

A Review of Traditional Methods and Deep Learning for Face Recognition

Hongyun Zeng

School of Computer and Network Security, Chengdu University of Technology,
Sichuan, China

zeng.hongyun@student.zy.cdut.edu.cn

Abstract. Face recognition technology is a key research direction in artificial intelligence and computer vision. It first originated in the 1960s, when Bledsoe et al. proposed and studied computer programming based on face recognition. It is divided into three steps: face detection, feature extraction, and face verification. This paper briefly introduces the wide application of face recognition in various fields such as vulnerable groups, analyses its characteristics and progress from the viewpoints of traditional learning and deep learning, and focuses on the methods based on face feature extraction (including geometric features, texture features, local features, etc.). From the previous studies appear some problems connected to environmental changes, such as light change, facial occlusion, posture and privacy leakage, etc. This paper will give the possible reasons for these problems and give the possible development direction of the future by exploring the current face recognition technology to provide a reference for the related research of face recognition technology.

Keywords: Face recognition, Deep learning, Traditional learning.

1. Introduction

Nowadays, everyone's identity is important in social life. Traditional forms of identification, such as ID cards, driving licenses, etc., can easily be lost, stolen, or even forged, leading to identity theft or unauthorized access to personal information [1]. Therefore, biometric authentication is essential to confirm status. Facial recognition is considered to be the most natural and effective compared to earlier forms of authentication such as fingerprints and iris scans, as people often use their faces to recognize people rather than other features [2]. Facial recognition technology first originated in the 1960s, when Bledsoe et al. proposed and investigated face recognition-based computer programming [3]. Unlike classical computer programming, face recognition involves training a machine and allowing the computer to learn from the database and make empirical predictions about the result [4]. This is used for solving complex tasks and reducing the need for human intervention [4]. Along with the development of computer vision and artificial intelligence, face recognition technology has gone through multiple stages of development. From the early ensemble feature matching to the artificial feature extraction method based on statistics, which improves the accuracy of the face recognition system. In recent years, the deep learning of memory neural networks has further solved the problems of facial images due to light changes or facial occlusion to a certain extent, which significantly improves the accuracy of face recognition and has become an indispensable technology in today's society.

At present, face recognition technology is mainly divided into two categories: traditional learning methods and deep learning methods. Traditional learning methods are mainly based on geometric measurements and feature extraction principles, combined with statistical methods such as Principal component analysis [5], (Linear Discriminant Analysis) LDA, etc. for face recognition [6]. Although these methods perform well in simple environments, once the angle is changed, or the face is rotated, or even the lighting changes, the methods do not perform well. Some of these techniques, such as LBP, SIFT, SURF, SVM, KNN, etc., can adapt to more complex environments and perform better to changes in light [3]. However, there are still limitations in unconstrained environments, e.g., pose changes, camouflage, and facial occlusion [3]. To overcome these problems, the convolutional neural network (CNN) technique was first proposed in 1997[7]. Since 2001, deep learning-based systems have made breakthroughs in performance and accuracy, and have demonstrated excellent performance in a variety of complex scenes [7].

With the advancement of society and the need to make life more convenient for people traveling, facial recognition technology is increasingly being used in daily life, such as improving the retail experience, healthcare, recording student attendance, unlocking mobile phones, automatic welding fault detection, lip reading (to help the hearing impaired with CNN to analyze lip movements), recognizing human emotions, driver fatigue detection, blind modeling (to help a blind person alert a match by voice and tell him if he knows the person), reducing crime, apprehending criminals, helping with border control, etc[4, 8].

The purpose of this paper is to review the development of face recognition technology to better demonstrate the characteristics of different stages of face recognition technologies and their differences. This paper will be from perspectives of the stage of technological development: according to the traditional method, deep learning two phases of face recognition technology, and explore the representative methods of each stage, focusing on the method based on face feature extraction and its advantages and shortages. Through this comprehensive analysis, the paper explores the current problems and propose possible future research directions to help readers understand the evolution of face recognition technology and help future research development and speed up technology improvement.

