

A survey of face recognition based on machine learning

Yufei Wang

Southwest Jiaotong University-University of Leeds Joint School, Chengdu, Sichuan
Province, China, 611730

sc212yw@leeds.ac.uk

Abstract. Face recognition technology based on machine learning methods is the most popular nowadays. This research, based on existing literature and data, was mainly focused on the face recognition system. The paper is written on the basis of existing face detection methods. Then, according to the difference in feature information, five methods were used to extract face features to apply them to face recognition. Then, this paper focused on two kinds of neural networks: support vector machines (SVM) and integrated learning methods, especially the Adaboost algorithm. Finally, the hidden danger of privacy disclosure was specially discussed. The results showed that the main direction of future research should be to expand the functional dimension and promote more applications to the public on the basis of ensuring the real-time and accuracy of the system.

Keywords: Face Recognition, Artificial Intelligence, Machine Learning, Survey, Face Detection, Feature Extraction.

1. Introduction

The research on face recognition can be traced back to the paper published by Galton in the science journal *Nature* in 1888 [1]. In 1910, Galton proposed transforming the portrait into a feature vector by using the reference point of the human face and the spatial relationship between the reference points and identifying the identity through the feature vector [2]. However, this was the rudiment of face recognition based on geometric features, not automatic face recognition (AFR). The earliest research on automatic face was recognized to have begun in the 1960s. The technical report, published by Bledsoe and Chan in 1965, opened with its prelude [3]. So far, research on face recognition has gone through several decades. With the passing of years, the technology of face recognition is also constantly updated. Early classification methods based on facial geometric feature extraction had poor performance. Until the 1980s, neural networks was successfully applied to face recognition, and then it was effectively improved in scientific research, gradually eliminating the technology of face recognition based solely on facial geometrical features [4]. Since then, the research field of machine learning has received more attention from face recognition researchers. Nowadays, the face recognition algorithm based on machine learning has become the focus of face recognition research because of its good recognition performance. This paper mainly studies face recognition skills on the base of machine learning and achieves purposes of the survey by summarizing and analyzing some existing relevant research results. In this paper, we studied the representative algorithms of each part of the face recognition system in detail, explained their applicable scenarios based on the principle,

and objectively evaluated the advantages and disadvantages of different methods. These studies not only meet some of the needs of scholars to explore face recognition technology but also enhance human understanding of the visual system itself, laying a solid foundation for the widespread application of this technology in life.

2. Face recognition system

The face recognition system is mainly composed of face detection, feature extraction and face recognition.

2.1. Face detection

Face detection is to search the input image to determine whether it contains faces. It is the premise and foundation of face information processing. After continuous updating and development, typical face detection methods include the feature invariant method and the template matching method.

2.1.1. Feature invariant approaches. The invariant face contains many inherent features, such as color features, contour features, structural features and so on. The feature invariant approach is to correctly select and use these features. It is mainly divided into two categories: single feature analysis method and combined feature analysis method. The classification is based on the type and number of selected features [5]. The single feature analysis method is a feature vector composed of a single type of features. It needs to map the feature value of a single feature to the face object according to certain rules. Gruhazali et al. established a skin color Gaussian model based on skin color features to match and calculate candidates [6]. Because the model adopts the skin color feature of fixed value, the robustness of face detection is poor when the illumination changes greatly, and it is easy to generate false detection [4]. The combined feature method is to combine multiple features of the face to form a face feature vector, so that the feature representation ability is more comprehensive, thus improving the accuracy of face detection. Chakraborty formulated rules for face detection based on the relationship between face shape and skin color [7]. The addition of face shape features improves the effect of face detection under posture changes, while it is complicated to extract multiple features and establish the relationship between features, which is likely to miss detection.

2.1.2. Template matching methods. The main template matching methods include the fixed and deformation template methods [8]. The fixed template method is mostly applied to the early stage of face detection exploration. It is based on a pre-established standard template library that contains multiple sub templates describing the local features of the face. It is needed that, firstly, matching the input image with template windows of different sizes, calculating the correlation coefficient between the matching result and different parts of the standard template, and then comparing the correlation coefficient with a preset threshold to judge whether the image window contains faces. The advantage of this method is that it is easy to implement. However, because there are huge changes in face features, it has bother in obtaining an valid template to exemplify the commonality of faces, and it also can not effectively handle the changes in scale, posture, and shape. Therefore, its practicability is not strong. The deformable template method is composed of adjustable parameter templates and corresponding energy functions. The parameter template with the minimum energy function value is obtained by the nonlinear optimization method, i.e., the optimal matching template. This method takes full account of the characteristics of the human face as a deformable body and is independent of posture and illumination, so it is stable and reliable.

