Student Attendance System in Crowded Classrooms using a Smartphone Camera

Domingo Mery
Universidad Católica de Chile
domingo.mery@uc.cl

Ignacio Mackenney
Universidad Católica de Chile
imackenney@uc.cl

Esteban Villalobos
Universidad Católica de Chile
egvillalobos@uc.cl

Abstract

To follow the attendance of students is a major concern in many educational institutions. The manual management of the attendance sheets is laborious for crowded classrooms. In this paper, we propose and evaluate a general methodology for the automated student attendance system that can be used in crowded classrooms, in which the session images are taken by a smartphone camera. We release a realistic full-annotated dataset of images of a classroom with around 70 students in 25 sessions, taken during 15 weeks. Ten face recognition algorithms based on learned and handcrafted features are evaluated using a protocol that takes into account the number of face images per subject used in the gallery. In our experiments, the best one has been FaceNet, a method based on deep learning features, achieving around 95% of accuracy with only one enrollment image per subject. We believe that our automated student attendance system based on face recognition can be used to save time for both teacher and students and to prevent fake attendance.

1. Introduction

To follow the attendance of students is a major concern in many educational institutions. The manual management of the attendance sheets is laborious and tedious for crowded classrooms. In our experience, the time invested for this task in a 70-student classroom is about 4 minutes, *i.e.* in the whole semester the total invested time could be longer than the duration of one lecture of 80 minutes. We believe that an automated student attendance system –based on face recognition– can be used to save time for both teacher and students, and to prevent fake attendance. This system could be part of a *Next Generation Smart Classrooms*[55], in order to improve teaching and learning experience in the classroom.

In face recognition, the task is to identify a subject appearing in an image as a unique individual. Over the last decade, we have witnessed tremendous improvements in face recognition algorithms. Some applications, that might have been considered science fiction in the past, have become reality now. However, it is clear that face recognition, is far from perfect when tackling more challenging images such as faces taken in unconstrained environments *e.g.* face images acquired by long-distance cameras. Although innovative methods in computer vision have improved the state of the art, the performance obtained in low-quality images remains unsatisfactory for many applications. This has been the case of face recognition in crowded classrooms for a student attendance system as illustrated in Fig. 1.

In this paper, we propose an automated student attendance system that can be used in crowded (and small) classrooms. In this application, after an enrollment stage, in which a face image of each student is acquired and the corresponding ID is registered, the user, *e.g.* the teacher, can take one or several pictures of the classroom using his/her smartphone in order to capture all students that are present. The proposed algorithm detects the faces in the picture(s) and recognize which students are present or absent in order to record the attendance of the class.

The main contributions of the paper are the following:

- A full-annotated dataset of images of a classroom with 67 students in 25 sessions, taken by a smartphone camera during 15 weeks. An example is shown in Fig. 1.
- A simple method based on known deep learning models implemented in Python that can be used as Student Attendance System.
- An evaluation protocol that takes into account the number of face images per subject in the gallery to compute the average accuracy in 25 sessions.
- A comparison of ten different face recognition methods in this task.

All enrollment images, session images, cropped face images, extracted descriptors and code are available at our webpage.

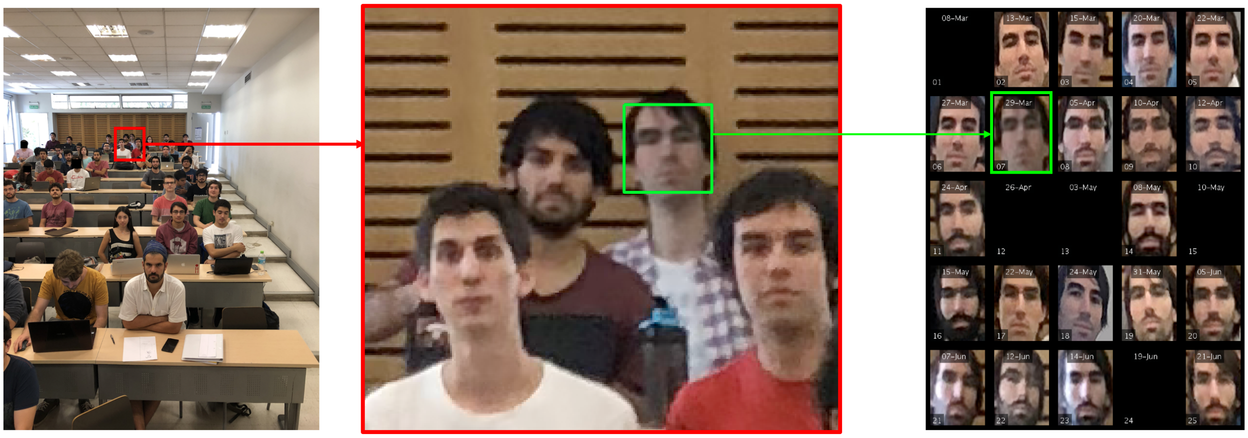


Figure 1: Results of our Student Attendance System. In this example of 25 sessions, the selected student was not present in 5 sessions (see black squares in the right image), *i.e.*, the attendance for him is 20/25 = 80.0%. In the middle image, the session of March 29th is shown, where the student was recognized in the last row of the classroom (see zoom of the red square in the left panel). In these experiments, the images were acquired using a smartphone camera.

The rest of the paper is organized as follows. In Section 2, a literature review in face recognition in low-quality images and in student attendance systems is presented. In Section 3, the proposed method is explained in further detail. In Section 4, the experiments, results and implementation are presented. Finally, in Section 5, concluding remarks are provided.

2. Related Work

In this Section, the literature review is focused on face recognition in low-quality images and student attendance systems.

2.1. Face Recognition in Low-quality Images

Over the last decade, face recognition algorithms shifted to deal with unconstrained conditions. In recent years, we have witnessed tremendous improvements in face recognition by using complex deep neural network architectures trained with millions of face images [53][43] [50][63][33][45][1]. In this Section, however, the focus is on face recognition in low-quality images [36].