2. Methods

Usually, face technology is processed by three steps: 1) face detection 2) feature extraction 3) face recognition to verify the identity of the face see Figure. 1. In these three steps, there are different methods to achieve the final aim of face detection. This section discusses the main methods of face technology and how these methods can be used to further solve the face recognition problem efficiently by using different feature extraction and classification techniques. Divided the methods into two classifications based on the need of the data size and complexity of the processing task, as well as the difference in the characteristics of dependence. These are described in detail as follows

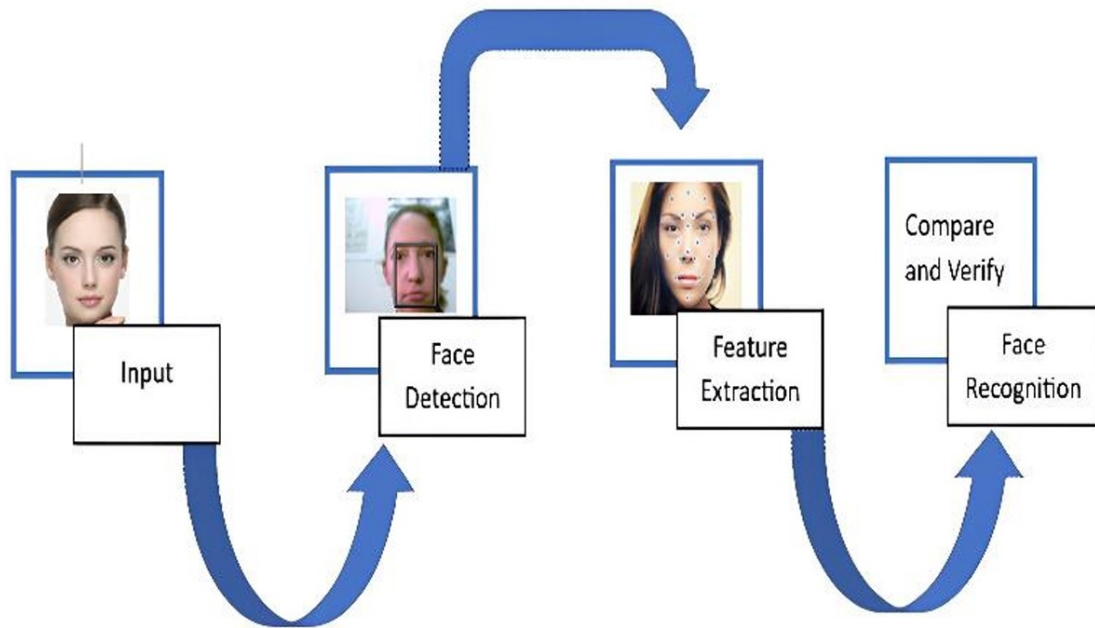


Figure 1. Steps of Face Recognition [1].

2.1. Traditional methods

Due to the limitations of computer power and hardware, inchoate face recognition techniques needed to be performed simply and feasibly by using existing image processing techniques and mathematical methods. In 1991, Turk and Pentland introduced PCA into facial recognition for the first time [1]. PCA is used by extracting the main features of facial images (also called feature faces) and projecting them into a low-dimensional space, preserving the fundamental features of the data [1]. In this way, it can largely speed up the process of data, also save time and cost. However, because of the different facial expressions of faces, light changes, or face occlusions, these may make PCA difficult to obtain low-dimensional data, and the requirements of the data must be large enough when training computers [9]. Besides, this method is not suitable for nonlinear data [1]. To solve the problem of light changes, researchers created one method called Local Binary Patterns (LBP), which is robust to light changes [9]. LBP is a technique based on texture feature extraction, proposed by Ahonen et al. in 2004-2006, in which faces are divided into different blocks, and the elements of each local image block are modeled to form a class of feature vectors for classification [3]. This established the foundation for a new LBP computation called one-dimensional local binary pattern(1DLBP), proposed by Benzaoui and Boukrouche in 2013-2014, to further improve the accuracy of CRR databases [3]. Similarly, Gabor filters extract features by choosing their frequency, orientation, and size and do not need to detect key points, but rather use the filters directly at each pixel location, addressing the need for local feature detection, such as LBP [3, 5, 9]. However, traditional learning is easy to train on small scales and difficult to train on massive data. To overcome the limitations of traditional learning in feature extraction, it is necessary to further improve the feature extraction technique of traditional learning. Scale-invariant feature transform (SIFT) is an algorithm for detecting and describing local features, which is invariant to scale and rotation[5]. However, the drawbacks are also obvious: high computational complexity and time required. Inspired by this technique, Speeded-up robust features (SURF) solve the problem of long execution time by using integral images to seek to find the maximum of the Hessian determinant approximation [5]. Unlike traditional learning of PCA, linear discriminate analysis (LDA) requires that the within-group variance of the low-dimensional data be as small as possible and the between-group variance be as large as possible [9]. The main objective is to classify the face images using the most accurate features to describe the face images and maximize the discriminate categories while reducing

