j}^{\hat{k}}$ should belong to the same category. They will be considered as the same person. Otherwise, they belong to different face categories.

To achieve this identification, the feature vector $v_{i,j}^{k}$ of each face in the meeting m_i will be taken from an output label of the *Conv_FF* to distinguish the unique characteristic of each face. Based on this design, the size of each face image $f_{i,j}^{k}$ should be $X = (256 \times 256 \times 8)$ taken as an input of the *I*-layer while the length of the feature vector $v_{i,j}^{k}$ of the output layer *O-layer* is $Y = (128 \times 1)$. The hidden layers build with more combinations of convolution and max-pooling operations, whose sizes are, *H-layer*$=(256 \times 256 \times 8)$, $(128 \times 128 \times 8)$, and so on, as shown in Fig. 3. The primary purpose of *Conv_FF* layers is to extract all generic features of the images. The main aim of the max-pooling layer is to decrease the number of parameters, reduce the image size, avoid over-fitting as well as reduce computational cost. The resulting size after performing the max-pooling function is shown in Exp. (20). That is,

$$\left\lfloor \frac{n_H - f_s}{s} + 1 \right\rfloor + \left\lfloor \frac{n_w - f_s}{s} + 1 \right\rfloor * n_c \tag{20}$$

where n_H and n_w denote height and width of the face image $f_{i,j}^{k}$, respectively, and f_s, s and n_C be the filter size, stride and channel number respectively. The feature vectors of all faces' F_i outputs of meeting m_i can be represented as $V_i = (v_{i,j}^{1}, v_{i,j}^{2} \ldots, v_{i,j}^{|F_i|}) \in V$, where V is the set of feature vectors of all faces' F outputs of all meetings M.

4.2 Cosine similarity (CS) and similar face grouping

The above-mentioned operations can help the constructed model *Conv_FF* learn the features of faces. After that, the proposed *AFRM* method adopts the cosine function to calculate the difference among extracted high-dimension features and uses them to determine the face group. Recall that $V_i = (v_{i,j}^{1}, v_{i,j}^{2} \ldots, v_{i,j}^{|F_i|})$ is with length 128×1 where $|F_i|$ denotes the number of feature vectors of all faces F_i in meeting m_i. Let the extracted face vector is $v_{i,j}^{k} = (v_{i,j}^{k,1}, \ldots, v_{i,j}^{k,128})$. The cosine similarity is applied as shown in Eq. (21). That is,

$$\begin{aligned} \underset{k \neq \hat{k}}{\text{Cosine similarity}}(v_{i,j}^{k}, v_{i,j}^{\hat{k}}) &= \frac{v_{i,j}^{k} \cdot v_{i,j}^{\hat{k}}}{\left|\left|v_{i,j}^{k}\right|\right| \left|\left|v_{i,j}^{\hat{k}}\right|\right|} \\ &= \frac{\sum_{q=1}^{128} v_{i,j}^{k,q} \times v_{i,j}^{\hat{k},q}}{\sqrt{\sum_{q=1}^{128} (v_{i,j}^{k,q})^2} \times \sqrt{\sum_{q=1}^{128} (v_{i,j}^{\hat{k},q})^2}} \end{aligned} \tag{21}$$

where $(v_{i,j}^{k}, v_{i,j}^{\hat{k}})$ is a pair of two face vectors in V_i. If their cosine similarity is equal to or greater than the predefined *thresholdvalue*, the two faces should belong to the same face vector category. Otherwise, a new face category will be created in V_i. The set of face vector categories is denoted by $g_i = \{g_{i,1}, g_{i,2}, \ldots, g_{i,n}\}$, where n is the number of face vector categories in vectors V_i in meeting m_i.

4.3 Probability distribution (PD)

To map the face to the name in a manner of one-to-one, the proposed *AFRM* applies a name inference algorithm, which depends on the probability distribution of names. The name-based probability distribution generates name lists L in meetings M. The probability of each name will be accumulated according to the number of participants in a meeting list L_i. In addition, the same participant can attend more than one meeting. In this case, the following equation calculates the probability of occurrence for each name. That is,

$$prob(name_{i,l}) = prob(name_{i,l}) + 1/len(L_i) \tag{22}$$

where $prob(name_{i,l})$ is the probability of each name $name_{i,l}$ and i and l are the meeting number and

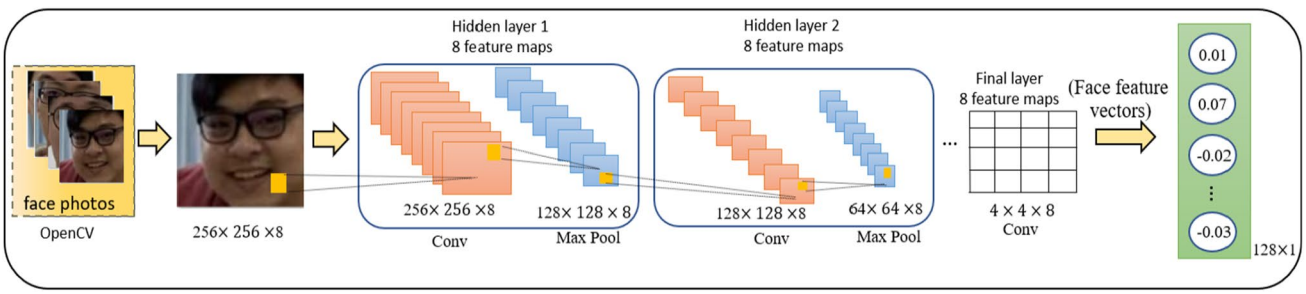


Fig. 3 Face feature extraction *Conv_FF* model

participant number in a name list, respectively, and the $name_{i,l} \in L_i$ are names in a meeting list L_i. Each name probability in a name list is represented by $prob(L_i) = prob(name_{i,1}), prob(name_{i,2}), \ldots, prob(name_{i,l})$ The conceptual diagram of the name inference system is shown in Fig. 4.