2.2. Feature extraction

The essence of feature extraction is to extract discriminative features that are different from other individuals from images containing faces, mainly including geometric and algebraic features [9]. According to different feature information, researchers divided feature extraction methods into

methods consisting of priori rules, color information, geometric shape, statistics, and association information.

2.2.1. Methods based on priori rules. The priori rule is an empirical description of human facial feature points. Compared with the unique features of the human face, it is more general. The feature extraction method based on a priori rule does not extract the single organ features of the face but extracts the relationship between the facial organs. Asymmetry and gray difference of facial organs are examples of common prior rules [5]. The implementation process of this method is relatively easy, but it must have strong dependence on stored prior rules, so it has great limitations in practical application. For example, only one person can appear in the image, the resolution cannot be too low, and the external environment cannot change too much. The methods based on priori rules are only suitable for rough location of features, and other methods need to be combined in feature extraction.

2.2.2. Methods based on color information. The essence of this method is to model and match the color information of human face with statistical methods. After the face color information model is established, the color information of the face to be detected is extracted, and the possible position of the target point is determined according to the matching degree between the color information of the point to be detected and the model. This method is widely used because of its low computational complexity, but it is affected by many factors when extracting features [5]. For example, the complex color feature structure of the human face makes modeling difficult, and the method based on color information has poor robustness to illumination changes, which reduces its stability.

2.2.3. Methods based on geometric shape. According to the geometric features of the face image, this method constructs a geometric parametric statistical model based on variable parameters. The variable part of the target object feature information is represented by the variable parameters in the model [9]. At the same time, an energy evaluation function matching the model is set to measure the fit between the image detection area and the statistical model. When the energy function value is the smallest, the statistical model will converge to the characteristic shape area of the image. This method may fully utilize the gray and texture information of the image to easily find the shape features around the facial features. However, the model is highly dependent on the initial position and parameters, which leads to the algorithm easily falling into the local optimum, thus deviating from the ideal feature position.

2.2.4. Methods based on statistics. The method based on parametric statistics is to treat the feature part as a class pattern and train the feature class samples and non feature class samples respectively, so as to build a classifier with different effects, which extracts the feature information of the target object [9]. Such methods mainly include principle component analysis (PCA), AdaBoost iterative algorithm, support vector machine (SVM), etc.

2.2.5. Methods based on association information. The four methods that have been described above mainly extract features from local images. Due to the limited information types of local images, there may be a large number of similar feature information, which interferes with feature selection. The feature extraction method based on association information can narrow the range of candidate feature points by using the correlation between different features and extract features more accurately.

2.3. Face recognition

Back propagation (BP) neural network, radial basis function (RBF) neural network, support vector machine (SVM) and ensemble learning are the four most commonly applied machine learning means at present, and are the most widely used machine learning techniques in face recognition research as well [10].

2.3.1. Back propagation neural network. Back propagation (BP) neural network is divided into three types of layers, as shown in Figure 1, namely the input layer, hidden layer, and output layer [11].

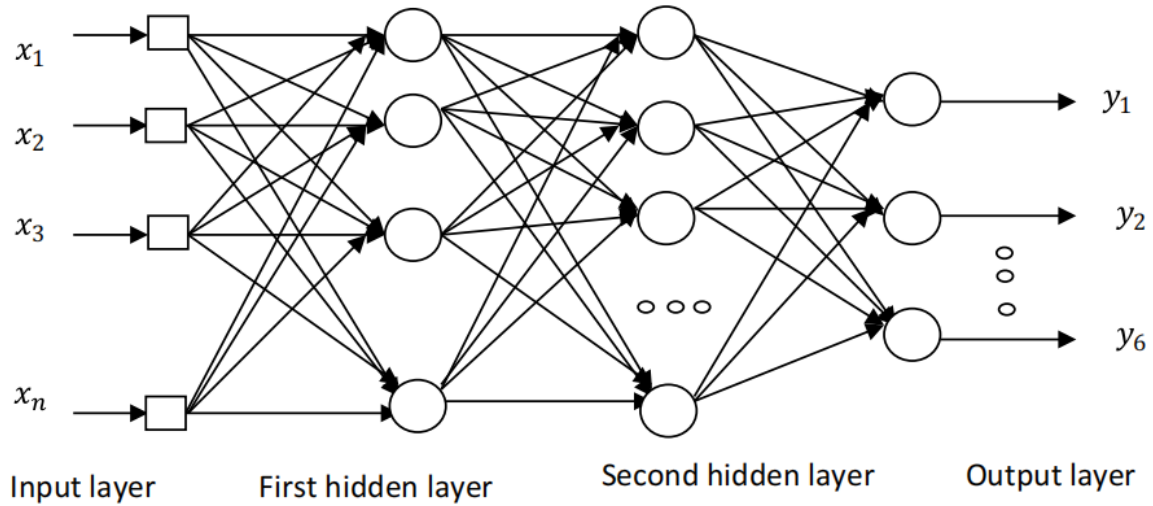


Figure1. BP neural network structure [11].