the dimensionality [1]. Similar to PCA, independent component analysis (ICA) is considered an extension of PCA, which is mainly applied to solve the challenges in signal processing, which is highly favorable to non-Gaussian distributions, and then maximize the non-Gaussian distributions and isolate the independent components to process feature extraction and classification [1]. Subsequently, kernel discriminant analysis (KDA) further enhances the scope of application of LDA and is suitable for dealing with non-linear problems to solve the mentioned problems of PCA.

2.2. Deep learning

To solve more complex classification tasks and to apply computer learning to big data models, a learning technique based on neural network architecture, deep learning, has been created[4]. Its typical representative called convolutional neural network(CNN) has been widely used in various fields in recent years, such as medical treatment, surveillance, and so on. It is inspired by the visual recognition of objects in animals and humans [4]. Unlike traditional neural networks, which have only one layer, CNN is made of multiple layers: a convolutional layer, a pooling layer, and a fully connected layer [5]. Sometimes there are activation layers, batch normalization layers, and dropout layers, where convolution is used to extract features from images using CNN [4]. There are also many different architectures for CNNs, some classic representatives, such as Yann L.C. et al. in 1998 created LeNet, AlexNet proposed by Alex Krizhevsky et al. in 2012, Google team proposed GoogLeNet in 2014, Kaiming He et al. in 2015 proposed ResNet, Gao Huang et al. in 2017 proposed DenseNet, and also existing some of their networks showed great advantages in ImageNet competition [4, 10], CNN can automatically learn the hierarchical features of input images and build high-level features over low-level features. Compared to some other methods, such as FCNN, RNN, and GNN. CNN can capture the spatial relationships between pixels in an image, the performance that none of them have. There are some other shortages: FCNN is easier to overfit than CNN, RNN is computationally more expensive than CNN and is impractical for some human tasks and GNN is more expensive than CNN when it comes to dealing with large graphs [4]. In addition, CNN can generalize well to new data, which means that they can make predictions about data that has never been seen before, and they offer significant improvements in accuracy and efficiency compared to traditional methods[4]. However, CNN also faces the costly and time-consuming task of training models on large number of GPUs.

3. Experiment

3.1. Database

Diverse databases can ensure the accuracy of face recognition technology in different complex tasks and real-world scenes, and thus improve the practicability of face recognition technology, such databases play an important role in training the machine and evaluating the performance of the model. Generally, the database is divided into test set and training set, in which the training set is used to train the model, while the test set is used to evaluate the generalization ability of the model and to detect whether the overfitting phenomenon occurs [5]. Currently, there are the following typical databases.

- Labelled faces in wild (LWF) is a commonly used database. This contains 13333 images, of which 1680 are composed of faces with at least two images of faces in different environments [5]. Most of them are in colour and are presented in JPG format [5].
- CASIA-WebFace is a database of half a million images about 10,000 topics, covering a rich variety of images [1].
- VGGFace was introduced by the Oxford Group to train learning systems [1]. Contains 2.6 million faces from about 2622 individuals, but suffers from some noise-related problems resulting in outliers in within-group variance [1].
- MS-Celeb-1M consists of approximately 10 million images, including 100,000 celebrities [1].
- Yale Face contains 16 images of 15 people, unlike LWF, which are all grey images but with different facial expressions or occlusions [1].

- Olivetti Research Laboratory (ORL) collected at the Cambridge Laboratory in the 1990s [5]. Ten images of each individual were taken at different times, with variations in expression and lighting conditions [5].
- FERET was a project sponsored by the United States Department of Defence when it was discovered that a small number of face recognition systems could be used to recognize faces in still images in controlled environments [8]. Containing 1564 sets of images, covering a wide range of images from different angles and facial expression variations, and are in color [5].