It is clear that image degradation, like blurriness, affects the performance of face recognition algorithms based on deep learning [30][14][13]. Image enhancement or considering of degraded samples in training dataset could lead to better deep learning models [20]. In order to cast the problem of face recognition in low-quality images, a straightforward approach is to estimate a high-quality image from one that is low-quality. Face recognition is then performed as normal with high-quality face images. This process involves image restoration techniques in the case of blurred images, and super-resolution techniques in the case of low-resolution images. Among image restoration techniques, we can identify blind deconvolution [35], non-blind deconvolution [65], regularization methods on total variation [48], and Tikhonov regularization [54]. In addition, there are more direct methods based on features of the image that are invariant to blurriness, such as processing images in spatial and frequency domains [15][18]. Nevertheless, the level of restoration is not satisfactory enough for severe blurriness. Conversely, super-resolution techniques, known as *face-hallucination* for low-resolution face images [4], attempt to estimate a high-resolution face image from one that is low-resolution. We can identify techniques based on sparse representations [62], patch-oriented strategies [7] and deep learning features [58], among others. Unfortunately, these methods do not obtain an adequate reconstruction of the high-quality face image when the resolution of the input image is very low, *e.g.* less than 22×15 pixels.

Novel methods that do not follow the above mentioned straightforward approach have been proposed in recent years. Some of these have attempted to perform face recognition by simultaneously computing super-resolution and feature extraction so as to measure the low and high frequencies of the face images [24]. Other methods extract features from face images in resized formats [25][34]. Finally, there are also methods that construct a common feature space (called *inter-feature space* [59]) for matching between low- and high-resolution features [42][5][47][60][23]. Although these innovative methods have improved the state of the art, the performance obtained in low-quality images remains unsatisfactory for many applications such as forensics and video surveillance.

[1] For a literature review of face recognition, see for example [61][39], among others.

Table 1: Related works of automated attendance systems

Reference	Year	Approach	Input	Number of subjects in the query image	Number of enrolled subjects	Size of query face image	Accuracy^{+}	Sessions
Kar et al. [29]	2012	Eigen-faces	Still images	1	10	50×50	95.0%	–
Chintalapati et al. [9]	2013	LBP, PCA, SVM	Still images	2	80	100×100	78.0%	–
Wagh et al. [56]	2015	Eigen-faces	Still images	–	–	–	–	–
Lukas et al. [40]	2016	DWT+DCT	Still images	13	16	64×64*	82.0%	–
Assarasee [3]	2017	Microsoft API	Still images	5	5	–	80.0%	1
Fu et al. [16]	2017	Deep learning	Still images	5	7	80×60*	–	–
He et al. [22]	2017	Deep learning	Still images	1	64	–	100.0%	–
Kawagucgi [31]	2017	–	Still images	2	15	–	80.0%	1
Lim et al. [38]	2017	Fisher-faces	Video	9	9	45×34*	81.9%	1
Rekha et al. [46]	2017	Eigen-faces	Still images	1	15	64×48*	–	–
Surekha et al. [52]	2017	MKD-SRC	Video	20	20	–	60.0%	1
Polamarasetty et al. [44]	2018	HOG	Still images	3	14	112×92	–	–
Sarkar et al. [49]	2018	Deep learning	Still images	14	14	120×117	100.0%	1
Ours	2018	Deep learning	Still images	45	67	50×40	95.0%	25

+ Accuracies are not comparable, because experiments are different. * Estimated size from test images presented in the published paper. '–' means 'not given'.

In the last three years, novel methods based on deep learning for low-quality face images have been developed: In [57], Partially Coupled Networks are proposed for unsupervised super-resolution pre-training. The classification is by fine-tuning on a different dataset for specific domain super-resolution and recognition simultaneously. In [27][28], an attention model that shifts the network's attention during training by blurring the images with various percentage of blurriness is presented for gender recognition. In [41], three obfuscation techniques are proposed to restore face images that have been degraded by *mosaicing* (or pixalation) and blurring processes. In [6], a multitask deep model is proposed to simultaneously learn face super-resolution and facial landmark localization. The face super-resolution subnet is trained using a Generative Adversarial Network (GAN) [11]. In [26], inspired by the traditional wavelet that can depict the contextual and textural information of an image at different levels, a deep architecture is proposed. In [8], a network that contains a coarse super-resolution network to recover a coarse high-resolution image is presented. It is the first deep face super-resolution network utilizing facial geometry prior to end-to-end training and testing. In [10], a deblurring network for deblurring facial images is proposed using a Resnet-based non-maxpooling architecture. In [64], a face hallucination method based on an upsampling network and a discriminative network is proposed. In [51], global semantic priors of the faces are exploited in order to restore blurred face images. In all these methods, we see that computer face recognition is far from perfect when tackling more challenging such as faces taken in unconstrained environments, surveillance, forensics, etc.

In the literature review presented in this Section, we concluded that finding techniques to improve face recognition in low-quality images is an important contemporary research topic. A very challenging application is to manage a student attendance record in a crowded classroom using a smartphone camera as illustrated in Fig. 1.

2.2. Student Attendance System

In the literature, there are some works that reported student attendance systems. A summary of them (since 2012) is presented in Table 1. The most relevant are discussed in the following. In [46], a face recognition approach based on *eigen-faces* is presented for a classroom of 15 students. In this approach, the testing images are individual face images instead of images of the whole classroom, that means, no face detection is required. In [38], not only the students are recognized, but also their behavior (*e.g.* activities like entering or leaving the classroom). The experiments are conducted in a 9-student classroom. In [31], an observation camera with fisheye lens is used on the ceiling of the classroom to detect where the students are siting, and another camera is directed to the selected seat using the pan/tilt/zoom. In [49], a method based on deep learning is presented. The size of trained model is small enough to be installed in simple microprocessors. The method achieves excellent results in small classrooms using a high-resolution reflex camera. In [9], a method based on LBP and SVM is presented for recognition of faces at the entrance of the classroom. In the query images, few students are present and the resolution of the face images is 100×100. In [52], a comparison is given between controlled and uncontrolled environments. Unsurprisingly, the uncontrolled environments are prone to error. In [22], experiments were conducted with different illuminations. They concluded that the more the quality of the images the more the accuracy of the recognition.