4.4 Name-to-face mapping and *KNN* for learning

After determining all face vector categories ing_i, and calculating the probabilities of all names $prob(L_i)$ in the meetingm_i, the proposed *AFRM* can assist face category-to-name mapping according to the number of face vectors in the categories and their name probabilities$prob(name_{i,l})$. Let $G = \{G_1, G_2, \ldots, G_s\}$ denote the set of face vector categories created using cosine similarity by Eq. (21), where s is the total set of face vector categories in all meetings$M$. After applying Eq. (22), assume that the set of name probabilities is: $prob(L) = prob(L_1), prob(L_2), \ldots, prob(L_{|L|})$, where $|L|$ is the total set of probabilities in L. The name with the highest probability is the name tag of the highest number of the face vector category. The face vector categories G_s are sorted in an ascending order depending on the number of face vectors in$G$. According to the probability of each name, the $prob(L)$ list is also sorted in ascending order. The proposed *AFRM* approach has mapped these names and face vector categories according to their positions. Due to the existence of non-participants' faces, the number of categories of face vectors may be more than the number of names which are in the global name list$L$. Hence the proposed method discards those categories of the face vectors with fewer non-participants' faces. The proposed *AFRM* further adopts the *KNN* supervised clustering approach to learn the sets of face vector categories s with the set of the probability of name lists $prob(L)$ to determine the target person's name in the future. The input instances of *KNN* are $G_1, G_2, \ldots, G_s$ with their labels $prob(L_1), prob(L_2), \ldots, prob(L_{|L|})$. The training of the *KNN* clustering model is accomplished when obtained face vectors with their corresponding name labels. In case some new face photos are collected in the future, the face recognition model can infer the face feature groups and determine the names of these faces. Then the proposed *AFRM* directly inputs these data points into the *KNN* model to complete the training process. The proposed *AFRM* can always learn and reinforce itself through new face photos.

The following example shown in Fig. 5 is given to illustrate the execution details of the proposed *AFRM*. Assume that the number of total meetings is $M = 2$. The two meetings are denoted by m_1 and m_2. Many photos $p_{i,j}$ are captured in meetings m_1 and m_2. Those are the first photo $p_{1,1}$, and the second photo $p_{1,2}$ of first meeting m_1 and the first photo $p_{2,1}$, and the second photo $p_{2,2}$ of second meeting m_2. The extracted frontal faces of the meeting m_1 are

$$(f^1_{1,1}, f^2_{1,1}, f^3_{1,1}, f^4_{1,1}) \in p_{1,1},$$
$$\text{and } \left(f^1_{1,2}, f^2_{1,2}, f^3_{1,2}, f^4_{1,2}, f^5_{1,2}, f^6_{1,2}\right) \in p_{1,2}$$

The extracted frontal faces of the meeting m_2 are

$$\left(f^1_{2,1}, f^2_{2,1}, f^3_{2,1}, f^4_{2,1}, f^5_{2,1}, f^6_{2,1}\right) \in p_{2,1}, \text{and} (f^1_{2,2}, f^2_{2,2}, f^3_{2,2}, f^4_{2,2}) \in p_{2,2}.$$

The total frontal faces of the meeting m_1 and m_2 are

$$\left(f^1_{1,1}, f^2_{1,1}, f^3_{1,1}, f^4_{1,1}, f^1_{1,2}, f^2_{1,2}, f^3_{1,2}, f^4_{1,2}, f^5_{1,2}, f^6_{1,2}\right) \in F_1, \text{and}$$
$$\left(f^1_{2,1}, f^2_{2,1}, f^3_{2,1}, f^4_{2,1}, f^5_{2,1}, f^6_{2,1}, f^1_{2,2}, f^2_{2,2}, f^3_{2,2}, f^4_{2,2}\right) \in F_2.$$

Assume that there are two name lists from m_1 and m_2

$$(name_{1,1}, name_{1,2}, name_{1,3}, name_{1,4}, name_{1,5}, name_{1,6}) \in L_1, \text{and}$$
$$(name_{2,1}, name_{2,2}, name_{2,3}, name_{2,4}, name_{2,5}) \in L_2$$

According to the face sets and name lists, the asymmetric condition occurred because the number of frontal faces and

