

Different Age Face Recognition Techniques based on LBP and SVM Algorithm

Meng Xu ^{a,*}, Mingjie Wu ^b and Xin Su ^c

Northwestern Polytechnical University, Xian, Shaanxi, 710072, China

^a 549922905@qq.com, ^b1098754879@qq.com, ^c 1612078218@qq.com

Keywords: face feature, support vector machine, principal component analysis, ROC

Abstract. In view of the different styles of pictures, we first on the selected picture library for noise reduction, filtering and wrinkle and other treatment, so that the picture library to meet the requirements of the title, while reducing the impact of age factors on identification, simplified model calculation. In order to further improve the recognition rate of the system, this paper improves the facial feature extraction and the design of the classifier. The LBP algorithm is used to extract the microscopic spatial structure of the face region, and the SDM algorithm is used to locate the face The CSLBP operator extracts the neighborhood characteristics of each feature point. As the image pixel density is too large, we use the main component analysis method (PCA) for dimension reduction processing. Finally, the data into SVM and other classifiers for algorithm learning, comprehensive comparison, select the best effect of the classifier as a final training model method.

1. Introduction

Face recognition can be widely used in security verification, e-commerce and other fields, is the computer vision and artificial intelligence research hot spots, the traditional face recognition is based on the identification of the same age, and these face recognition algorithms are using a single feature and Classifier, only one-sided description of the face, resulting in different ages, postures, lighting and other environments, the robustness of the algorithm is not strong. Aiming at this problem, we propose a face recognition algorithm based on LBP and CSLBP operator to extract facial features and use SVM and other eight kinds of machine learning methods to train. By observing the experimental results on FG-NET and Yale face database and comparing with other methods, our proposed LBP and SVM-based algorithms have obvious advantages in identifying the effects of different ages on the same person.

For the same person, if there is no change in face disease, facial trauma or surgical experience, young and old face always have a great similarity. People in life are often able to distinguish between two different age photos are not the same person. Of course, the greater the age difference, the more difficult to identify, because the facial contours and textures will change with age. But the facial features of the change will not be particularly large, there is always a great similarity. This paper aims to establish a reasonable mathematical model, through the algorithm automatically identify two different age photos are the same person.

2. Selection of Face Database

By searching the face database, we found that most of the existing database is under the constraints of the shooting according to different needs of the collection of photos, such as facial expressions, multi-gesture, different lighting conditions such as collection of object photos. For example, the ORL face database includes changes in gestures, facial expressions and facial ornaments; But limited to tracking the acquisition of difficult, long time span and other issues, these large face database are not collected objects of different ages of the photo, coupled with the Internet although there are different age objects of the photo but will involve the problem of portrait rights. For these reasons, the development of face-based recognition technology based on age has been limited.

In a further search, we found that as part of the EU Face and Attitude Recognition Program, the FG-NET age database collected 82 face images of different ages and detailed age information, available for scientific research. In addition, we also found the FERET face database, contains 200 different objects of different gestures and different age photo library, we can only keep the database of different age face database for later feature extraction and training model. In this paper, FG-NET and FERET part of the face image, as the basic database of the experiment..

3. Face Image Preprocessing

The preprocessing of face images is critical because the face objects that need to be identified in this paper are face features of different ages, so the image may contain noise of unreasonable levels, as well as wrinkles on the face, Equilibrium is what we need to consider. Therefore, pre-processing such as noise reduction, wrinkle and equalization can improve the quality of image and reduce the influence of age factor on image recognition accuracy before image recognition..

In the process of face image acquisition, the acquisition device and the light restriction will make the facial feature extraction noise, so we use the filter denoising method to make face feature extraction and face recognition more convenient. In this paper, we use the filter denoising method is Gaussian filtering. Gaussian filtering is used to eliminate Gaussian noise, and the weighted average gray value of the pixels in the field determined by the template is determined by the convolutional pixel of the government face image instead of the template center pixel. In our chosen face database, the same object's face rotation is different, which makes it difficult to identify, face rotation as shown in Figure 1.