In all these methods, the evaluation protocol used to estimate the accuracy is not clear enough to be reproduced.

That means, number of sessions, number of enrolled face images per subject, days between enrollment and testing, etc. are not clearly reported. It seems, that many of these recognition experiments have been conducted on images taken in only one session. Moreover, the subjects that are present in the query images are no more than 20, in many cases less than 10. In addition, the datasets and the implemented codes are not public, *i.e*. comparisons with the proposed methods is not possible. It is worthwhile to note that an attendance system must be used in a long period (sometimes a semester or a year), so the enrolled face images and the query face images can differ significantly in the time as shown in our dataset presented in Fig. 1. Thus, a dataset and a protocol that include more realistic scenarios with more sessions are required to design a robust attendance system.

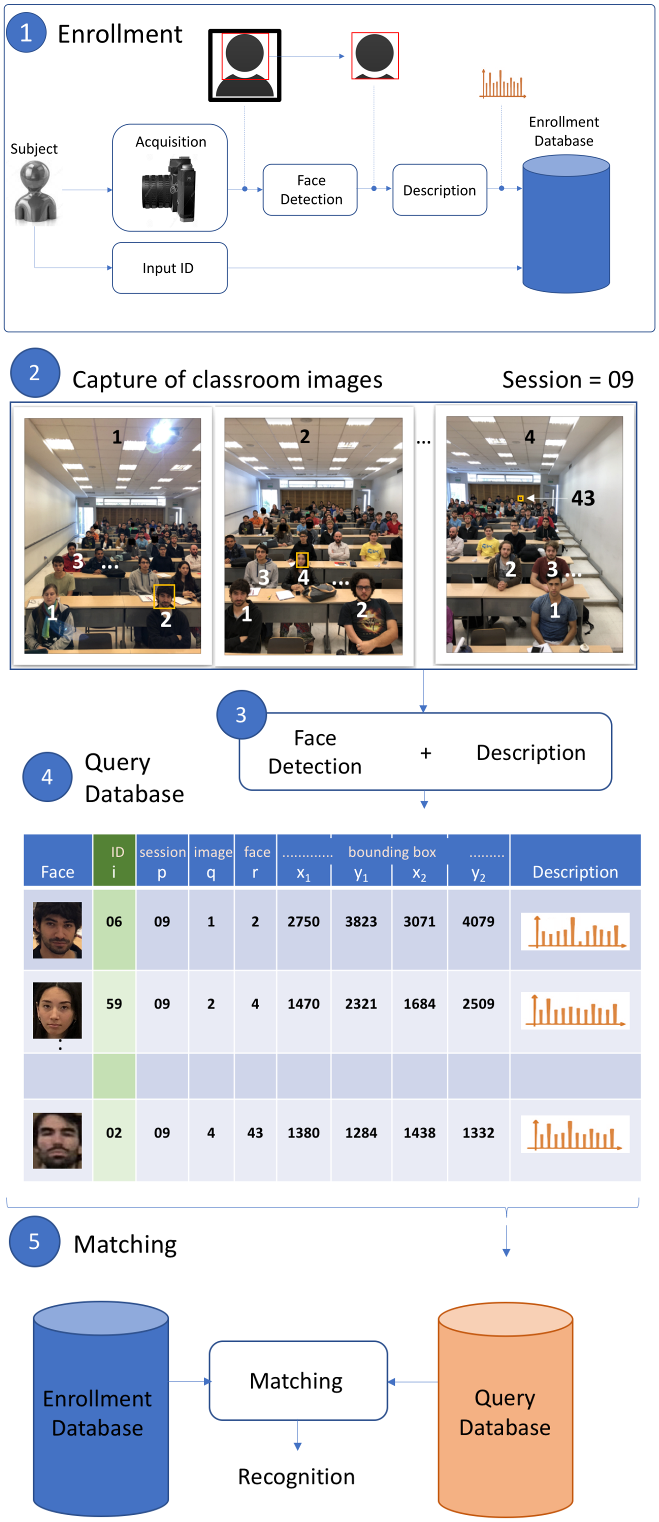
3. Proposed Method

In this work, we explain the proposed algorithm for Student Attendance System. It is presented in Fig. 2. It consists of five steps: ① enrollment,② capture of classroom images, ③ face detection and description, ④ query database and ⑤ matching algorithm. They will be addressed in the following five sub-sections in further details (one sub-section and one panel in Fig. 2 per step):

3.1. Enrollment

The enrollment of the participants is the first step in a face recognition system of a Student Attendance System. In this step, the biometric information of every subject of the classroom is captured and stored. As illustrated in Fig. 2, panel ①, we build the *Enrollment Database* by storing the face image, a description of the image (descriptor) and the ID of the subject. The last one is built using the identity information, such as name or ID number, given by the enrolled subject. The descriptor is a discriminative vector of d elements, *e.g*. for VGG-face model [43], $d = 4.096$. The information of enrollment is necessary for the recognition. In the recognition stage, face images that belong to the same/different subject might have similar/different vectors. Thus, Euclidean distance or cosine similarity can be used to match the image faces [37].

In our work, we asked for a selfie of each student that was sent per e-mail to the teacher[2]. Usually, only one face image is required by a recognition system, however, in our experiments, due to the low image quality of the face images of the students that are sitting in the last rows of the classrooms (see Fig. 1), we improved the accuracy of the system by adding more face images to the gallery. The new face images can be those images that are detected in the classroom images and added manually to the database.



Face	ID i	session p	image q	face r	.... bounding box x_1	 y_1	 x_2	.... y_2	Description
	06	09	1	2	2750	3823	3071	4079	
:	59	09	2	4	1470	2321	1684	2509	
	02	09	4	43	1380	1284	1438	1332	

Figure 2: Proposed method (see Section 3.1–3.5): ① Enrollment. ② Capture Session images. ③ Face detection and Description. ④ Query database. ⑤ Matching algorithm.