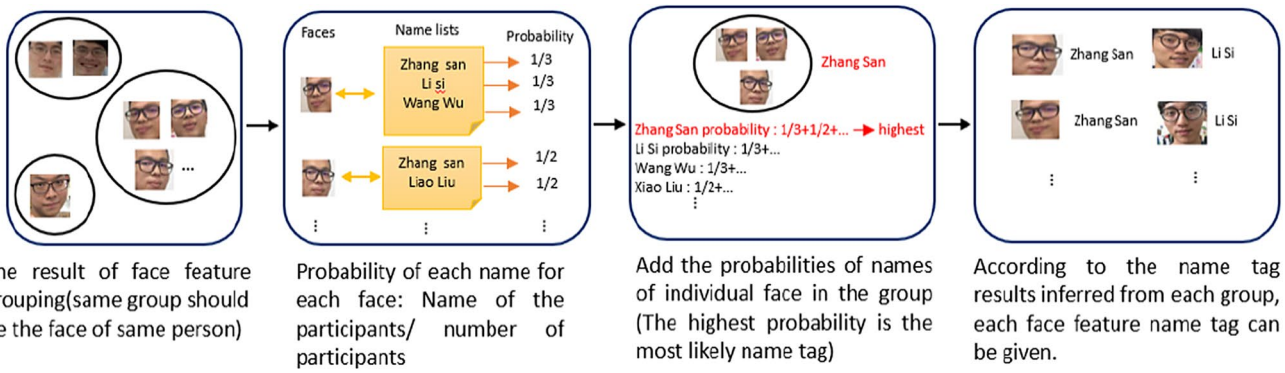


Fig. 4 Conceptual diagram of name inference

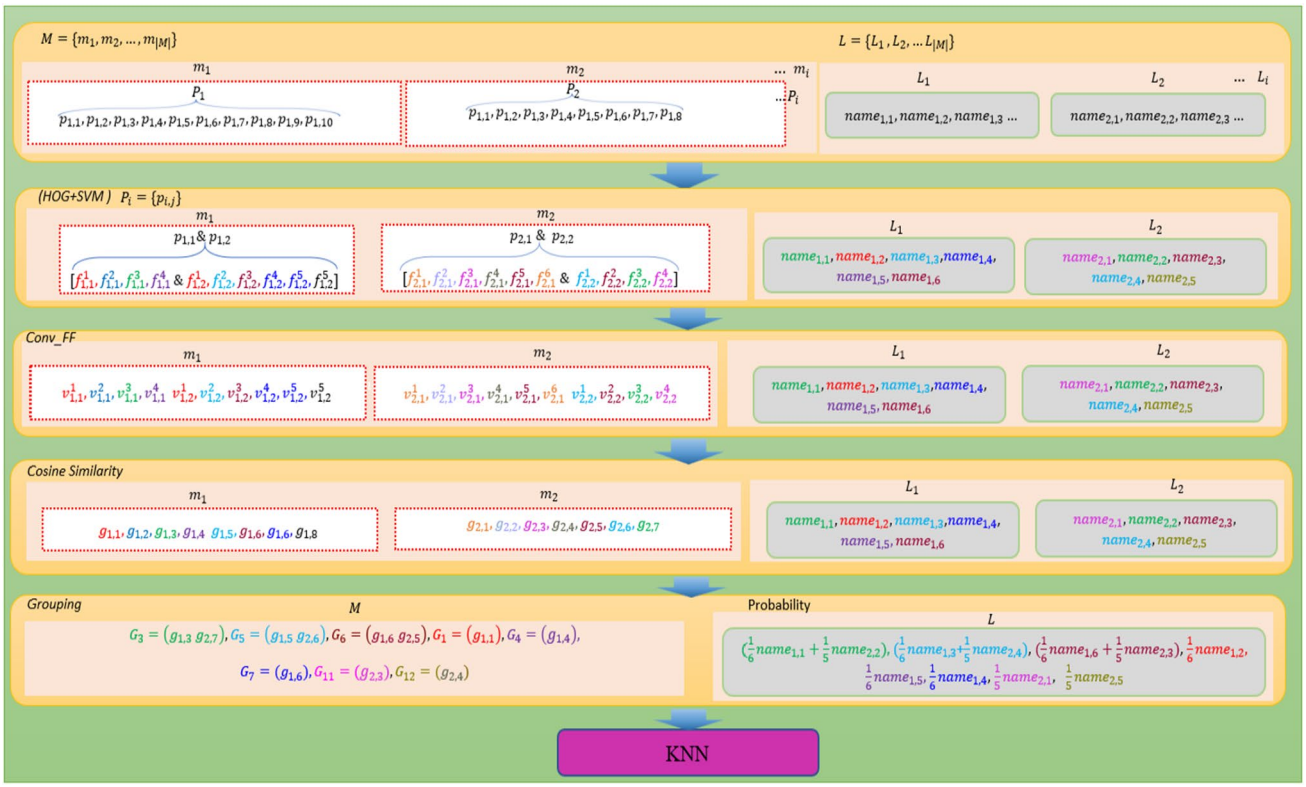


Fig. 5 A flowchart of *Conv_FF-KNN*

participant names are 10 and 6, respectively, in the meeting m_1 and 10 and 5, respectively, in the meeting m_2. The two face sets F_1 and F_2 are collected from four photos $p_{1,1}$, $p_{1,2}$, $p_{2,1}$, and $p_{2,2}$ of two meetings m_1, and m_2, respectively. The training data, $X=20$ frontal face images, are fed as the inputs of the constructed *Conv_FF,* while the length of column vectors 128×1 will be the corresponding output label Y. The extracted face vectors in meetings m_1 and m_2 are

$$\left(v_{1,1}^1, v_{1,1}^2, v_{1,1}^3, v_{1,1}^4, v_{1,2}^1, v_{1,2}^2, v_{1,2}^3, v_{1,2}^4, v_{1,2}^5, v_{1,2}^6\right) \in V_1, \text{and}$$
$$\left(v_{2,1}^1, v_{2,1}^2, v_{2,1}^3, v_{2,1}^4, v_{2,1}^5, v_{2,1}^6, v_{2,2}^1, v_{2,2}^2, v_{2,2}^3, v_{2,2}^4\right) \in V_2.$$

After extracting the face feature vectors, cosine similarity is calculated by Eq. (21) to create categories of the same faces $f_{i,j}^k$ with corresponding face vectors of two meetings to know the duplicate person. The same colors indicate the same categories. The face vector categories are in m_1 and m_2:

$$\left(g_{1,1}, g_{1,2}, g_{1,3}, g_{1,4}, g_{1,5}, g_{1,6}, g_{1,7}, g_{1,8}\right) \in g_1, \text{and}$$
$$\left(g_{2,1}, g_{2,2}, g_{2,3}, g_{2,4}, g_{2,5}, g_{2,6}, g_{2,7}\right) \in g_2.$$

All face vector categories are in all meetings M :

$$\{G_1 = (g_{1,1}), G_2 = (g_{1,2}), G_3 = (g_{1,3}g_{2,7}), G_4 = (g_{1,4}),$$
$$G_5 = (g_{1,5}g_{2,6}), G_6 = (g_{1,6}g_{2,5}), G_7 = (g_{1,7}), G_8 = (g_{1,8}),$$
$$G_9 = (g_{2,1}), G_{10} = (g_{2,2}), G_{11} = (g_{2,3}), G_{12} = (g_{2,4})\} \in G.$$

After that, two name lists are taken from meetings m_1 and m_2:

$$\left(name_{1,1}, name_{1,2}, name_{1,3}, name_{1,4}, name_{1,5}, name_{1,6}\right)$$
$$\in L_1, \text{ and } \left\{name_{2,1}, name_{2,2}, name_{2,3}, name_{2,4}, name_{2,5}\right\} \in L_2.$$