BP algorithm includes the procedures of forward propagation and error back propagation. In the forward propagation, the eigenvalue is transferred from the input layer, processed layer by layer by the hidden layer, and then transferred to the output layer. If the actual output is the same as the expected output, the learning algorithm is ended. Furthermore, if the actual value is not consistent with the expected value, the error is back propagated, and the error of the output layer is back propagated layer by layer through the hidden layer to the input layer in some form. In the process, the error is assigned to each node of each layer in the network so as to obtain the error signal of each cell as the basis for revising the weight of each cell. These two processes run round and round. The procedure of continuous weight adjustment is the training and learning process of neural networks. The trained neural network could approximately realize an arbitrary continuous nonlinear mapping from input to output, transform the input space into a space stretched by output, make the classification of output space simple, and be suitable for solving problems with complex internal mechanisms [12]. BP neural network itself also has some defects. For example, because the learning rate is fixed, the network has slow converging and long training time. The weights converged by the gradient descent method are light to sink into local minimum. In addition, the number of layers and cells of the hidden layer could only be fixed by experience or repeated experiments, which may produce greater redundancy, increasing the learning burden of the network [12].

2.3.2. Radial basis function neural network. The radial basis function (RBF) neural network is also composed of three types of layers, but only one hidden layer can be included. The input layer is directly linked to the hidden layer instead of being linked by weight. The hidden layer neurons generally adopt Gaussian function as the activation function, namely

$$R_i(x) = \exp\left(\frac{-\|x - c_i\|^2}{2\sigma_i^2}\right) \quad (1)$$

where $\|\cdot\|$ represents the Euclidean norm in the input space and σ_i is the width of the i th RBF unit [13].

The j th output $y_j(x)$ of an RBF neural network is

$$y_i(x) = \sum_{i=1}^u R_i(x) \times \omega(j, i) \quad (2)$$

where $R_0=1$, $\omega(j,i)$ is the weight or strength of the i th receptive field to the j th output and $\omega(j,0)$ is the bias of j th output. If the network complexity was intentionally reduced, the bias could not be considered in the analysis [13].

It can be seen from (1) and (2) that the mapping from the hidden layer to the output layer is linear, that is, the output of the network is characterized by a linear discriminant function. They generate linear decision boundaries (hyperplanes) in the output space. Therefore, the property of RBF neural networks is mainly decided by the separability of classes in n -dimensional space generated by nonlinear transformation performed by n RBF units [13]. Compared with the BP neural network, RBF neural network has significant advantages in strong nonlinear approximation ability, fast convergence speed, and global convergence [14]. The training of the RBF neural network is also simpler, because only the weight needs to be corrected through learning among its undetermined parameters. In the classification problem, the RBF neural network classifier will take precedence over the BP neural network classifier [10].

2.3.3. Support vector machine. Support vector machine (SVM) is a generalized linear classifier (GLC) that performs binary classification on samples in the light of supervised learning, the main principle of which is to find an optimal hyperplane in the space where the training set samples are located to separate the two types of samples in the training set. The optimal hyperplane is defined as the linear decision function with the maximum margin between the two classes of vectors, as shown in Figure 2 [15]. The linear classifier whose function is defined is known as the maximum-margin classifier. To construct such an optimal hyperplane, only minor training data, namely the support vector (SV), needs to be considered, which determines the margin. Sometimes the data is not linearly separable, and the data points on the wrong side are called outliers. In this case, the hinge loss function is used to compute the empirical risk of a soft-margin SVM.

Structural risk minimization principle (SRM) minimizes the upper limit of Vapnik-Chervonenkis (VC) dimension, which makes SVM based on SRM has better generalization ability than artificial neural network based on empirical risk minimization principle (ERM), and therefore has advantages in classification and regression problems [10].