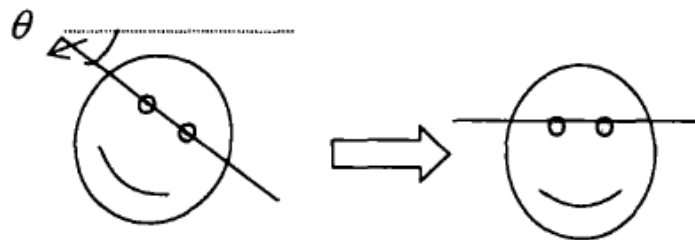


Figure 1. Rotation adjustment processing diagram

In this algorithm, we use the manual selection of the eye position and record the coordinates, find the angle between the two eyes and horizontal lines, and then follow the angle of rotation can be corrected after the picture. Figure 2 is an example of the correction before and after.



Figure 2. Face contrast before and after comparison

The use of photoshop software and download the Kodak filter plug can achieve anti-acne, wrinkle effect, to help remove the characteristics caused by age, facial recognition feature extraction easier, more accurate face recognition. In the image acquisition, the change of light makes the image unevenness, so it is not easy to get the characteristic in the same difficulty when extracting the feature. So enhance the face image contrast, making the gray distribution evenly, as shown in Figure 3, can be more convenient for face recognition.



Figure 3. Gray scale equalization comparison chart

4. Treatment with LBP and CSLBP

This paper deals with the standard position, the standard light shot, so we can choose to reflect the main facial differences in different eyebrows, eyes, nose, mouth area of the local details of the details and face color, contour of the overall properties. We use the SDM algorithm to locate the face feature points, and use the CSLBP operator to extract the features in the neighborhood of the face feature points. All the feature points' neighborhood features a fine face Feature; at the same time in the face area to extract the partition LBP histogram feature description of the human face space structure, reflecting the face of the texture features.

The dimension of the LBP feature is $2^8 = 256$ and the dimension of the CSLBP feature is $2^4 = 16$. It can be seen that the feature dimension of the CSLBP is significantly lower than the LBP feature dimension, which makes the CSLBP feature have lower computational complexity and higher feature extraction efficiency.

$$LBP_{P,R} = \sum_{i=0}^7 s(n_i - n_c) 2^i \quad (0-1)$$

$$CSLBP_{P,R} = \sum_{i=0}^3 s(n_i - n_{i+4}) 2^i \quad (0-2)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (0-3)$$

LBP histogram sequence belongs to the first-order statistical feature, can better describe the image texture characteristics, the use of operator scanning the entire face image, you can get the encoded image. As shown in Figure 4, the first four pictures of the FG-NET library pretreatment after two different ages of the photo, the middle of the four images of the LBP operator extracted from the face texture map. From Fig. 5, we can see that the face texture map extracted by LBP not only has good local characteristics, but also maintains the visual characteristics of the original image. More importantly, the extracted facial face information is obviously prominent, especially Mouth, eyes and nose area, and texture information is also more obvious. Thus, this algorithm has good reliability. According to this algorithm to generate LBP histogram (Figure 6), from the figure we can see that the same person difference is small, and different people two photos LBP histogram difference is large, which also shows that the algorithm has a good reliability Sex.



Figure 4. Examples of face images of the same person at different ages



Figure 5. The LBP algorithm processes the image

The feature points are able to locate the exact position of facial features and facial contours, which can extract the features that best serve the classification pattern from the original image and reduce noise interference. In this paper, the SDM algorithm is used to locate the feature points. The face and the feature point model are fitted in the detected face region, and the location of the 69 feature points is obtained. The positioning effect of the feature points is shown in Fig. 7. The image subgraph is taken in each feature point neighborhood (15×15 pixel rectangle with feature point as center). After preprocessing, the CSLBP feature of the sub-image is extracted, and finally 69 feature points Neighborhood characteristics are concatenated to form feature points Neighborhood CSLBP features. The feature extraction area of the face feature point is shown in Fig. 7.