From these two name lists, it knows that three participants have attended both meetings. They are

$$name_{1,1} = name_{2,2}, name_{1,6}$$
$$= name_{2,3}, \text{ and } name_{1,3} = name_{2,4}.$$

Hence, the probability of occurrence of each name is

$$
\begin{aligned}
prob(name_{1,1}orname_{2,2}) &= \left(\frac{1}{6}+\frac{1}{5}\right), prob(name_{1,2}) \\
&= \left(\frac{1}{6}\right), prob(name_{1,3}orname_{2,4}) \\
&= \left(\frac{1}{6}+\frac{1}{5}\right), prob(name_{1,4}) \\
&= \left(\frac{1}{6}\right), prob(name_{1,5}) \\
&= \left(\frac{1}{6}\right), prob(name_{1,6}orname_{2,3}) \\
&= \left(\frac{1}{6}+\frac{1}{5}\right), prob(name_{2,1}) \\
&= \left(\frac{1}{5}\right), prob(name_{2,5}) = \left(\frac{1}{5}\right)
\end{aligned}
$$

which is calculated by Eq. (22). The face vectors tag with their corresponding names according to similar face vectors in categories and the probability value of the names. Those are

$$
\begin{aligned}
&\{G_3 = (g_{1,3}g_{2,7}), prob(name_{1,1}orname_{2,2}) = \left(\frac{1}{6}+\frac{1}{5}\right)\}, \\
&\{G_5 = (g_{1,5}g_{2,6}), prob(name_{1,3}orname_{2,4}) = \left(\frac{1}{6}+\frac{1}{5}\right)\}, \\
&\left\{G_6 = (g_{1,6}g_{2,5}\right), prob(name_{1,6}orname_{2,3}) = \left(\frac{1}{6}+\frac{1}{5}\right)\}, \\
&\{G_1 = \left(g_{1,1}\right), prob(name_{1,2}) = \left(\frac{1}{6}\right)\}, \\
&\{G_4 = (g_{1,4}), prob(name_{1,5}) = \left(\frac{1}{6}\right)\}, \\
&\{G_7 = (g_{1,7}), prob(name_{1,4}) = \left(\frac{1}{6}\right)\}, \\
&\{G_{11} = (g_{2,3}), prob(name_{2,1}) = \left(\frac{1}{5}\right)\}, \\
&\{G_{12} = (g_{2,4}), prob(name_{2,5}) = \left(\frac{1}{5}\right)\}.
\end{aligned}
$$

Those face vector categories feed into the *KNN* cluster with their names for learning. The rest of the face vector categories are

$$G_2 = (g_{1,2}), G_8 = (g_{1,8}), G_9 = (g_{2,1}), \text{and} G_{10} = (g_{2,2})$$

Those vector categories cannot map with names because they are not participants in any meeting. They cannot enter the *KNN* cluster. The result of the *KNN* cluster determines the name of the face.

Based on the above-mentioned designing concept, the proposed *AFRM* training algorithm has presented step by step in Table 3. The main target of the proposed algorithm is to extract all faces and find out the names of all participants of all meetings$M$. At the beginning, all photos and all name lists are collected from all meetings. In step 1, each meeting m_i considers in the meeting set ($m_i \in M$,$i = 1, 2, \ldots M$). Further, step 2 considers each photo $p_{i,j}$ in the i-th meeting ($p_{i,j} \in m_i$, $p_{i,j} = 1, 2, \ldots |P_i|$). All frontal faces F_i are extracted by algorithms HOG and SVM. Each face $f_{i,j}^k$ is extracted from photos $p_{i,j}$ in the i-th meeting. All extracted faces of each i-th meeting are stored in F_i ,($F_i = \bigcup_{p_{i,j} \in P_i} \bigcup_{f_{i,j}^k \in p_{i,j}} f_{i,j}^k$). Steps 3 to 5 extract feature vectors of all faces of the i-th meeting by *Conv_FF* and store them in V_i, $V_i = \bigcup_{v_{i,j}^k} Conv\_FF(f_{i,j}^k)$ where ,$f_{i,j}^k = 1, 2, \ldots |F_i|$. In steps 6 to 17, the face vector category $g_{i,n} = \varnothing$ is initialized with n=0 in meetingm_i. The similar face vectors are grouped by $groupby\left(g_{i,n}, v_{i,j}^k\right)$ and a face category is newly created in the meeting m_i for the high-dimensional face vector $v_{i,j}^k$. The face category is dynamically incremented. Next, the threshold value is calculated for grouping all face vectors. If the θ is greater than the predefined threshold value, this indicates that two face vectors might be the same. Therefore, they should belong to the same category. Otherwise, a new category will be newly created for that face in the meetingm_i. In steps 18 to 30, similarly the face vector categories $G_s \in G$ are created by $groupby\left(G_s, g_{i,n}\right)$ in meeting$M$. The face vector categories are sorted in G in step 31. After learning all face characteristics and face categories, all name lists L in all meetings M and each name $name_{i,l}$ in name list L_i are considered in steps 32 to 34. Steps 35 and 36 calculate the probability of each name $name_{i,l}$ in L_i in i-th meeting. According to the probability value of each name sorted $prob\left(name_{i,l}\right)$ in L as shown in step 37. Step 38 maps face vector categories with names by probability distribution algorithm and one-to-one name tags have been employed into *KNN* cluster for learning. The *AFRM* algorithm helps obtain the name of each face. Table 4 summarizes the common notations used in this paper.

5 Performance evaluation

This section mainly describes the performance improvement of the proposed *AFRM* algorithm for asymmetric face identification. The proposed algorithm collects name lists and historically stored face photos, which can determine the name of the target person. After extracting all faces by HOG and SVM algorithms, the proposed *AFRM* algorithm assigns feature vectors for each face of past meetings based on the *Conv_FF* and *KNN* clustering algorithms. The training of the face vector is mainly based on extracting the faces with names. It is guaranteed that the similarity of two face vectors indicates the same participant present in the meeting. Then, the two face vectors are used to identify the same person in the meeting.