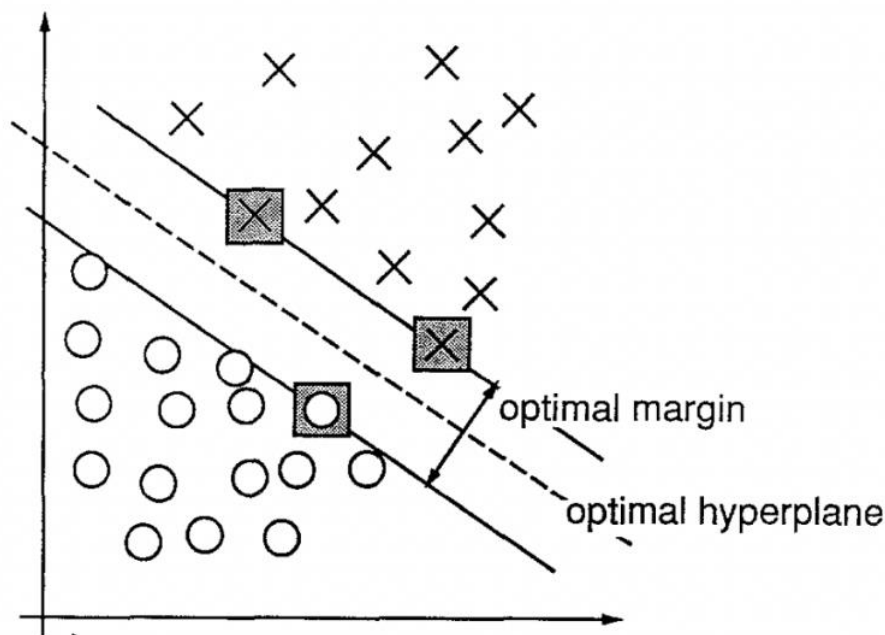


Figure2. Support vector machine (SVM) [15].

2.3.4. Ensemble learning. Ensemble learning means that in supervised learning, several single classifiers (weak supervised model) are integrated (forming a strong supervised model), and the classification results of multiple classifiers jointly determine the final classification, so as to achieve more comprehensive performance than a single classifier. In the distribution-free model (also known as the possible approximate correct or PAC model), if there is a polynomial learning algorithm which could learn a certain group of concepts and has a high accuracy, then this algorithm is a strong learning algorithm; However, if there is another learning algorithm which could also learn this group of concepts, but the correct rate is only slightly higher than the random guess, then this algorithm is a weak learning algorithm [16].

AdaBoost is designed to settle the classification issues. The main idea is to give a weak learning algorithm and a training set, assign equal weight $1/N$ to each group of training data at the same time of initialization, and then use the learning algorithm to drill the training set for T rounds [10]. After each drill, the training data that fails to be trained is given a greater weight, and on the contrary, and vice versa, so that the learning algorithm can focus on the difficult training data in the subsequent rounds to obtain a prediction function sequence. Among them, each prediction function has its own weight. The weight with a good prediction effect is larger, and vice versa. The final prediction function merges all predictions through a weighted majority vote to generate the final prediction result [4].

Viola and Jones applied AdaBoost to face detection for the first time [17]. They constructed multiple weak classifiers using multiple rectangular block features on the face image, as shown in Figure 3 [17]. Where, (a) and (b) show boundary features, (c) shows thin line features, and (d) show diagonal features. Then, AdaBoost is used to merge and promote these weak classifiers to generate a strong classifier, and thus the world's first real-time face detection system is designed. Later, AdaBoost was extended to the field of face recognition successfully as well.

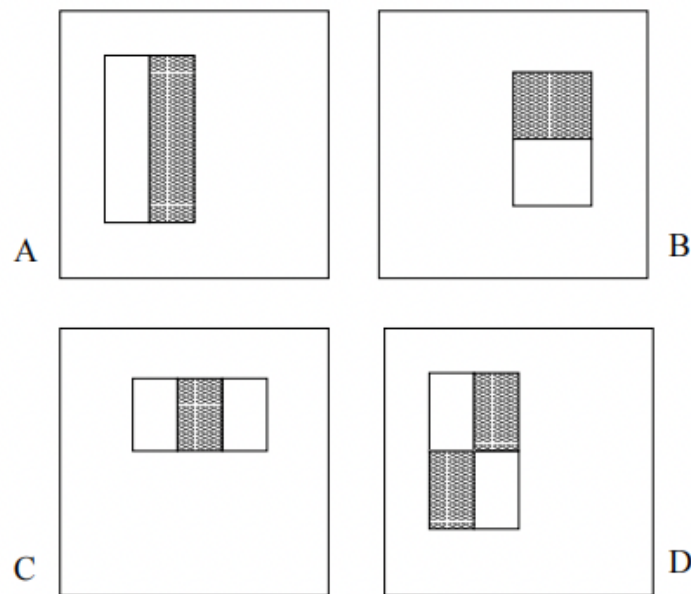


Figure3. Rectangular block features on the face image [17].