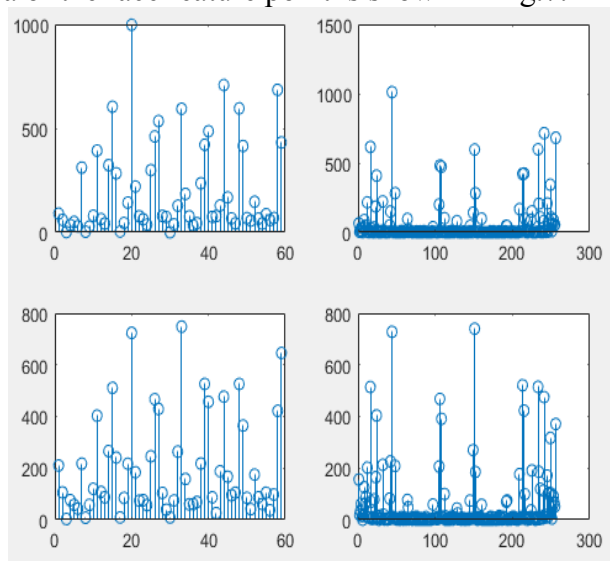


Figure 6. Same person with different age face histogram

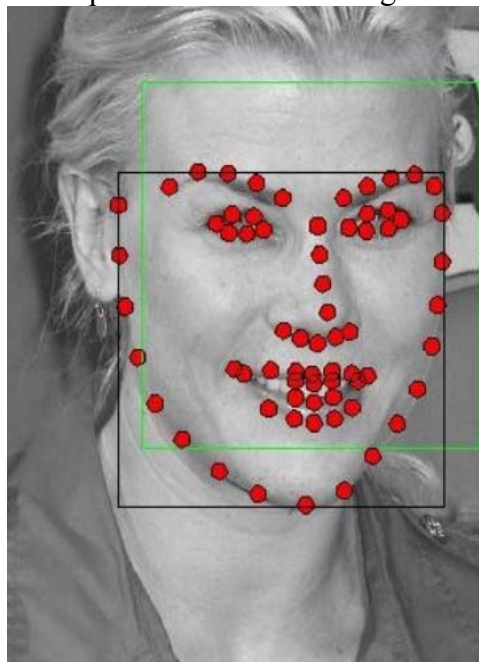


Figure 7. Feature point extraction

5. Model Establishment

A. Main analysis method Dimension reduction (PCA)

Principal component analysis (PCA) is the most commonly used linear dimensionality reduction method. Its goal is to map high-dimensional data to low-dimensional space representation by some kind of linear projection and expect data on the projected dimension Variance is the largest, in order to use less data dimension, while retaining more of the original data points of the characteristics.

It is popular to understand that if all the points are mapped together, almost all of the information (such as the distance between points and points) is lost, and if the mapping variance is as large as possible, then the data points are scattered Open to keep more information. It can be shown that PCA is the least linear way to lose the original data information. (In fact, is closest to the original data, but PCA does not attempt to explore the data inherent structure)

Let the n-dimensional vector w be the coordinate axis direction (called the mapping vector) of the target subspace, and maximize the variance after the data mapping.

$$\max_w \frac{1}{m-1} \sum_{i=1}^m (w^T (x_i - \bar{x}))^2 \quad (0-4)$$

Where m is the number of data instances, x_i is the vector representation of the data instance i , and \bar{x} is the average vector for all data instances. The definition of W is a matrix containing all the mapping vectors as column vectors. After linear algebraic transformation, the following objective functions can be obtained:

$$\min_w t_r(W^T A W), s.t. W^T W = I \quad (0-5)$$

$$A = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^T \quad (0-6)$$

Which t_r represents the rank of the matrix, and A is the data covariance matrix. It is easy to get the optimal W is composed of the eigenvector corresponding to the k largest eigenvalues of the data covariance matrix as the column vector. These feature vectors form a set of orthogonal bases and best retain the information in the data.

The PCA pursues the inherent information that maximizes the retention of data after dimensionality and measures the importance of the direction by measuring the size of the variance of the data in the projection direction. But this projection after the distinction between the roles of the data is not large, but may make the data points rubbing together cannot distinguish. This is also the biggest problem with the existence of PCA, which leads to the use of PCA in many cases the classification effect is not good. As shown in the figure below, when using PCA to project data points onto one-dimensional space, the PCA chooses the 2-axis, which makes it difficult to distinguish between the two points that were easily distinguishable. 1 axis will get a good distinction between the results.

