Studies Advanced in Development and Application of Facial Expression Recognition

Piao Guo *

Xiamen University, Selangor, Malaysia

* Corresponding author: EEE2109272@xmu.edu.my

Abstract. Facial expression is an important way to convey people's inner emotional changes, as the basis of emotional understanding and the prerequisite for computers to understand human emotions, expression recognition has attracted more and more research attention from academia and industry in recent years. Most of the early expression recognition methods relied on manual features such as texture, geometry, and contour. Thanks to the powerful feature representation capability of convolutional neural networks, deep learning-based face recognition technology has made breakthroughs in recognition accuracy and speed. This paper details the latest research progress in the field of face expression recognition, including the design ideas, key steps, advantages and disadvantages of representative methods. This paper also introduces the common face expression recognition datasets and quantitatively compares the results of different methods on these datasets. Finally, this paper discusses the problems in the face expression recognition research field and look forward to the future development.

Keywords: Facial expression recognition, deep learning, application.

1. Introduction

The degree of automation across the board is rising with the quick growth of technological advances in computers, artificial intelligence technology, and related fields. Conversation between humans is growing in importance, much like the need for human-computer contact does. The connection between humans and computers will be drastically altered by computers and robots that can perceive, feel, and act like humans. This will allow computers to assist humans more effectively. Facial expressions are a vital means of communicating people's inner emotional changes since they serve as the foundation for emotion interpretation and are necessary for computers to interpret human emotions. They also play a significant role in interpersonal communication by allowing people to sense and comprehend the motives of others. Therefore, expression recognition has attracted more and more research attention in academia and industry.

In the 19th century, research on face expressions was underway. The similarities and differences between animal and human facial expressions were described by Darwin in 1872. The first work on modern facial expression recognition was done in 1971 by Ekman and Friesen, who defined six basic human expressions (happiness, sadness, surprise, fear, anger, and disgust) and created the Facial Action Coding System (FACS), which allows researchers to categorize and categorize facial expressions. Action Units (AUs) are used to characterize a face's movements and, via the link between expressions and actions, to recognize subtle movements in the face. Suwa presented automatic facial expression analysis in image sequences and made the first attempt at face expression identification on a face video animation in 1978. A number of studies were conducted on face expression video sequences. Computerized automated processing of facial expression identification became feasible in the mid-1990s with the advancement of image processing and pattern recognition technologies. After applying the suggested optical flow method for facial expression recognition, the recognition rate was close to 80%, and automatic facial expression recognition entered a new era. Mase et al. were the pioneers who first used optical flow to determine the main direction of muscle movement.

The expression recognition technology based on machine learning and deep learning keeps making advancements in recognition accuracy and speed because of the quick growth of machine learning technologies like convolutional neural networks. In order to characterize the face information, texture,

geometry, contour, and other characteristics are extracted. Classifiers like support vector machines (SVM) are then used to recognize various expression pairs. This is the fundamental notion of machine learning-based expression recognition. However, limited by the insufficient granularity of manual features to describe subtle expressions, it cannot meet the needs of practical applications. Most of the recent works are along the framework of deep learning, which utilizes the powerful high-dimensional nonlinear variation ability of convolutional neural networks to adaptively extract discriminative face features, thus realizing accurate face expression recognition. The current body of study on face recognition is divided into five primary groups based on differences in design ideas: (1) methods based on geometric features; (2) methods based on statistics; (3) methods based on subspace; (4) methods based on elasticity theories; and (5) methods based on neural networks.

Focusing on the aforementioned aspects, this paper presents an in-depth research and analysis of the research on face expression recognition and presents the latest research progress in detail. Specifically, this paper first introduces representative recognition algorithms of different categories, including their design ideas, key steps, advantages and disadvantages. Second, the common face expression recognition datasets are introduced and the results of different methods are quantitatively compared. Finally, this paper discusses the problems in the face expression recognition research field and look forward to the future development.