3. Discussion

In the Internet era, balancing technology empowerment and privacy protection is a new problem faced by face recognition. We can neither give up the application of technology for privacy security nor abuse it without restraint. Data security is always the key. By focusing on data flow, strengthening legal supervision, optimizing product privacy standards, and cultivating people's privacy literacy, the risk of personal privacy disclosure could be effectively reduced [18]. Besides, ensuring the real-time

and accuracy of the face recognition system and even further expanding the functional dimension of recognition are all the main goals of future research.

4. Conclusion

Face recognition is a biometric recognition technology based on facial feature information, which is most widely used in the field of artificial intelligence in recent years. The article introduces and analyzes the face recognition system from three parts on the basis of machine learning algorithm: face detection, feature extraction, and face recognition. From this, it is concluded that the prospect of face recognition lies in improving accuracy and expanding dimensions. Although this paper has made some achievements through theoretical analysis, there is still a lot of research space in the future. For example, for the sake of the integrity of the article structure, this paper discusses almost all links of the face recognition system but does not carry out in-depth research one by one. The next work will focus on experiments, and more practical analysis will be carried out according to the experimental data set.

Acknowledgment

Firstly, I desire to express my sincere appreciation to the professors, both of my university and the overseas project team, whose wonderful courses have provided valuable materials and guidance for my thesis writing. Further, I am also grateful for the support and encouragement comes from all my families, friends and classmates. Without their enlightening suggestions and impressive goodwill, it would be difficult for me to successfully complete my thesis.

References

- [1] Galton, F.S. (1888). Personal Identification and Description. *Nature*, 38, 173-177.
- [2] Galton, F.S. (1910). Numeralised Profiles for Classification and Recognition. *Nature*, 83, 127-130.
- [3] Bledsoe, W. W., and Chan, H. (1965). A Man-Machine Facial Recognition System-Some Preliminary Results, Technical Report PRI 19A, Panoramic Research, Inc., Palo Alto, California.
- [4] Tianwen, Z. (2008). Research on face recognition based on machine learning methods. Doctoral dissertation, Shanghai Jiaotong University.
- [5] Shengxi, L. (2018). Research on face detection and feature extraction algorithm in face recognition. Doctoral dissertation, Nanjing University of Posts and Telecommunications.
- [6] Ghazali, K.H., Ma, J., Xiao, R., & Lubis, S.A. (2012). An Innovative Face Detection Based on YCgCr Color Space. *Physics Procedia*, 25, 2116-2124.
- [7] Chakraborty, D. (2012). An Illumination Invariant Face Detection Based on Human Shape Analysis and Skin Color Information. *Signal & Image Processing : An International Journal*, 3, 55-62.
- [8] Kunhao, W. (2006). Face detection based on machine learning. Doctoral dissertation, Jilin University.
- [9] Zheng, J. (2016). Research and implementation of feature extraction algorithm in face recognition. Doctoral dissertation, Nanjing University of Posts and Telecommunications.
- [10] Changsheng, Y., & Liang, T. (2009). Performance comparison of several machine learning methods in face recognition. *Computer Engineering and Applications*, 45(4), 4.
- [11] Moganam, P. K., & Sathia Seelan, D. A. (2022). Deep learning and machine learning neural network approaches for multi class leather texture defect classification and segmentation. *Journal of Leather Science and Engineering*, 4(1).
- [12] Shan, Q. (2014). Research on Face Recognition Method Based on BP Neural Network. Doctoral dissertation, China University of Geosciences (Beijing).
- [13] Er, M. J. , Wu, S. , Lu, J. , & Toh, H. L. (2002). Face recognition with radial basis function (rbf) neural networks. *IEEE Trans Neural Netw*, 13(3), 697-710.
- [14] Zhong, W. (2005). Research on Face Recognition Based on Gabor Wavelet and RBF Neural

Network. Doctoral dissertation, Fuzhou University.

- [15] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- [16] Kearns, M., & Valiant, L. (1994). Cryptographic limitations on learning boolean formulae and finite automata. *Association for Computing Machinery. Journal of the Association for Computing Machinery*, 41(1), 67.
- [17] Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*(Vol. 1, pp. I-I). IEEE.
- [18] Tao, L., & Yanan, G. (2021). Research on privacy governance and evaluation of face recognition in the era of artificial intelligence. *Evaluation & Management*(019-004).