[2] In our classroom, there were 74 attendees, 67 of them gave their permission to be part of this research.

Table 2: Details of the dataset

Session (p)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Images (m_p)	6	6	5	4	8	6	5	11	10	7	5	8	4	8	6	6	4	6	4	4	4	7	6	5	8
Day (t_p)	0	5	7	12	14	19	21	28	33	35	47	49	56	61	63	68	75	77	84	89	91	96	98	103	105
Subjects (s_p)	60	60	61	62	56	58	52	54	56	47	55	49	45	47	48	56	48	52	46	46	49	50	39	39	50

In the enrollment database there are n subjects. In our experiments, $n = 67$. Thus, the ID of the enrolled students will be a number i for $i = 0, \cdots, n-1$. The enrolled color face images, detected by a face detection algorithm, are defined as $\mathbf{X}_{ij}$, for $j = 1 \cdots n_i$, where n_i is the number of enrolled face images of subject i. The corresponding descriptions are defined as:

$$\mathbf{x}_{ij} = f(\mathbf{X}_{ij}) \qquad (1)$$

where $f(\cdot)$ is the function that resizes the image if necessary and extracts the descriptor (a column vector of d elements) from the face image. In order to use the cosine similarity, function f returns a descriptor that has norm one. Thus, the similarity is computed by a simple dot product that corresponds to a normalized scalar product (cosine of angle). In our case, all enrolled face images $\{\mathbf{X}_{ij}\}$ have the same size: 165×120 pixels. The size of the original face image was changed using bicubic interpolation [17].

3.2. Capture of Classroom Images

In each session, we take several images of the classroom in order to capture all students that are present in the classroom. The capture is a collaborative process because the students that are present want to be recognized by the attendance system. The attendees were asked to look at the camera and make a neutral expression. If the classroom is small enough, only one image is required, however, in our classroom, the angle of view of the camera did not cover the whole classroom, for this reason we had to take several images as shown in Fig. 2, panel ②. For this end, we used an smartphone (iPhone-8, iOS 11.2.6) with a resolution of 4.032×3.024 pixels without flash. The color images were stored in HEIC format and converted to PNG using iMazing HEIC Converter 1.06, with 97% quality. Each converted image is stored in a file of 10MB approximately.

We define each captured session image as $\mathbf{S}_{pq}$, for $p = 1 \cdots M$ and $q = 1 \cdots m_p$, where M is the number of sessions (in our case, $M = 25$ sessions) and m_p is the number of images captured in session p. The number of images of each session is given in Table 2. Totally, we captured 153 images. In average, there were 6.1 images per session. All session images were captured in a period of 105 days (15 weeks). The days between consecutive sessions, $t_{p+1} - t_p$, were in average 4.4 (that correspond to two sessions per week with some exceptions). The number of attendees that are present in each session is given as s_p in Table 2, in average, there are 51 subjects per session.

3.3. Face Detection and Description

In each image session, faces are detected automatically with a face detection algorithm (*e.g.* Dlib[3]) and then manually checked as illustrated in Fig. 2 , panel ③. The number of detected faces in session image $\mathbf{S}_{pq}$, *i.e.* image q of session p, is defined as n_{pq}. The detected faces in image $\mathbf{S}_{pq}$ are stored as $\mathbf{Y}_{pqr}$, for $r = 1 \cdots n_{pq}$. For the description, we use the same function defined in (1):

$$\mathbf{y}_{pqr} = f(\mathbf{Y}_{pqr}). \qquad (2)$$

After this step, for each detected face defined, we have the location of its bounding box and its descriptor.

3.4. Query Database

In this step, we build the *Query Database* (see Fig. 2, panel ④) by storing all detected faces and their corresponding information. In the database, we store the following information for each detected face:

1. The detected image face.
2. The ground truth ID (i), that is the ID of the face.
3. The number of the session (p).
4. The number of the image of the session (q).
5. The number of the detected face image in the image of the session (r).
6. The location of the detected bounding box (coordinates (x_1, y_1) and (x_2, y_2) – the left-top and right-bottom pixel of the bounding box).
7. The descriptor ($\mathbf{y}$), a d-element vector computed by (2).

The ground truth ID of the Query Database is used for two purposes: *i)* For evaluation purposes, that is to determine the correctness of a matching; and *ii)* For testing purposes in which certain detected faces can be included in the gallery. For example, we can add to the Enrollment Database all face images of the first p sessions of the Query Database, and test the accuracy in the recognition with the rest, *i.e.* sessions $p + 1 \cdots M$. The ground truth can be established manually or by a semi-supervised algorithm that checks for special consistency.

[3] See http://dlib.net.

3.5. Matching

As shown in Fig. 2, panel ⑤, a matching algorithm is used to establish if a subject is present in a session. A simple algorithm –based on cosine similarity– can be used to determine if subject i is present in session p as follows: we find if the maximal value of the dot product of all combinations $\langle \mathbf{x}_{ij}, \mathbf{y}_{pqr} \rangle$ is greater than a threshold θ. We recall the reader that both vectors has norm one. This can be easily implemented using two matrices: $\mathbb{X}_i$ and $\mathbb{Y}_p$ defined by:

$$\mathbb{X}_i = [\mathbf{x}_{i,1} \cdots \mathbf{x}_{i,j} \cdots \mathbf{x}_{i,n_i}] , \tag{3}$$

where $\mathbf{x}_{i,j}$ is a $d \times 1$ vector defined in (1), and

$$\mathbb{Y}_p = \left[\{\mathbf{y}_{p,1}\} \cdots \{\mathbf{y}_{p,q}\} \cdots \{\mathbf{y}_{p,m_p}\} \right] , \tag{4}$$

where $\{\mathbf{y}_{p,q}\}$ is a matrix of $d \times n_{pq}$ elements defined as:

$$\{\mathbf{y}_{p,q}\} = \left[\mathbf{y}_{p,q,1} \cdots \mathbf{y}_{p,q,r} \cdots \mathbf{y}_{p,q,n_{pq}} \right] . \tag{5}$$