2. Method

In order to better understand the development and application of facial expression recognition technology, two development perspectives of expression recognition, horizontal and vertical, can be an important starting point. For the horizontal aspect, comparing the existing different kinds of expression recognition methods and means, analyzing and summarizing the advantages and disadvantages of each, which is conducive to combining different methods to achieve better recognition application effects. Vertically, it is also an important part of the development process of expression recognition to deduce the future development trend and possible difficulties through the historical development and application of expression recognition.

2.1. Geometric Features Based Facial Recognition Methods

The method of recognizing face expressions by geometric features is one of the earliest face recognition methods [1]. The location of key feature points, such the mouth, nose, and eyes, as well as the geometrical features of key organs, like the eyes, are typically extracted when classifying faces using this method. Important characteristics for distinguishing between faces include the mathematical description of shape and structural relationships. In essence, the extracted geometric characteristics of a face are matches between feature vectors, which typically comprise the length, curvature, angle, and other parameters between two places that the face specifies. The recognition process is comparatively straightforward and easy to comprehend, but in essence, because feature extraction makes it possible to overlook local subtleties, some information is lost, and accuracy needs to be increased.

2.2. Statistically Based Facial Recognition Methods

The Hidden Markov (HMM) method and the KL algorithm are two statistically based facial recognition techniques. Since the KL transform views a high-dimensional vector created by unfolding a face image in rows (columns) as a random vector, it can be used to identify the orthogonal K-L base of a high-dimensional vector, the bigger eigenvalue base of which has a face-like shape. The Pentland group at MIT's Media Laboratory is one of the more prominent organizations in the field of face identification using still or video pictures. They specialize in KL transform-based feature extraction in eigenspace, or "Elgenface".

The Cambridge University researchers Samaria and Fallside proposed the Hidden Markov Model (HMM), whose parameters are the eigenvalues, on a spatial sequence of several sample images.

Initially, Samatia employed 1-D HMM and 2-D Pseudo HMIM for facial identification, utilizing the facial features that run from top to bottom and left to right. As seen in Fig 1, Kohir achieved good recognition results by using low-frequency DCT coefficients as observation vectors [2].

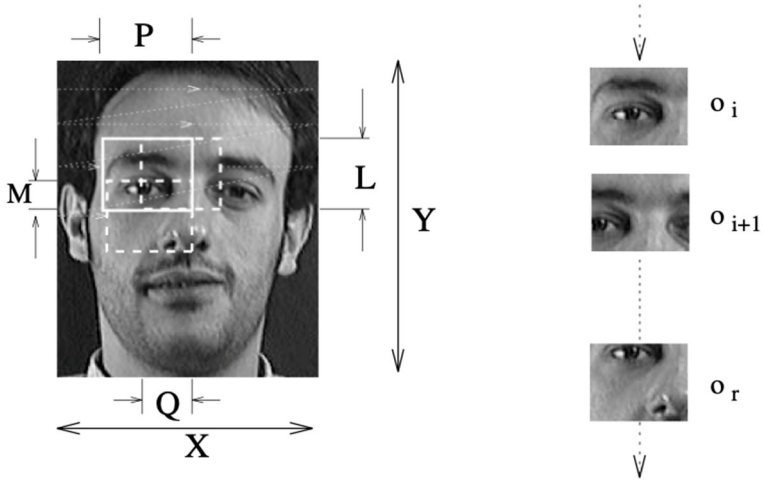


Figure 1. Sampling technique for an ergodic HMM

Eickeler used 2-DPseudo HMIM to recognize face images in DCT compressed JPEG images; Nefian used embedded HMM to recognize faces, as shown in Fig 2 [2].