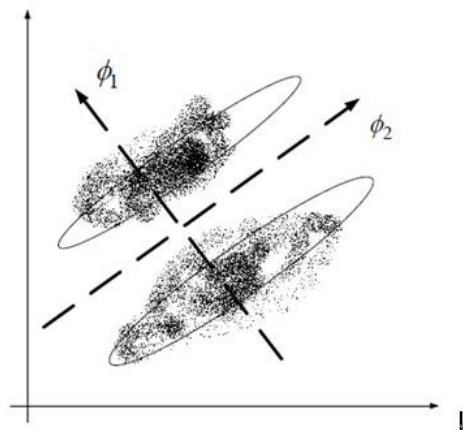


Figure 8. PCA schematic

B. Support vector machine method classification

Machine learning has achieved significant success in many areas of knowledge engineering and has been widely used in a variety of visual tasks. K-nearest neighbor (KNN), that is, given a training data set, for the new input instance, in the training data set to find the nearest instance of the k examples, the k examples of the majority of a Class, the input instance is classified into this class.

Decision Tree (Decision Tree) through the training data to build a decision tree, you can efficiently classify unknown data. There are two advantages to the decision-making number: 1) the decision tree model can be read well and descriptive, and it is helpful for manual analysis. 2) High efficiency, decision tree only needs to be constructed and used repeatedly. The maximum number of calculations for each forecast is not more than the depth of the decision tree.

Support vector machine SVM system has a strong general learning ability, to overcome the neural network and other methods of learning and lack of learning problems, there is a strong non-linear classification ability. If x is the data point and y is the category (y can take 1 or -1, representing two different classes), the learning objective of a linear classifier is to find a hyperplane in the n -dimensional data space (Hyper plane), and the hyperplane equation is:

$$w^T x + b = 0 \quad (0-7)$$

Logistic regression aims to learn a 0/1 classification model from the feature, and this model is a linear combination of the characteristics as an independent variable, because the range of independent variables is negative to infinity. Thus, the logarithm function (or sigmoid function) is used to map the argument to (0, 1), and the mapped value is considered to be the probability of $y = 1$.

$$h_{\theta}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \quad (1-8)$$

$$A = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^T \quad (1-9)$$

Where x is the n -dimensional eigenvector and the function g is the logistic function.

Define the Lagrangian function (by combining the Lagrangian function to the objective function, so that only one function expression can clearly express our problem):

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (w^T x_i + b) - 1) \quad (1-10)$$

$$\theta(w) = \max_{\alpha_i \geq 0} (w, b, \alpha) \quad (1-11)$$

When all constraints are met, the optimal value is $\theta(w) = \frac{1}{2} \|w\|^2$, that is, the amount $\frac{1}{2} \|w\|^2$ that is initially minimized.

Thus, minimizing the requirement that the constraint is $\theta(w)$ met, it is actually equivalent to direct minimization (the constraint is $\alpha_i \geq 0, i=1, \dots, n \geq 0$), because if the constraint is not met, it will be equal to infinity and will not be what we ask the minimum value.

$$\min_{w, b} \theta(w) = \min_{w, b} \max_{\alpha_i \geq 0} L(w, b, \alpha) = p^* \quad (1-12)$$

Here is the optimal value for the problem, and p^* is equivalent to the original problem. If the direct solution, then one will have to face the w and b two parameters, but α_i also inequality constraints, the solution process is not good to do. May wish to exchange the smallest and largest position, become:

$$\max_{\alpha_i \geq 0} \min_{w, b} L(w, b, \alpha) = d^* \quad (1-13)$$

The new problem after the exchange is the dual problem of the original problem, and the optimal value of the new problem is used to represent. And there are, in the case of certain conditions, the two are equal, this time by solving the dual problem to indirectly solve the original problem.