where $\mathbf{y}_{p,q,r}$ is a $d \times 1$ vector defined in (2). In this case, $\mathbb{X}_i$ is a matrix of size $d \times n_i$, in which the n_i descriptors of enrolled face images of subject i are stored in columns of d elements; and matrix $\mathbb{Y}_p$ is of size $d \times n_p$, in which the descriptors of n_p detected faces in session p are stored in columns of d elements, where n_p is defined as $n_p = n_{p1} + n_{p2} + \cdots n_{pm_p}$. All combinations of the dot product can be computed by

$$\mathbb{Z}_{ip} = \mathbb{X}_i^{\mathsf{T}} \mathbb{Y}_p, \tag{6}$$

where the result is a matrix of $n_i \times n_p$ elements. Thus, if

$$\max(\mathbb{Z}_{ip}) > \theta, \tag{7}$$

we could say that subject i is present in session p, where parameter θ is a threshold given in the algorithm. In our experiments, $\theta = 0.5$ achieves good results.

3.6. Attendance Algorithm

The attendance algorithm is based on the method described in Section 3.5. The key-idea of the approach is not to recognize the identity of every detected face in the images of a session, but to recognize if an specific subject is present in the session, that is to find the most similar detected face in the session given an ID. Thus, given the descriptors of the enrolled face images of subject ID, we have to look for the most similar descriptor of the detected faces in the session. The algorithm that computes the attendance percentage of subject ID is presented in Algorithm 1. For example, the attendance percentage of subject #02 in sessions 8, 9, $\cdots$, 16, can be estimated by setting in the input of Algorithm 1: ID = 02 and $\mathbf{p} = [8\ 9 \cdots 16]$ (see for example the result for the same subject for $\mathbf{p} = [1 \cdots 25]$ in Fig. 1). If we want to compute the attendance sheet in all sessions for the whole class, we have to set $\mathbf{p} = [1 \cdots M]$, and repeat Algorithm 1 for every enrolled student, that is for $i = 1 \cdots n$.

```
input  : ID and sessions $\mathbf{p}$
output: Attendance percentage of ID
begin
    $i \leftarrow$ ID
    Compute $\mathbb{X}_i$ using (3)
    $s \leftarrow 0$ % number of sessions
    $a \leftarrow 0$ % number of attended sessions
    for $p$ in $\mathbf{p}$ do
        $s \leftarrow s + 1$
        Compute $\mathbb{Y}_p$ using (4)
        $\mathbb{Z}_{ip} \leftarrow \mathbb{X}_i^{\mathsf{T}} \mathbb{Y}_p$
        if $\max(\mathbb{Z}_{ip}) > \theta$ then
            $a \leftarrow a + 1$
        end
    end
    AttendancePercentage $\leftarrow a/s \times 100$
end
```

Algorithm 1: Algorithm that computes the attendance percentage of subject ID in the sessions given in array $\mathbf{p}$.

4. Experimental Results

In order to present the experiments used in our work, in this Section, we give further details of the dataset, the experimental protocol, the obtained results with analysis and the implementation.

4.1. Dataset

The dataset consists of 3 sets of images, the database of detected faces, the set of descriptors for the enrolled and query faces and the attendance sheet:

- Enrollment Images Database: The Enrollment Database contains the face images of 67 subjects and their IDs.
- Session Images: For each session there are several color images (4.032×3.024 pixels) of the classroom. The number of images per session are given in Table 2. Totally, there are 153 session images in 25 sessions.
- Query Images Database: For each detected face, a crop is saved for the purpose of descriptor extraction. There is a total of 4.898 detected face images.
- Query Database: The Query Database has one entry per detected face in the session images. The descriptor vectors are stored separately. See more details in Section 3.4.
- Descriptor Database: for each face the Enrollment and Query Databases there is a descriptor vector. To test each one of the ten different descriptors this database is loaded accordingly. For further details see Section 4.3.
- Attendance sheet: For this end, we define a binary array $\mathbf{A}$ of size $n \times M$, where $n = 67$ is the number of enrolled subjects and $M = 25$ is the number of sessions. Thus, $A(i,p)$ is 1 or 0 if subject i was present or absent in session p.

```
input  : Descriptors of all sessions
output: Accuracy average $\bar{\eta}$
begin
    $\mathbb{X}_i \leftarrow$ Enrollment for $i = 1 \cdots n$ using (3)
    $\mathbb{Y}_p \leftarrow$ Query for $p = 1 \cdots M$ using (4)
    for $s = 0 \cdots M - 1$ do
        % s: number of sessions in the gallery
        $c \leftarrow 0$ % number of samples to be detected
        $t \leftarrow 0$ % number of samples correctly detected
        for $p = s + 1 \cdots M$ do
            for $i = 1 \cdots n$ do
                $\mathbb{Z}_{ip} \leftarrow \mathbb{X}_i^{\mathsf{T}} \mathbb{Y}_p$
                if $\max(\mathbb{Z}_{ip}) > \theta$ then
                    $a \leftarrow 1$
                else
                    $a \leftarrow 0$
                end
                $c \leftarrow c + 1$
                if $A(i,p) = a$ then
                    $t \leftarrow t + 1$
                end
            end
        end
        $\bar{\eta}(s) = t/c$
        for $i = 1 \cdots n$ do
            $\mathbb{X}'_i \leftarrow$ descriptors in $\mathbb{Y}_{s+1}$ of subject $i$
            $\mathbb{X}_i \leftarrow [\mathbb{X}_i \; \mathbb{X}'_i]$
        end
    end
end
```

Algorithm 2: Evaluation protocol.

4.2. Experimental Protocol

There are $M = 25$ full annotated sessions, *i.e.* we have the detected faces with the corresponding ID's and the real attendance sheet $\mathbf{A}$ of every session as explained in Section 4.1. Following the idea of Algorithm 1 we define the evaluation protocol as presented in Algorithm 2.