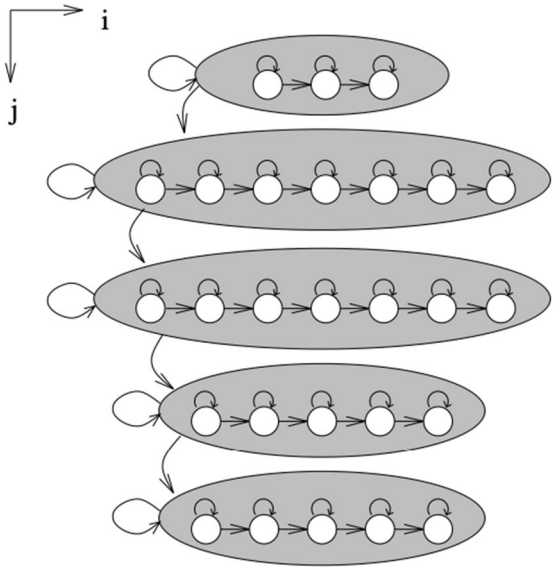


Figure 2. Structure of a P2D-HMM

Later integrated coupled HMIM and HMM form a hybrid system structure by using different models for the superstate and each embedded state. The HMIM-based facial recognition method has three primary advantages: (1) greater recognition rate; (2) good scalability (i.e., adding new samples does not need training of every bit of data); and (3) the capacity to allow the face to have larger head rotations and expression changes.

2.3. Subspace-based Facial Recognition Methods

Eigen subspace, distinguished subspace, independent component subspace, and so forth are instances of frequently used linear subspace techniques. Furthermore, there exist techniques for local eigen analysis, factor analysis, and other related fields. Additionally, hybrid linear and nonlinear subspaces have been included in the application of these techniques.

Turk was the first to apply the Eigenfaces approach to recognize faces [3]. The fundamental principle of the Eigenfaces analysis technique is to determine the eigenvectors of the sample set of face image variance matrix in order to do an approximation statistical analysis and characterize the

face image. To get the second order eigenspace, sometimes referred to as the second order eigenface, the difference pictures of the original and reconstructed images perform another K-L transformation [4]. Overall, the structural relationships of the face and the information inferred within the set of face samples are reflected by the eigenface approach. The advantage is that it reduces the amount of input data, but it still essentially relies on the training set and test set images, which has significant limitations. Alber proposed the TPCA (Topological PCA) method, which has improved the recognition rate [5]. The method known as the land area Local Feature Analysis was proposed by Penev. Compared to the eigenface technique, the recognition results are superior [6]. Belhumeur proposed the Fisher aces method, which produced better recognition results, based on Linear Discriminant Analysis (LDALinear Discriminant Analysis). The Eigenspace method does not account for the information between the sample categories when there are multiple sample images for each person. Bartlett recognized faces using the Independent Component Analysis (ICA, Independent Component Analysis) approach, which produced superior recognition results than the PCA method.

2.4. Elasticity model based facial recognition methods

Lades proposed a method to recognize faces using Dynamic Link Architecture (DLA, Dynamic Link Architecture) [3]. It puts the face in a lattice-like sparse graph as shown in Fig 3.