Accuracy is compared with the proposed *AFRM* and existing algorithms. The experimental configuration of the

Table 3 The algorithm proposes the face vector similarity and name probability

Algorithm: ***Face and Name Mapping***

Purpose: The *Conv_FF and KNN clustering* algorithm should learn the names $name_{i,l}$ of each face $f_{i,j}^{k}$ for $\forall\, name_{i,l} \ni L_i, \forall f_{i,j}^{k} \ni p_{i,j}$, $KNNclusters = 3$.

Input: Number of meetings M, Number of name lists L, and Set of photos P captured in meetings

Output: Based on facial features, assign the name of each participant

For $i = 1$ to M:
  $F_i = \bigcup_{p_{i,j} \in P_i} \bigcup_{f_{i,j}^{k} \in p_{i,j}} f_{i,j}^{k}$
For $i = 1$ to M:
  For each $f_{i,j}^{k}$ in F_i:
    $V_i = \bigcup_{v_{i,j}^{k}} Conv\_FF(f_{i,j}^{k})$
For each V_i in V:
  $n = 0$
  $g_{i,n} = \emptyset$
  For each $v_{i,j}^{k}$ in V_i :
    if $g_{i,n} == \emptyset$
      $g_{i,n} = groupby(g_{i,n}, v_{i,j}^{k})$
    for each $g_{i,n}$ in g_i:
      $\theta = Cosine\ similarity(g_{i,n}, v_{i,j}^{k})$
      $if\ \theta \geq thresholdvalue$
        $g_{i,n} = groupby(g_{i,n}, v_{i,j}^{k})$
      else
        $g_{i,n+1} = groupby(g_{i,n+1}, v_{i,j}^{k})$
$s = 0$
$G_s = \emptyset$
For $i = 1$ to M:
  For each $g_{i,n}$ in g_i:
    if $G_s == \emptyset$
      $G_s = groupby(G_s, g_{i,n})$
    else
      for each G_s in G:
        $\theta = Cosine\ similarity(G_s, g_{i,n})$
        $if\ \theta \geq thresholdvalue$
          $G_s = groupby(G_s, g_{i,n})$
        $else$
          $G_{s+1} = groupby(G_{s+1}, g_{i,n})$
$G = sort(G)$
For $i = 1$ to M:
  $L_i = L_i \cup name_{i,l}$
For $i = 1$ to L:
  For $name_{i,l}$ in L_i:
    $prob(name_{i,l}) = prob(name_{i,l}) + 1/len(L_i)$
$sort(prob(L)$
$Mapping = KNN(G, prob(L))$

proposed method is described in Table 5. The proposed model runs in the TensorFlow deep learning framework (1.13) to evaluate the performance using python (2.7). The experiment is conducted in the ubuntu 16.04 environment with CPU i5-8400.

We created an in-house meta dataset to evaluate the efficiency of the proposed *AFRM* algorithm. The in-house meta dataset is collected from 98 project meetings in the past years from a system-integrated information company in Taiwan. We collected numbers of photos ranging from

Table 4 Summary of common notations

Notations	Meanings
M	All meetings
m_i	Each meeting in M
P	All photos in all meetings M
P_i	The set of photos collected in the i-th meeting
$p_{i,j}$	j-th photo in meeting m_i
F	All faces in all meetings M
F_i	The set of faces in the i-th meeting
$f_{i,j}^k$	k-th face in j-th photo in the i-th meeting
V	Face vector of all faces in all meetings M
V_i	The set of face vectors in the i-th meeting
$v_{i,j}^k$	k-th face vector in j-th photo in the i-th meeting
$g_{i,n}$	Set of face categories in g_i for the i-th meeting
G_s	Set of face categories in G for all meeting M
L	All name list in all meeting M
L_i	The name list in the i-th meeting
$name_{i,l}$	The l-th name in a list L_i in the i-th meeting
$prob(name_{i,l})$	The l-th name probability in a list L_i in the i-th meeting

Table 5 The experimental settings

Parameters	Values
CPU	i5-8400
Ubuntu	16.04
Python	2.7
Tensorflow	1.13
Tool	Dlib
Photo extract library	OpenCV
Conv_FF	128-dimensional vector
Cosine similarity	$>=0.85$
KNN cluster	3

5 to 10 from all meetings with attendance lists. We have adopted photos from each meeting to train the proposed model and evaluate the effectiveness of the proposed *AFRM* algorithm.

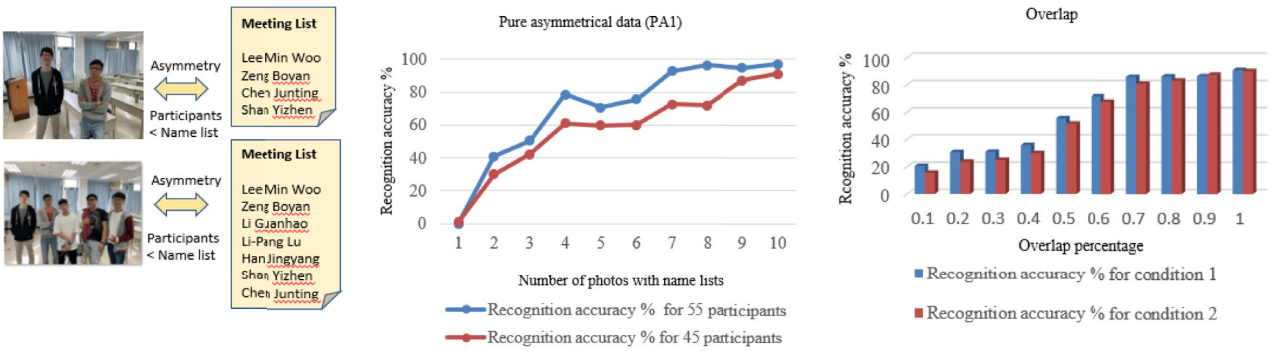
Since most of the related studies only can handle the symmetrical dataset, we prepare five cases of the dataset, aiming to compare the related studies with the proposed *AFRM* which can handle not only the symmetrical dataset but also the asymmetrical dataset. Recall that L_i denotes the name list of the meeting m_i. Let $|L_i|$ denote the number of names in the name list L_i. We also recall that $f_{i,j}^k$ denotes the k-th face in the j-th photo of the meeting m_i. It is noticed that in the same meeting, the numbers of faces in different photos are identical. Recall that F_i denotes the set of all photos in the meeting m_i. Let $|F_i|$ denote the number of faces in each photo collected from the meeting m_i. The following presents the considered five cases of datasets:

1. Completely symmetrical case (CS) ($|L_i|=|F_i|$).
2. Pure asymmetrical case 1 (PA1) ($|L_i|>|F_i|$).
3. Pure asymmetrical case 2 (PA2) ($|L_i|<|F_i|$).
4. Combination of asymmetrical cases 1 and 2 (CA) (combinations of $|L_i|>|F_i|$ case and $|L_i|<|F_i|$ case).
5. Combination of CS and PA case (CSA) (combinations of $|L_i|=|F_i|$ case and $|L_i|\neq|F_i|$ case).

The proposed *AFRM* algorithm considers only two cases among five cases. Those are pure asymmetrical case 1 (PA1) and pure asymmetrical case 2 (PA2). The experiment results are shown as follows:

Pure asymmetrical case 1 (PA1)

Figure 6 considers the pure asymmetrical training data and compares the performance of two datasets in Pure Asymmetrical Case 1 (PA1). Figure 6a shows a pure



(a) Pure asymmetrical training data

(b) Relationship between the number of meetings and the correct rate of the face recognition model

(c) Overlap

Fig. 6 Pure asymmetrical training data, compared the performance of two datasets, and overlap in pure asymmetrical case 1 (PA1)

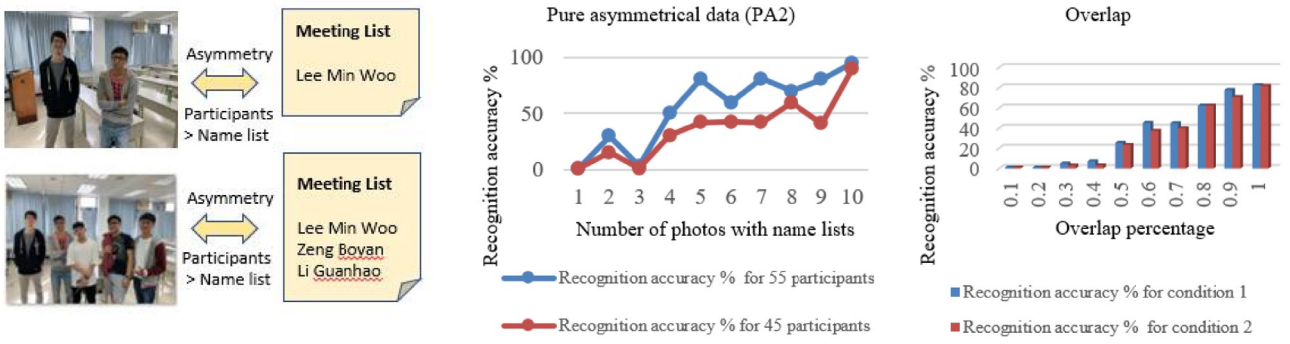
asymmetric training dataset which satisfies the condition $|L_i| > |F_i|$. To satisfy this condition, the proposed *AFRM* checks the name list and faces in photos and selects some data from 98 project meetings to organize the dataset, which guarantees the *PA1* condition. That is, the number of participants' names in the name list L_i is larger than that of faces F_i faces in each photo collected in the corresponding meeting m_i.

In Fig. 6b, the X-axis indicates the number of photos with corresponding name lists in all meetings, ranging from 1 to 10 while the Y-axis indicates the success rate of the proposed *AFRM* for different datasets. The two datasets are utilized, aiming to know the efficiency of the proposed *AFRM*. The resultant outcomes of two datasets 55 (indicated by the blue line in graphs) and 45 (indicated by the orange line in graphs) participants are represented by condition 1 and condition 2, respectively. In comparison, the proposed *AFRM* has achieved 97.1% and 91% accuracies for conditions 1 and 2, respectively. A common trend shows that if the number of participants increases, the face recognition accuracy can be improved. The major reason is that when the number of extracted faces is less than the number of participants, each extracted face can be assigned with different names easily. In this case, the proposed *AFRM* creates a cosine similarity gap between actual and fake assignments to achieve the highest recognition accuracy.

Figure 6c considers the overlapped condition for the PA1 case, which can help measure the accuracy by varying the number of participants present in more than one meeting. The X-axis indicates the overlap percentage, varying from 0.1 to 1.0 in all meetings. The overlap percentage is calculated by the number of duplicate participants divided by the total number of participants. The Y-axis indicates the success rate of the proposed *AFRM* model. The proposed *AFRM* model generates two bars in the graph. The blue bar displays the face recognition accuracy for condition 1 (in 55 participants), which is 90.9%. The orange bar indicates the face recognition accuracy for condition 2 (in 45 participants), which is 90.1%. It shows that the face recognition accuracies will increase if the overlap percentages increase. The primary reason is that the number of duplicate participants will increase when the overlap percentage increases. As a result, the proposed *AFRM* can assign faces with names easily. In summary, the proposed *AFRM* algorithm achieves better performance in the PA1 case. The reason is that the number of participants' names is greater than that of the participants' photos.