3.2. Assessment of indicators

In the above-mentioned steps of face detection, face detection occupies a key position in face recognition techniques. It is usually used to determine whether an image contains a face, and if it does, the data is then as input into a recognition model for identification. In this process, the accuracy, precision, and recall of the detection affect the performance of the whole system [5]. Therefore, evaluating the algorithms of the face recognition system plays a crucial role in the whole face recognition technology. Thus, the paper uses a variety of evaluation indexes to relatively comprehensively analyze the detection methods, compare and contrast the indexes to judge the algorithms, and then to improve the algorithms. The next section will introduce some typical different indicators.

- Receiver Operating Characteristic (ROC) obtains different curves by changing various thresholds [4]. In ROC, the similarity between images is calculated first, and then the recognition is done by local thresholding [9]. In this image (Table 1), the area under the ROC curve is called the area under curve(AUC), the larger its value, the better the performance of the algorithm [9]. Besides, the closer the ROC is to the upper left corner, the better the algorithm performs [9].
- Accuracy(Acc) is used to detect the percentage of correct classification and the formula is shown in Table 1 [4]. The higher the Acc value, the better the algorithm performance [9].
- Recall is the number of positive samples that can be detected correctly. A higher value means that fewer samples are not detected, this means the algorithm can find most of the face[4].
- The false match rate(FMR), also called false accept rate (FAR), is used to indicate the percentage of illegal aggressors that are incorrectly perceived as legitimate users by the system, so the lower the FMR, the higher the security performance of the algorithm [1].
- The false non-match rate(FNMR), also called false reject rate(FRR), is used to indicate the percentage that a correct sample will be incorrectly rejected by the system, and measures the inability of the system to effectively identify a legitimate user, so the lower the FNMR, the better it ensures that access rights can be correctly enabled [1].

Table 1. Metrics formula (In this table, TP is the number of correctly detected faces. FP is the number of incorrectly detected faces. FN is the number of missed faces. TN is the number of correctly recognized non-faces)

ACC	$\frac{TP + TN}{TP + TN + FN + FP}$
Recall	$\frac{TP}{TP + FN}$

3.3. Results

The algorithms and evaluation indexes for this part have been described in detail above. A comparison of the different algorithms, mainly in terms of accuracy, is presented in Table 2 below.

Table 2. Accuracy of different methods with different databases.

Method	Database	Acc
LBP [5]	LFW	90.95%
LBP and KNN [5]	LFW	85.71%
SIFT [5]	LFW	98.04%
SURF [5]	LFW	95.60%
SURF and SIFT [5]	LFW	78.86%
PCA [5]	YALE	84.21%
CNN [5]	ORL	95%
Deep CNN [10]	LFW	98.95%
VGGNet [10]	LFW	97.45%
CNN and SVM [10]	ORL	97.50%
AlexNet and SVM [10]	ORL	99.17%
ResNet and SVM [10]	ORL	98.50%
SIFT and KPCA [5]	ORL	80.34%

From Table 2, this paper can give the following conclusion.

1)On the LFW database, models using deep learning architectures, such as Deep CNN and VGGNet perform better than traditional methods such as LBP and PCA to a large extent, even the best-performing Deep CNN outperforms the worst PCA by more than 10%. This is also evidenced by the fact that in the ORL database, the accuracy of the hybrid SIFT and KPCA approach is lower than that of the Deep Learning CNN alone.

2)On the ORL database, the combination of deep learning models with SVM classifiers performs the best, especially the combination of AlexNet and SVM achieving the highest accuracy of 99.17%, which shows that this combination has a strong adaptive ability when dealing with the ORL database. By comparing the CNNs with and without SVM, it can be found that the set with SVM is higher, so this indicates that SVM as a classifier can further improve the accuracy of the model when combined with deep learning models.

3)On the YALE database, the performance of the PCA method is relatively weak, reaching only 84.21% accuracy, which may be due to the inefficiency of the method in dealing with more complex environments on this database, thus proving the limitations of the traditional learning methods in facing complex environments as mentioned above.

4)Traditional feature extraction methods such as LBP, SIFT, and SURF perform differently when combined with SVM. On the LFW database, the accuracy of LBP combined with KNN is 85.71%, which is lower than the 90.95% of LBP alone. The accuracy of SIFT combined with SURF decreases to 78.86%, which is lower than 98.04% and 95.60% of SIFT alone. All these indicate that the combination of feature extraction is not always beneficial to improve the accuracy.