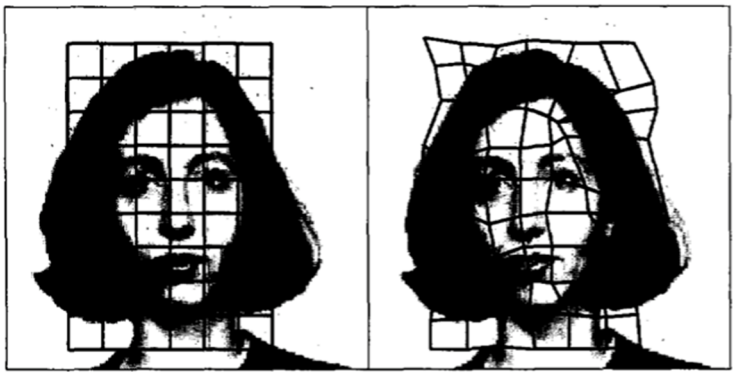


Figure 3. Dynamic Link Architecture

The edges of the graph in Figure 3 are labeled with distance vectors connecting the nodes, while the nodes themselves are labeled with feature vectors derived from the Gabor wavelet decomposition of image locations. Wiskott employed a 97.3% accurate elastic graph matching technique. Wiskott created an elastic graph by using several places on the face's characteristics as datums [7]. The system's storage capacity is decreased when a string of representative feature vectors is stored at each datum point. Wurtz employs a multilayered hierarchical structure and only uses features from the IC region of the face, further removing background and superfluous information from the structure [8]. Grudin also uses a hierarchically structured elasticity graph, which creates a sparse description of the face by removing some redundant nodes from the structure [9]. In addition to this, Nastar proposed to represent face images as a deformable 3D mesh table (x,y), convert the face matching problem into a surface matching problem, deform the surface using finite analysis, and recognize the face based on the degree of deformation matching between the two images [10]. In this method, the face images are modeled as deformable 3D mesh surfaces, thus transforming the face matching problem into an elastic matching problem with deformable surfaces. The algorithm better utilizes the structure and distribution information of the face and has a high recognition rate. However, it has the disadvantages of high time complexity, slow speed and complex implementation.

2.5. Neural network based facial recognition methods

An artificial neural network (ANN) is a type of network structure that imitates the structure of human brain neurons. Neurones are an adaptable nonlinear dynamic system made up of several basic,

interconnected neurons. A hybrid neural network was proposed by Gutta [11]. In order to achieve sample clustering, Lin used a convolutional neural network CNN; Lawrence used a multilevel SOM; Demers proposed using a principal neural network approach to extract features from a face image, then compressing those features using an autocorrelation neural network, and utilizing an MLP for face recognition [12]. Lin also used a neural network approach based on probabilistic decision making. Er used PCA for dimensional compression, then LDA is used to extract the features and then face recognition is performed based on RBF. Haddadnia performs face recognition based on PZMI features and uses RBF neural network with hybrid learning algorithm [13].

Using the coordinate positions of the face features and their corresponding gray values as inputs to the neural network, the ANN can provide complex inter-class interfaces that are hard to imagine. Artificial neural network is a nonlinear dynamical system with good self-organization and self-adaptation ability, which has a great advantage over appealing several methods.

3. Experiment And Performance Analysis

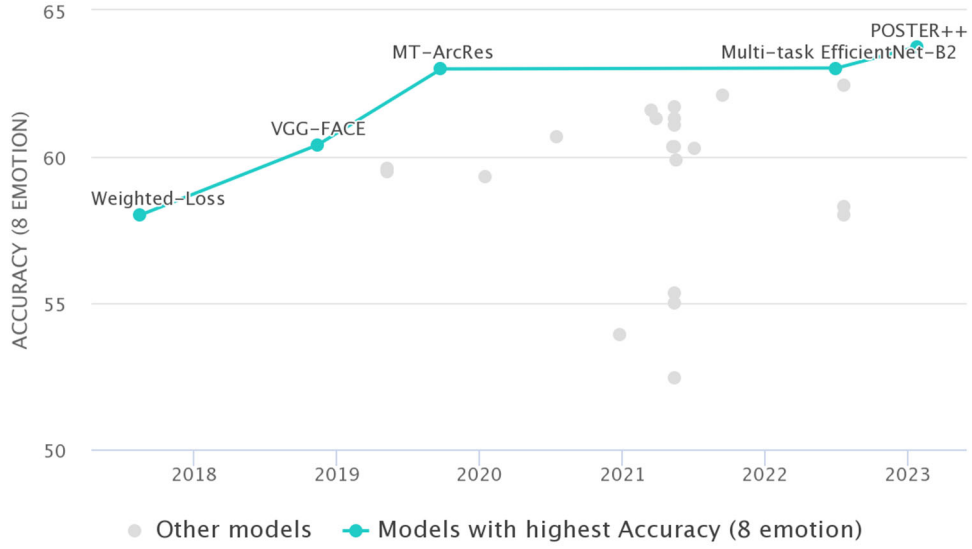
3.1. Common Datasets

This section introduces the existing commonly used facial expression recognition datasets, including the FER2013, AffectNet, and The Real-World Emotional Faces Database (RAF-DB).