In order to understand Algorithm 2, we start with the explanation of the accuracy of a session: In each session, there are positive and negative 'samples', that means subjects that are present and subjects that are absent. Thus, the accuracy in a session will be defined as:

$$\eta = \frac{TP + TN}{n} \tag{8}$$

where TP (true positives) is the number of present subjects correctly detected, TN (true negatives) is the number of absent subjects correctly detected, and n is the number of 'samples', that means number of enrolled subjects (in our case $n = 67$). Ideally, $TP + TN = n$, *i.e.* $\eta = 1$.

In our protocol of Algorithm 2, we define the average accuracy $\bar{\eta}(s)$, that is the average of the accuracies η according to (8) computed for sessions $s + 1 \cdots M$ when in the gallery we have the original enrolled face with the faces of the first s sessions. For example, $\bar{\eta}(0)$ means the average accuracy of all sessions $(1 \cdots M)$ when in the gallery we have only the original face image used in the enrollment. In addition, $\bar{\eta}(3)$ means the average accuracy of sessions $4 \cdots M$, when in the gallery we have the original face image used in the enrollment and the face images of the first three sessions. The idea is to establish, how many face images are necessary in the gallery to achieve robust results.

4.3. Experiments

We execute Algorithm 2 for ten handcrafted and learned descriptors as follows:

- **Handcrafted descriptors**: *i)* LBP: descriptor based on Local Binary Patterns [1]. As pre-processing, the images were resized to 224×224 pixels and converted to grayscale. Afterwards, the grayscale images were divided into 4×4 partitions. Thus, we extracted 16 LBP features of 59 elements each yielding a descriptor of 944 elements. *ii)* HOG: descriptor based on Histogram of Gradients [12]. We followed the pre-processing mentioned in LBP. Afterwards, HOG features were extracted from grayscale images using a cell size of 18×18 pixels. Thus, the HOG-descriptor has 4.356 elements. *iii)* Gabor: descriptor based on Gabor Transform [21]. We followed the pre-processing mentioned in LBP. Afterwards, Gabor features were extracted from grayscale images using a factor of 16 for downsampling along rows and columns, with 5 scales and 8 orientations. Thus, the Gabor-descriptor has 7.840 elements.
- **Learned descriptors**: *i)* VGG-Face: a deep learning model with a descriptor of 4.096 elements [43]. *ii)* Dlib: a deep learning model based on the ResNet architecture with a descriptor of 128 elements [32]. *iii)* FaceNet: a deep learning model with a descriptor of 128 elements [50]. *iv)* OpenFace: a deep learning model with a descriptor of 128 elements [2]. *v)* SqueezeNet: a deep learning model with a low number of layers with a 2.048-element descriptor [19]. *vi)* GoogleNet-F: a known model (GoogleNet) trained for faces with a 2.048-element descriptor [19]. *vii)* AlexNet-F: a known model (AlexNet) trained for faces with a 4.096-element descriptor[19].

For each descriptor, we choose the best parameter θ by maximizing the accuracy using exhaustive search. The idea is to report the maximal accuracy achieved by each method. In our experiments, we decided to focus on face recognition and not in face detection, because manual and automated face detection (using for example Dlib [32]) achieved very similar results in our session images. The most likely reason of this is the collaborative nature of the capture process in which all faces to be recognized were frontal with neutral face expressions.

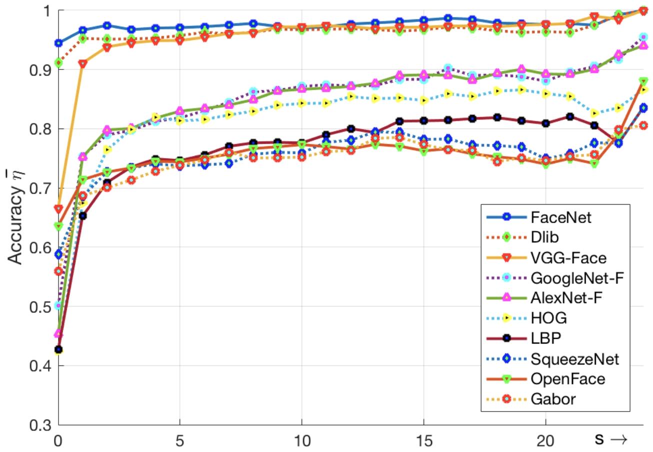


Figure 3: Average accuracy according to Algorithm 2.

Table 3: Average accuracy for $s = 0, 3, 5, 7$

Descriptor	θ	$\bar{\eta}(0)$	$\bar{\eta}(3)$	$\bar{\eta}(5)$	$\bar{\eta}(7)$
FaceNet	0.55	0.9445	0.9674	0.9709	0.9751
Dlib	0.50	0.9116	0.9518	0.9575	0.9602
VGG-Face	0.50	0.6657	0.9450	0.9493	0.9610
GoogleNet-F	0.35	0.5009	0.7972	0.8187	0.8449
AlexNet-F	0.50	0.4537	0.8005	0.8299	0.8400
HOG	0.50	0.4251	0.7992	0.8134	0.8234
LBP	0.51	0.4269	0.7354	0.7463	0.7703
SqueezeNet	0.50	0.5875	0.7341	0.7373	0.7413
OpenFace	0.65	0.6358	0.7341	0.7440	0.7570
Gabor	0.24	0.5594	0.7137	0.7381	0.7604

4.4. Results and Implementation

Results are summarized in Fig. 3. On the one hand, we observe that the best deep learning method in this experiments was FaceNet achieving 97% using as gallery the enrollment face image and the (labeled) face images of the first three sessions (see Fig. 1 for a perfect attendee record of ID #02 in all 25 sessions using our method). On the other hand, all handcrafted methods achieve low accuracies, however, the best one was HOG (more than 80% after the first three sessions). In order to compare numerically all methods, a summary is presented in Table 3. Here, the average accuracy after using in the gallery the enrollment faces and the face images of the first 0, 3, 5 and 7 sessions is shown.