Pure asymmetrical case 2 (PA2)

Figure 7 describes the pure asymmetrical training data and compares the performance of two datasets in Pure Asymmetrical Case 2 (PA2). Figure 7a shows a pure asymmetrical dataset which satisfies the condition $|L_i| < |F_i|$. According to this state, the participant names in the meeting list L_i must be less than the number of faces F_i in photos in the meeting m_i. In Fig. 7b, the X-axis indicates the number of meeting photos with corresponding name lists in all meetings, ranging from 1 to 10. The Y-axis indicates the success rate of two different datasets. In the experiment, two datasets, conditions 1 (including 55 participants) and 2 (including 45 participants) are considered. The resultant outcomes of two datasets, conditions 1 and 2 are indicated by the blue line and the orange line, respectively, in the graph. In comparison, the accuracies of both datasets i.e., conditions 1 and 2, are 95.4% and 90.0%, respectively. Although the accuracy of this pure asymmetrical condition PA2 is slightly lower than pure asymmetrical PA1, this situation frequently arises in most real cases. The reason is that the number of participants is less than the extracted faces. As a

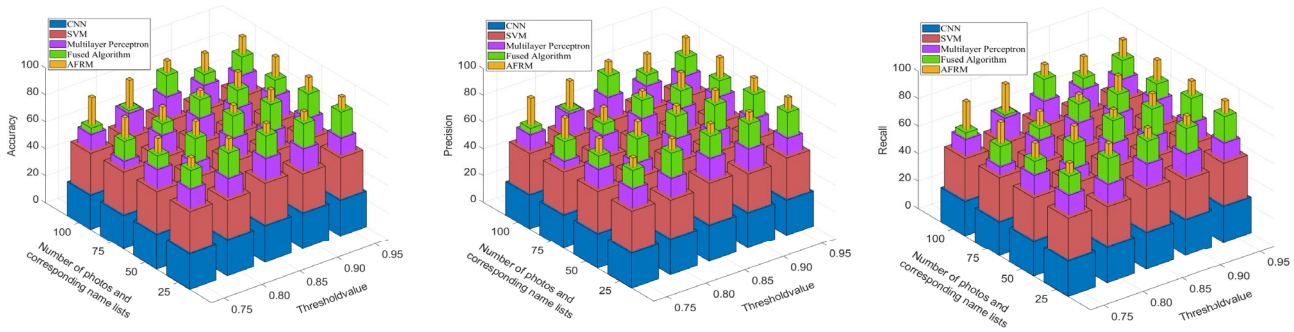


(a) Pure asymmetrical training data

(b) Relationship between the number of meetings and the correct rate of the face recognition model

(c) Overlap

Fig. 7 Pure asymmetrical training data, compared the performance of two datasets, and overlap in pure asymmetrical case 2 (PA2)



(a)Accuracy (b) Precision (c) Recall

Fig. 8 Performance of five compared algorithms in terms of accuracy, precision, and recall

Table 6 Accuracy comparison in different methods

Methods	Accuracy (%)
Support vector machine (SVM) [24]	58.7
Multilayer perceptron (MLP) [24]	74.6
Convolution neural network (CNN) [24]	28.8
Fused algorithm [24]	89.4
The proposed method (*AFRM*)	97.5

result, fewer classes are created, which leads to the situation that different participants' faces may be assigned to the same category, reducing the recognition accuracy.

Figure 7c describes the results of the overlap condition in the PA2 case, which helps measure duplicate participants in all meetings. The X-axis indicates the overlap percentage, varying from 0.1 to 1.0 in all meetings. The overlap percentage is calculated by the following formula: the number of duplicate participants/ total number of participants. The Y-axis indicates the success rate of the proposed *AFRM* model. The proposed *AFRM* model generates two bars in the graph. The blue bar indicates the face recognition accuracy for condition 1 (in 55 participants), which is 82.6%. The orange bar indicates the face recognition accuracy for condition 2 (in 45 participants), which is 81.9%. In summary, the proposed *AFRM* achieves better results, but recognition accuracies are slightly lower than PA1. The main reason is that the number of participants' names is smaller than that of the participants' photos.

Accuracy comparison in different traditional methods with various conditions

Figure 8 mainly compares the performances of the proposed *AFRM* and the existing algorithms in terms of accuracy, prediction and recall. To calculate the accuracy, the face recognition system is compared in different conditions (number of participants) with different parameters (threshold value). The performances of CNN, SVM, MP, Fused Algorithm, and proposed *AFRM* are obtained by setting the number of photos with their corresponding name lists in all meetings. The number of photos varies from 25 to 100, while the threshold value varies, ranging from 0.75 to 0.95. The experimental results show that the accuracies are changed depending on photo numbers and threshold values. When the number of photos increases, the accuracy values will increase. The main reason is that the number of faces is increased with the number of photos. In addition, the number of duplicate faces also increases. The values of the face vectors will be scattered in more dimensions, leading to low similarity even though it is suitable for grouping faces. On the other hand, a large threshold value will discard participants who are suitable for the same face group, leading to low accuracy. In summary, accuracy has reached better among different conditions when the threshold value is 0.85. The experimental results show the common trend that the performances of the proposed *AFRM* and four existing mechanisms are increased with the number of faces, in terms of accuracy, precision and recall. This is because the number of training data increases with the number of faces. As a result, it helps improves accuracy, precision as well as recall, which are shown in Fig. 8a–c, respectively.

The *AFRM* method has been compared with SVM, MLP, and CNN to verify its effectiveness in Table 6. In comparison, the accuracy, precision, and recall of the existing CNN, SVM, MP, and Fused are smaller than the proposed *AFRM*. As a result, the accuracy of the *AFRM* algorithm is 97.5%. The accuracies of the SVM, MLP, CNN, and Fused algorithms are 58.7%, 74.6%, 28.8%, and 89.4%, respectively. The results demonstrate that the proposed *AFRM* algorithm is better than the other four algorithms, and it is very potential to contribute to the future development of real-time face recognition systems.