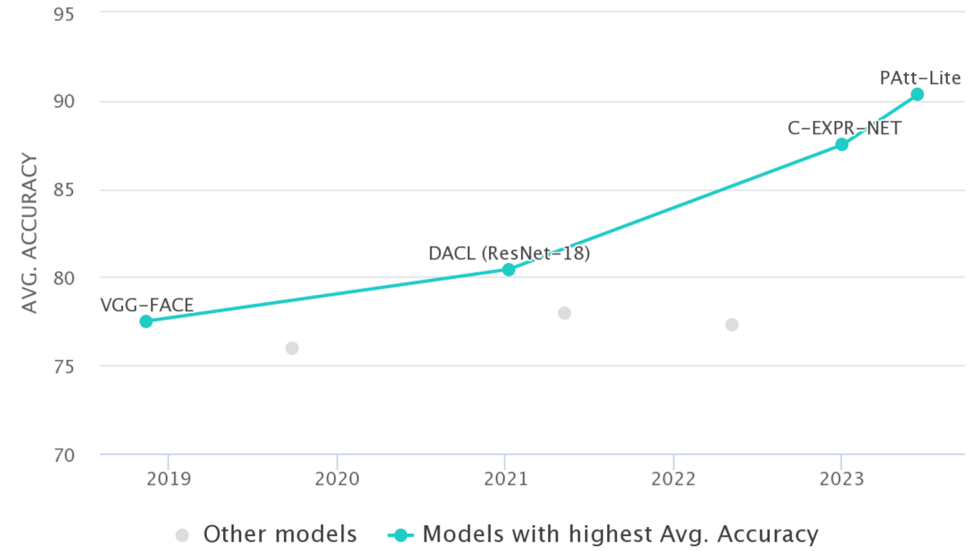
Anger, disgust, fear, happiness, neutrality, sorrow, and surprise are the seven basic facial expressions found in FER2013. The International Face Recognition Competition and academics at the University of California, Los Angeles (UCLA) have contributed the dataset (FERET). The FER2013 dataset is split into three sections: training, validation, and testing. It comprises 35, 887 48x48 pixel images. There are 28709 photos in the training section, 3589 in the validation area, and 3589 in the testing section. The appropriate face expression category was assigned to each image. Approximately 400 million manually categorized photos with eight different facial expressions-neutral, happy, angry, sad, afraid, shocked, disgusted, and contemptuous-make up the massive facial expression database known as Affect Net. A human annotator examines the photographs' emotions to determine these emotion descriptors. Additional data including face location, pose, and image quality scores are also provided by Affect Net. A collection of face expressions is called the Real-World Emotional Faces Database (RAF-DB). It has 29,672 face pictures with basic and compound expression labels applied by 40 different markers. This database arranges the photographs according to the following criteria: participants' age, gender, and ethnicity; head pose; lighting circumstances; occlusion (such as spectacles, facial hair, or self-occlusion); post-processing alterations (such as different filters and special effects); and so on.