We implemented our methods in Matlab (VGG-face and handcrafted features) and in Python (the rest of features and Algorithm 2)[4].

5. Conclusions

In this paper, we propose an automated student attendance system based on deep learning that can be used in crowded classrooms, where the session images are taken by a smartphone camera. To the best knowledge of the authors, this is the first work that presents a realistic solution in a crowded classroom (around 70 attendees) in so many sessions (25 sessions with images taken during 15 weeks). Ten well known face recognition algorithms based on learned and handcrafted features were evaluated using a protocol that takes into account the number of face images per subject used in the gallery. In our experiments, the best one has been FaceNet, a method based on deep learning features, achieving around 95% of accuracy using only one enrollment face image per subject. Both full annotated databases and codes are available on our webpage. We believe that our automated student attendance system based on face recognition, can be used to save time for both teacher and students, and to prevent fake attendance. As future work, we will implement more sophisticate algorithms based on reduction of dimensionality checking spatial consistency of the attendees.

Acknowledgments

This work was supported by Fondecyt Grant No. 1161314 from CONICYT, Chile.

References

[1] T. Ahonen, A. Hadid, and M. Pietikinen. Face description with local binary patterns: application to face recognition. *IEEE Trans Pattern Anal Mach Intell*, 28(12):2037–2041, dec 2006.

[2] B. Amos, B. Ludwiczuk, and M. Satyanarayanan. Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.

[3] P. Assarasee, W. Krathu, T. Triyason, V. Vanijja, and C. Arpnikanondt. Meerkat: A framework for developing presence monitoring software based on face recognition. In *2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media)*, pages 1–6, Aug 2017.

[4] S. Baker and T. Kanade. Hallucinating faces. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 83–88, 2000.

[5] S. Biswas, K. Bowyer, and P. Flynn. Multidimensional scaling for matching low-resolution face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10):2019–2030, 2012.

[6] A. Bulat and G. Tzimiropoulos. Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2018.

[4] The code is available on http://dmery.ing.puc.cl/index.php/material/ (available after publication).

[7] L. Chen, R. Hu, Z. Han, Q. Li, and Z. Lu. Face super resolution based on parent patch prior for VLQ scenarios. *Multimed Tools Appl*, 76(7):10231–10254, apr 2017.

[8] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang. FSRNet: End-to-end learning face super-resolution with facial priors. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2018.

[9] S. Chintalapati and M. V. Raghunadh. Automated attendance management system based on face recognition algorithms. In *2013 IEEE International Conference on Computational Intelligence and Computing Research*, pages 1–5. IEEE, dec 2013.

[10] G. G. Chrysos and S. Zafeiriou. Deep face deblurring. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017.

[11] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath. Generative adversarial networks: an overview. *IEEE Signal Process Mag*, 35(1):53–65, jan 2018.

[12] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[13] C. Ding and D. Tao. Trunk-branch ensemble convolutional neural networks for video-based face recognition. *IEEE Trans Pattern Anal Mach Intell*, 40(4):1002–1014, apr 2018.

[14] S. Dodge and L. Karam. Understanding how image quality affects deep neural networks. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, jun 2016.

[15] J. Flusser, S. Farokhi, C. Hoschl, T. Suk, B. Zitova, and M. Pedone. Recognition of images degraded by gaussian blur. *IEEE Trans Image Process*, dec 2015.

[16] R. Fu, D. Wang, D. Li, and Z. Luo. University classroom attendance based on deep learning. In *2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, pages 128–131. IEEE, oct 2017.

[17] R. Gonzalez and R. Woods. *Digital Image Processing*. Pearson, Prentice Hall, third edition, 2008.

[18] R. Gopalan, S. Taheri, P. Turaga, and R. Chellappa. A blur-robust descriptor with applications to face recognition. *IEEE Trans Pattern Anal Mach Intell*, 34(6):1220–1226, jun 2012.

[19] K. Grm, V. Štruc, A. Artiges, M. Caron, and H. K. Ekenel. Strengths and weaknesses of deep learning models for face recognition against image degradations. *IET Biometrics*, 7(1):81–89, 2017.

[20] K. Grm, V. Struc, A. Artiges, M. Caron, and H. K. Ekenel. Strengths and weaknesses of deep learning models for face recognition against image degradations. *IET Biometrics*, 7(1):81–89, jan 2018.

[21] M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb. Cloudid: Trustworthy cloud-based and cross-enterprise biometric identification. *Expert Systems with Applications*, 42(21):7905–7916, 2015.

[22] C. He, Y. Wang, and M. Zhu. A class participation enrollment system based on face recognition. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pages 254–258, June 2017.

[23] D. Heinsohn and D. Mery. Blur adaptive sparse representation of random patches for face recognition on blurred images. In *Workshop on Forensics Applications of Computer Vision and Pattern Recognition, in conjunction with International Conference on Computer Vision (ICCV2015), Santiago, Chile*, 2015.

[24] P. H. Hennings-Yeomans, S. Baker, and B. V. Kumar. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.

[25] C. Herrmann, D. Willersinn, and J. Beyerer. Residual vs. inception vs. classical networks for low-resolution face recognition. In P. Sharma and F. M. Bianchi, editors, *Image Analysis*, volume 10270 of *Lecture notes in computer science*, pages 377–388. Springer International Publishing, Cham, 2017.

[26] H. Huang, R. He, Z. Sun, and T. Tan. Wavelet-SRNet: A wavelet-based CNN for multi-scale face super resolution. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1698–1706. IEEE, oct 2017.

[27] F. Juefei-Xu, E. Verma, P. Goel, A. Cherodian, and M. Savvides. DeepGender: Occlusion and low resolution robust facial gender classification via progressively trained convolutional neural networks with attention. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 136–145. IEEE, jun 2016.

[28] F. Juefei-Xu, E. Verma, and M. Savvides. DeepGender2: A generative approach toward occlusion and low-resolution robust facial gender classification via progressively trained attention shift convolutional neural networks (PTAS-CNN) and deep convolutional generative adversarial networks (DCGAN). In *Deep Learning for Biometrics*, pages 183–218. Springer, 2017.