6 Conclusion

Generally speaking, after each meeting, it is very common for meeting members to take pictures as souvenirs. However, it is difficult to map each name to each face in the photo since the name list and the faces in the photo are asymmetric. Given the name lists and the collected photos from past meetings, this paper proposed an *AFRM* algorithm aiming to map each name to each face accurately. Unlike the existing studies, this paper takes the benefits of feature vector extraction, cosine similarity calculation, probability estimation as well as grouping, which contribute significantly to face recognition for the systematic dataset. The proposed *AFRM* is mainly composed of four phases. In the *initial phase*, all face images are extracted from meeting photos, and then face vectors are extracted from all face images. Then the *grouping phase* calculates the cosine similarity to know the duplicate participants among all face vectors. Next, the *statistic mechanism phase* is adopted to calculate the name probability of each name in the name lists. Finally, the *mapping phase* tags the names with face vector categories according to probability values by the name inference algorithm and feeds the face vector categories with names into the *KNN* clustering mechanism for training. After training, participants' names instantly come out if new photos are fed into the recognition model. The performance study reveals that the proposed *AFRM* algorithm performs better than the existing algorithms in real-time face recognition cases. Future work will consider the companies' databases to improve face recognition accuracy. As a result, the system will provide the name of the participant's face photo with yield background information such as contact information, cooperation cases, and experience in the meeting.

Author contributions All the authors have equally contributed, and all the authors have read and agreed to the published version of the manuscript.

Funding This study was not funded by any institution.

Data availability The datasets generated during the current study are not publicly available but are available from the corresponding author.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This study does not involve either human subjects or animals.

Informed consent All participating authors have been informed.

References

1. Damale, R. C. and Pathak, B. V.: Face recognition based attendance system using machine learning algorithms. In: 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, pp. 414–419 (2018)
2. Wang, H. and Guo, L.: Research on face recognition based on deep learning. In: 2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM), IEEE, pp. 540–546 (2021)
3. Haji, S. and Varol, A.: Real time face recognition system (RTFRS). In: 2016 4th International Symposium on Digital Forensic and Security (ISDFS), pp. 107–111 (2016)
4. Stekas, N. and Heuvel, D.: Face recognition using local binary patterns histograms (LBPH) on an FPGA-Based System on Chip (SoC). In: 2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Chicago, pp. 300–304 (2016)
5. Ghorbel, A., Tajouri, I., Elaydi, W. and Masmoudi, N.: The effect of the similarity measures and the interpolation techniques on fractional eigenfaces algorithm. In: 2015 World Symposium on Computer Networks and Information Security (WSCNIS), IEEE, pp. 1–4 (2015)
6. Wan, M.H., Lai, Z.H.: Generalized discriminant local median preserving projections (GDLMPP) for face recognition. Neural Process **49**(3), 951–963 (2019)
7. Rose, J., and Bourlai, T.: On designing a forensic toolkit for rapid detection of factors that impact face recognition performance when processing large scale face datasets. Securing Social Identity in Mobile Platforms. Springer, pp. 61–76 (2020)
8. Bao, J., Wang, Y., Cheng, Y.: Asymmetry label correlation for multi-label learning. Appl. Intell. **52**(6), 6093–6105 (2022)
9. Zhang, X., Yang, T., and Srinivasan, P.: Online asymmetric active learning with imbalanced data. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2055–2064 (2016)
10. Zheng, L., Yang, Y. and Hauptmann, A.G.: Person reidentification: Past, present and future. arXiv preprint arXiv:1610.02984 (2016).
11. Fu, B., Chen, C., Henniger, O., and Damer, N.: The relative contributions of facial parts qualities to the face image utility. In: 2021 International Conference of the Biometrics Special Interest Group (BIOSIG), IEEE, isbn. 978-1-66542-693-0, pp. 1–5, (2021)
12. Fu, B., Chen, C., Henniger, O., and Damer, N.: A deep insight into measuring face image utility with general and face-specific image quality metrics. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 905–914 (2022)
13. Chen, K., Yi, T., Lv, Q.: LightQNet: lightweight deep face quality assessment for risk-controlled face recognition. IEEE Signal Process. Lett. **28**, 1878–1882 (2021)
14. Meng, Q., Zhao, S., Huang, Z. and Zhou, F.: Magface: a universal representation for face recognition and quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14225–14234 (2021)
15. Ge, Y., Zhu, F., Chen, D., Zhao, R., Li, H.: Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. Adv. Neural. Inf. Process. Syst. **33**, 11309–11321 (2020)
16. Lin, Y., Dong, X., Zheng, L., Yan, Y., Yang, Y.: A bottom-up clustering approach to unsupervised person reidentification. Proc. AAAI Conf. Artif. Intell. **33**(01), 8738–8745 (2019)
17. Fan, H., Zheng, L., Yan, C., Yang, Y.: Unsupervised person reidentification: clustering and fine-tuning. ACM Trans. Multimed. Comput. Commun. Appl. **14**(4), 1–18 (2018)
18. Lin, Y., Xie, L., Wu, Y., Yan, C. and Tian, Q.: Unsupervised person reidentification via softened similarity learning. In Proceedings of

the IEEE/CVF conference on computer vision and pattern recognition, pp. 3390–3399 (2020)
19. Zeng, K., Ning, M., Wang, Y., and Guo, Y.: Hierarchical clustering with hard-batch triplet loss for person reidentification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13657–13665 (2020)
20. Wang, D. and Zhang, S.: Unsupervised person reidentification via multi-label classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10981–10990 (2020)
21. Cheng, E.J., Chou, K.P., et al.: Deep sparse representation classifier for facial recognition and detection system. Pattern Recogn. Lett. **125**, 71–77 (2019)
22. Zangeneh, E., Rahmati, M., Mohsenzadeh, Y.: Low resolution face recognition using a two-branch deep convolutional neural network architecture. Expert Syst. Appl. **139**, art. no. 112854 (2020)
23. Nguyen, D.T., Pham, T.D., Lee, M.B., Park, K.R.: Visible-light camera sensor-based presentation attack detection for face recognition by combining spatial and temporal information. Sensors **19**(2), 410 (2019)
24. He, G. and Jiang, Y.: Real-time Face Recognition using SVM, MLP and CNN. In 2022 International Conference on Big Data, Information and Computer Network (BDICN) . IEEE, pp. 762–767 (2022)