3.2. Performance Comparison

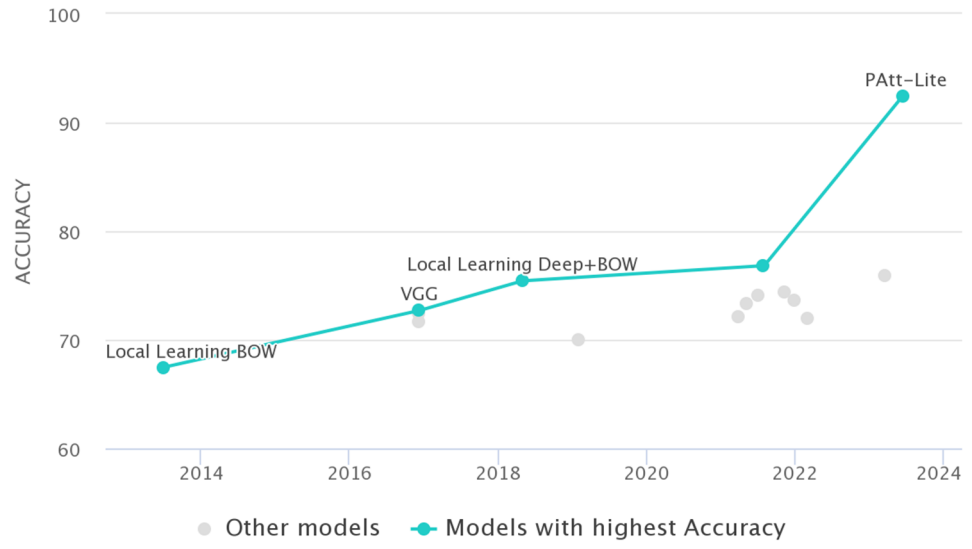
In order to visualize the recognition accuracy of the face expression recognition methods, additional experiments are conducted in this section to quantitatively compare the results of representative methods on different datasets, as shown in Fig 4. On AffectNet, it can be observed that the best performance comes from POSTER++, which can achieve a accuracy of 63.7%. Compared to the second-place method EfficientNet-B2, it obtains a gain of 0.7%. On RAF-DB dataset, the average is more promising than the challenging AffectNet, where simple VGG can obtain a accuracy of 77.5% while the best PAtt-Lite can obtain a accuracy over 90%. PAtt-Lite introduces the lightweight patch and attention MobileNet to produce a more powerful feature embedding, which can explain the significant improvement for recognition. For the classical FER2013 dataset, PAtt-Lite also can achieve the best accuracy of 92.5%, which outperforms previous state-of-the-art method by a large margin, such as 19.8% than VGG.



(a) facial expression recognition on AffectNet



(b) facial expression recognition on RAF-DB



(c) facial expression recognition on FER2013

Figure 4. Recognition accuracy of different methods on three datasets

4. Discussion

Facial expression recognition is based on facial recognition technology and research. Since humans have not had much experience with facial expression recognition, it is more difficult to recognize facial expressions because of their diversity, complexity, and involvement of physiology and psychology. As a result, compared to other biometrics like fingerprint, iris, and face recognition, facial expression recognition has developed more slowly and is not yet widely used. As a result, the development is moving more slowly than other biometrics like fingerprint, iris, face, and so forth, and the application is still relatively fresh. Future prospects and challenges for facial expression recognition technology will be greater given its complexity.

First of all, it is a given that facial expression recognition technology, still in its infancy, will advance. In the future, this technology will be more intelligent, adaptable, and able to handle more complex scenes and data, improving recognition accuracy and reliability.

Secondly, there is a large market and development space for expression recognition, and what needs to be done next is to apply expression recognition to actual scenes and combine it with real needs. For example, in the production of games, it can make real-time reflections based on human emotions to enhance player immersion; in terms of safe driving, it can be based on the driver's expression to determine the driver's driving status to avoid accidents. In terms of public safety monitoring, it can be used to prevent crime by judging whether there are abnormal emotions based on expressions. In distance education, teachers can indirectly infer students' just mastery ability through real-time expression recognition of students. Expression recognition is a very promising direction of development, will be closely linked with the daily needs of such products need to consider the important factors, rather than just give a test result, perhaps one of the future direction of development.