[29] N. Kar, M. K. Debbarma, A. Saha, and D. R. Pal. Study of implementing automated attendance system using face recognition technique. *IJCCE*, pages 100–103, 2012.

[30] S. Karahan, M. Kilinc Yildirum, K. Kirtac, F. S. Rende, G. Butun, and H. K. Ekenel. How image degradations affect deep CNN-based face recognition? In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE, sep 2016.

[31] Y. Kawaguchi, T. Shoji, L. Weijane, K. Kakusho, and M. Minoh. Face recognition-based lecture attendance system. In *The 3rd AEARU workshop on network education*, pages 70–75. Citeseer, 2005.

[32] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10(Jul):1755–1758, 2009.

[33] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa. Disguised faces in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, volume 8, 2018.

[34] Z. Lei, T. Ahonen, M. Pietikainen, and S. Z. Li. Local frequency descriptor for low-resolution face recognition. In *Face and Gesture 2011*, pages 161–166. IEEE, mar 2011.

[35] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Computer science and artificial intelligence laboratory technical report;

efficient marginal likelihood optimization in blind deconvolution. *Optimization*, 2011.

[36] P. Li, L. Prieto, D. Mery, and P. Flynn. Low resolution face recognition in the wild. *arXiv preprint arXiv:1805.11529*, 2018.

[37] S. Li and A. Jain. *Handbook of face recognition*. Springer, 2011. Second Edition.

[38] J. H. Lim, E. Y. Teh, M. H. Geh, and C. H. Lim. Automated classroom monitoring with connected visioning system. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 386–393. IEEE, dec 2017.

[39] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26, 2017.

[40] S. Lukas, A. R. Mitra, R. I. Desanti, and D. Krisnadi. Student attendance system in classroom using face recognition technique. In *2016 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1032–1035. IEEE, oct 2016.

[41] R. McPherson, R. Shokri, and V. Shmatikov. Defeating image obfuscation with deep learning. *arXiv preprint arXiv:1609.00408*, 2016.

[42] S. P. Mudunuri and S. Biswas. Low resolution face recognition across variations in pose and illumination. *IEEE Trans Pattern Anal Mach Intell*, 38(5):1034–1040, may 2016.

[43] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference (BMVC2015)*, volume 1, page 6, 2015.

[44] V. K. Polamarasetty, M. R. Reddem, D. Ravi, and M. S. Madala. Attendance system based on face recognition. *Work*, 5(04), 2018.

[45] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J.-C. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa. Deep learning for understanding faces: Machines may be just as good, or better, than humans. *IEEE Signal Process Mag*, 35(1):66–83, jan 2018.

[46] E. Rekha and P. Ramaprasad. An efficient automated attendance management system based on eigen face recognition. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, pages 605–608. IEEE, jan 2017.

[47] R. Ren, T. Hung, and K. C. Tan. A generic deep-learning-based approach for automated surface inspection. *IEEE Trans Cybern*, feb 2017.

[48] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, nov 1992.

[49] P. R. Sarkar, D. Mishra, and G. R. K. S. Subhramanyam. Automatic attendance system using deep learning framework. In M. Tanveer and R. B. Pachori, editors, *Machine intelligence and signal analysis*, volume 748 of *Advances in intelligent systems and computing*, pages 335–346. Springer Singapore, Singapore, 2019.

[50] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823. IEEE, jun 2015.

[51] Z. Shen, W. Lai, T. Xu, and J. Kautz. Deep semantic face deblurring. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. openaccess.thecvf.com, 2018.

[52] B. Surekha, K. J. Nazare, S. Viswanadha Raju, and N. Dey. Attendance recording system using partial face recognition algorithm. In N. Dey and V. Santhi, editors, *Intelligent techniques in signal processing for multimedia security*, volume 660 of *Studies in computational intelligence*, pages 293–319. Springer International Publishing, Cham, 2017.

[53] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. DeepFace: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708. IEEE, jun 2014.

[54] A. Tikhonov and V. Arsenin. *Solutions of ill-posed problems*. Vh Winston, 1977.

[55] V. L. Uskov, J. P. Bakken, and A. Pandey. The ontology of next generation smart classrooms. In V. L. Uskov, R. J. Howlett, and L. C. Jain, editors, *Smart Education and Smart e-Learning*, pages 3–14, Cham, 2015. Springer International Publishing.

[56] P. Wagh, R. Thakare, J. Chaudhari, and S. Patil. Attendance system based on face recognition using eigen face and PCA algorithms. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, pages 303–308. IEEE, oct 2015.

[57] Z. Wang, S. Chang, Y. Yang, D. Liu, and T. S. Huang. Studying very low resolution recognition using deep networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4792–4800. IEEE, jun 2016.

[58] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang. Deep networks for image super-resolution with sparse prior. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 370–378. IEEE, dec 2015.

[59] Z. Wang, Z. Miao, Q. M. JonathanWu, Y. Wan, and Z. Tang. Low-resolution face recognition: a review. *Vis Comput*, 30(4):359–386, apr 2014.

[60] Y. Xiao, Z. Cao, L. Wang, and T. Li. Local phase quantization plus: A principled method for embedding local phase quantization into fisher vector for blurred image recognition. *Information Sciences*, 2017.

[61] Y. Xu, Z. Li, J. Yang, and D. Zhang. A survey of dictionary learning algorithms for face recognition. *IEEE Access*, 5:8502–8514, 2017.

[62] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Trans Image Process*, 19(11):2861–2873, nov 2010.

[63] X. Yin and X. Liu. Multi-task convolutional neural network for pose-invariant face recognition. *IEEE Trans on Image Processing*, 27(2):964–975, feb 2018.

[64] X. Yu, B. Fernando, R. Hartley, and F. Porikli. Super-resolving very low-resolution face images with supplementary attributes. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 908–917. IEEE, 2018.

[65] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Progressive inter-scale and intra-scale non-blind image deconvolution. In *ACM Transactions on Graphics (TOG)*, page 74, 2008.