Through the above comparison, it can be found that deep learning performs well on different databases, especially LFW and ORL with a high accuracy rate. The traditional learning methods, such as LBP, SIFT, SURF, etc., although they perform well on some databases, the overall accuracy is not as good as the deep learning models. This may be due to the limitations of traditional methods, which cannot efficiently handle complex tasks when facing problems such as light changes. Moreover, not all combinations of methods are favourable for face recognition, and the combination of feature extraction and classifiers can lead to a decrease in accuracy.

4. Discussion and future research directions.

Face technology is widely used in our daily life and this technology has reached a certain degree of maturity, but face recognition technology still has certain limitations. Furthermore, based on the review of two different stages of algorithms in this paper, it is found that there are still problems such as changes in facial expression, changes in light and darkness, and low resolution that led to unsuccessful

recognition. Among these problems, there is a great concern about the leakage of data privacy to the extent of illegal use of face technology and property leakage due to face synthesis for financial fraud [6,8,11]. To solve these problems, researchers have proposed various approaches to address them and explore possible future directions.

1) Solve problems of posture, lighting, and facial expression changes.

The first step in face detection is to determine whether a given image contains a face or not, which in turn allows for subsequent feature extraction. In this step, the recognition rate is often greatly reduced in response to changing lighting and expressions. Although CNN can improve the accuracy of face recognition in these situations. However, CNN is susceptible to suffering adversarial attacks where small changes in inputs can lead to erroneous results, and CNN is considered a black-box model where understanding how they make decisions can be difficult [4,9]. Furthermore, CNN requires the high hardware that makes the model training time too long and requires constant iterations to optimize the model, resulting in high cost [4,9]. These problems are supported by a number of techniques: quantisation and compression can be used to reduce the computational requirements of CNN by reducing the amount of needed memory for storage and using CNN to remove unnecessary neurons to obtain smaller and more efficient models, thus reducing the demand on GPU [4]. Some researchers also developed some technologies to solve this problem to some extent, such as, Chen Qin et al. used deep CNN VGG16 to extract faces with different postures and achieved good results [9], some researchers used invariant features to process, which means features that do not change with illumination, the effect is good [9] and there are formal CNN that help to ensure the model meets the specifications, and detect the defects of the structure or the training process before and after the deployment to enhance the accuracy of face recognition [4].

In future research, researchers can explore more multimodal fusion techniques such as thermal imaging and 3D techniques to enhance face recognition. Develop more powerful CNN techniques for more complex tasks and environments, such as video recognition. Alternatively, a combination of techniques can be used to increase the efficiency and accuracy of the algorithm, but as mentioned in the experimental section above, not all combinations of techniques are beneficial to face recognition technology, and it is necessary to judge the effectiveness of the combination of methods by the corresponding indicators.

2) Addressing Privacy Issues:

With the popularity of face technology, privacy and security issues have become a growing concern. Currently, data samples are usually stored on smartphones and model training is usually performed on centralized servers [12]. Such a setup can easily lead to unauthorized access. A number of techniques have also contributed to this problem: homomorphic encryption, which is an authentication method using deep learning proposed by Ma et al. that uses CNN to extract facial features, and Paillier encryption to protect the data and synthetic faces, which generates synthetic faces to create an independent database that is not related to the original database, then replacing the real face but retaining the attributes of the original face [13].

In future research, privacy-preserving techniques, such as federated learning can be developed in greater depth. Federated learning can achieve data protection by avoiding the transmission of sensitive user data to a central server and still provide efficient face services while protecting it [12]. It also can supplement legal loopholes and strictly enforce laws to protect public privacy.

3) Resolve occlusion and low-resolution issues.

Typically, a camera is used to record face images and uploaded to a database. Commonly, the object is too far away from the camera or there is an occlusion problem, which leads to unclear and low-resolution images. Moreover, image processing at long distances is computationally intensive and requires powerful hardware. Furthermore, there is a lack of datasets such as LWF, which are taken at a distance for researchers to train models [14]. Some methods have also tried to use adversarial networks and full-product structures, or super-resolution to improve low-resolution face recognition, but these techniques suffer from some limitations such as performance reduction [9,14]. In order to solve the above problems, it is proposed to develop a framework to create a face diffusion database different from

the conventional database for model training [14]. And some researches shown that improving the image sensor of the camera can significantly improve the accuracy, especially in conditions up to 20 meters [14]. In the SIFC paper, automatic pose invariants for face recognition at a distance (FRAD - face recognition at a distance), 3D reconstruction of faces, and classification using LBP were proposed to improve the performance of face recognition at low resolution.