In addition, the standardization and openness of facial expression recognition will be further promoted. In the future, facial expression recognition technology will establish more standardized and open standards, thus promoting the interoperability and repeatability of the technology, and improving the reliability and sustainability of the technology.

Even though expression recognition technology has advanced significantly, practical application still faces challenges. Improving the elements that follow is necessary in addition to concentrating on and achieving accurate and quick face segmentation and detection. First, more thorough study and a more precise description of the expression pattern are necessary for facial recognition using the hybrid model method. Second, despite the similarities between faces and expressions as well as the impact of different lighting conditions, angles, and other variables, it is still challenging to accurately recognize facial expressions. Third, 3D deformation modeling has a bright future ahead of it and can adapt to many changing conditions. It has also been demonstrated that improved outcomes are obtained by using simulation or compensating for different changing factors. While still in its exploratory phase, the selection of 3D face recognition algorithms requires innovation and improvement based on the first traditional recognition algorithms. Ultimately, even though the surface texture recognition algorithm is relatively new, it has a lot of potential for facial emotion identification, thus further learning and study is needed to find a better approach.

5. Conclusion

The current focus of expression recognition shifts to challenging real scene conditions, where deep learning techniques are utilized to solve problems such as lighting variations, occlusion, and non-frontal head poses. Another major issue to be considered is that, although expression recognition techniques have been extensively studied so far, the expressions we have defined cover only a small fraction of specific kinds, mainly facial expressions, while in reality there are many other human expressions. The study of expressions is much more difficult and the applications are much broader compared to face age and so on, and more and more applications will surely emerge for these. In summary, face recognition is an extremely difficult topic. It is difficult to achieve good recognition

results simply using the current method. Future research should focus on how to combine with other technologies, increase recognition rate and speed, decrease computation, use embedded and hardware implementation, and be practical.

References

[1] Cheng Y. Liu K, Yang J, etal. Human face recognition method based on the statistical model of small sample size. SPIE Proc, Intell. Robots and Computer Vision X: Algorithms and Techn. 1991, 1606: 85 - 95.

[2] Samaria F S. Face recognition using hidden Markov models [D]. University of Cambridge, 1994.

[3] M.Lades, J. C. Vorbruggen, J. Buhmann, ect. Distortion invariant object recognition in the dynamic link architecture. IEEE Trans. on Computer, 1993, 42 (3): 300 - 311.

[4] Nastar C, Moghaddam B A. Flexible Images: Matching and Recognition Using Learned Deformations [J]. Computer Vision and Image Understanding, 1997, 6 5(2): 179 - 191.

[5] Tibbalds A D. Three-dimensional human face acquisition for recognition [D]. University of Cambridge, 1998.

[6] Wenyi Zhao. Robust image-based 3D face recognition [D]. PhD. Thesis. University of Maryland, College Park, 1999.

[7] Wiskott L, Fellous J M, Krüger N, et al. Face recognition by elastic bunch graph matching [M]//Intelligent biometric techniques in fingerprint and face recognition. Routledge, 2022: 355 - 396.

[8] Wurtz R H. Neuronal mechanisms of visual stability[J]. Vision research, 2008, 48 (20): 2070 - 2089.

[9] Grudin M A. On internal representations in face recognition systems [J]. Pattern recognition, 2000, 33 (7): 1161 - 1177.

[10] Nastar C, Mitschke M. Real-time face recognition using feature combination [C]//Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. IEEE, 1998: 312 - 317.

[11] Gutta S, Wechsler H. Face recognition using hybrid classifiers [J]. Pattern Recognition, 1997, 30 (4): 539 - 553.

[12] DeMers D, Cottrell G. Non-linear dimensionality reduction [J]. Advances in neural information processing systems, 1992, 5.

[13] Haddadnia J, Ahmadi M. N-feature neural network human face recognition [J]. Image and Vision Computing, 2004, 22 (12): 1071 - 1082.