In future research, it is possible to develop sensors and cameras specifically designed for long-distance face detection to increase the resolution of the image. Alternatively, GAN can be used to augment the data to produce high-quality facial images to reduce some of the losses due to low resolution or occlusion.

5. Conclusion

This paper reviews different methods at different stages as well as their advantages and disadvantages, and analyses their features and progress from two perspectives: traditional learning and deep learning, and focuses on methods based on face feature extraction (including geometric features, texture features, local features, etc.). At present, deep learning, especially CNN, has become the mainstream of face recognition and has achieved remarkable results in many fields. However, face recognition technology is still facing challenges such as lighting, posture, expression changes, occlusion, low resolution, etc., and has certain limitations. To better develop this, face recognition technology needs to make further breakthroughs in improving data security, protecting face privacy, adapting to more complex application environments, etc., so that face recognition technology can have a wider range of application scenarios. In the future, attempts to enhance the computational efficiency of algorithms, multimodal fusion, construction of more diversified datasets (e.g., long-distance images), and cross-domain integration (e.g., optical technology) may promote the further development of face recognition technology. In the future, the combination of CNN with other deep learning methods may further advance the progress of AI in the field of data processing combined of multiple models.

References

- [1] Gururaj, H. L. et al. (2024). A comprehensive review of face recognition techniques, trends, and challenges, *IEEE Access*, 12, pp. 107903–107926. doi:10.1109/access.2024.3424933.
- [2] Kaspersky (2020). What is facial recognition & how does it work? Available at: <https://usa.kaspersky.com/resource-center/definitions/what-is-facial-recognition> (Accessed: 24 August 2024).
- [3] Adjabi, I. et al. (2020). Past, present, and future of Face Recognition: A Review, *MDPI*. Available at: <https://www.mdpi.com/2079-9292/9/8/1188>.
- [4] Krichen, M. (2023). Convolutional Neural Networks: A survey, *MDPI*. Available at: <https://www.mdpi.com/2073-431X/12/8/151>.
- [5] Kortli, Y. et al. (2020). Face recognition systems: A survey, *MDPI*. Available at: <https://www.mdpi.com/1424-8220/20/2/342>.
- [6] Zhou, X. and Zhu, T. (2024). *IOPscience, Journal of Physics: Conference Series*. Available at: <https://iopscience.iop.org/article/10.1088/1742-6596/2717/1/012027/meta>.
- [7] Ali-Kutlug, M. A. and Sirin, Y. (2023). Reducing false positive rate with the help of scene change indicator in deep learning based real-time face recognition systems - multimedia tools and applications, *SpringerLink*. Available at: <https://link.springer.com/article/10.1007/s11042-023-15769-0>.
- [8] Facial recognition system (2024). *Wikipedia*. Available at: https://en.wikipedia.org/wiki/Facial_recognition_system.
- [9] Li, L. et al. (2020). A review of Face Recognition Technology, *IEEE Access*, 8, pp. 139110–139120. doi:10.1109.
- [10] Almadby, S. and Elrefaei, L. (2019). Deep convolutional neural network-based approaches for face recognition, *MDPI*. Available at: <https://www.mdpi.com/2076-3417/9/20/4397>.

- [11] Wang, Y. et al. (2024). Machine learning-based facial recognition for financial fraud prevention, Journal of Computer Technology and Applied Mathematics. Available at: <https://www.suaspress.org/ojs/index.php/JCTAM/article/view/v1n1a11>.
- [12] Woubie, A., Solomon, E. and Attieh, J. (2024). Maintaining privacy in face recognition using Federated Learning Method, IEEE Access, 12, pp. 39603–39613. doi:10.1109.
- [13] Laishram, L. et al. (2024). Toward a privacy-preserving face recognition system: A survey of leakages and solutions, ACM Computing Surveys [Preprint]. doi:10.1145/3673224.
- [14] Llaurodo, J. M. et al. (2023). Study of image sensors for enhanced face recognition at a distance in the smart city context - scientific reports, SpringerLink.