For the Gaussian kernel function, if the selection σ is large, the weight of the higher order feature actually decays very quickly, so in fact (the numerical approximation) is equivalent to a low-dimensional subspace; in turn, if the σ election is Small, you can map any data to linear separability - of course, this is not necessarily a good thing, because the attendant may be a very serious over-fitting problem. However, in general, by adjusting the parameters σ , the Gaussian kernel actually has a very high degree of flexibility, is also one of the most widely used kernel functions. The example

shown below is the low-dimensional linear inseparable data through the Gaussian kernel function mapped to the high-dimensional space:

Gaussian kernel function:

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (1-4)$$

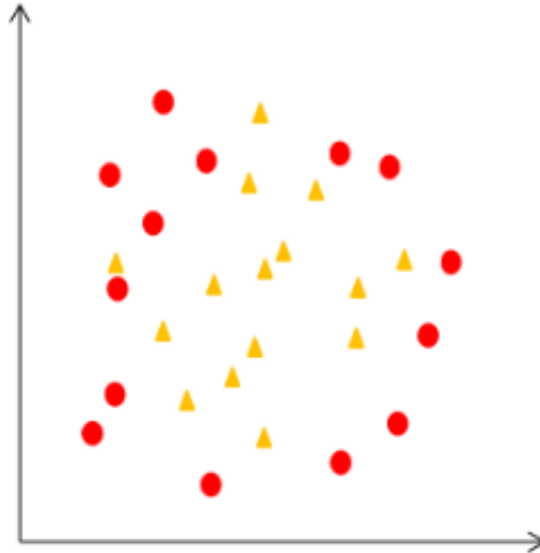


Figure 9. Input vector SVM Gaussian kernel classification strategy

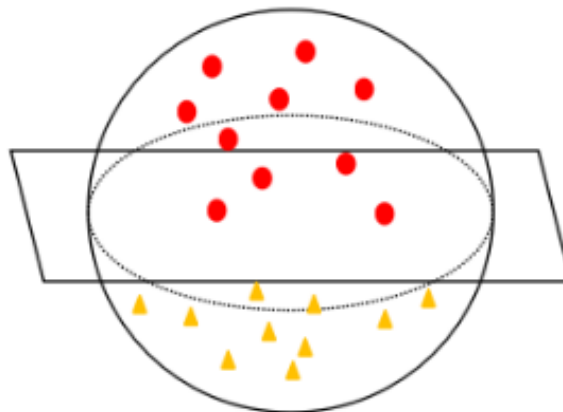


Figure 10. SVM classifier

In view of the different training methods of the machine learning methods have to be the same, we selected the K neighbor, support vector machine, decision tree, Logistic regression, maximum entropy model five machine learning methods for comprehensive analysis. As a result of the learning model of the classifier, we can see that the support vector machine method has a good classification effect, so we choose the support vector machine as the final classifier of our method.

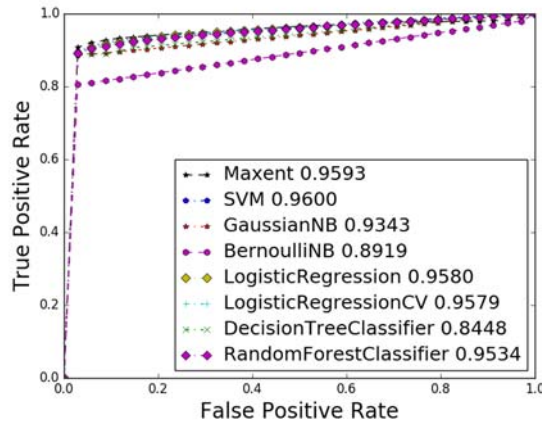


Figure 11. Effect Analysis of Various Machine Learning Methods

6. Data Processing

We obtained from the network FG-NET, Yale two data sets, each data set for a separate analysis, training set and test set are randomly selected, the following table is the data set related parameters.

Table 1. Data set selection and parameters

Data set	Number of samples	Data set size	Training set size	Test set size
FG-NET	82	1500	1000	500
Yale	64	1200	600	600

In order to further verify the effectiveness of the proposed algorithm for face recognition under complex conditions, the algorithm is compared with the existing common algorithms such as PCA, LBP, LTP and HOG. Because in practice, the conditions of the training set are usually ideal, so in the Yale face database using the subset 1 for the training set, the other subset of the test set for the experiment. In the experiment, the local binary model LBP adopts the radius window of 2 and extracts the characteristics of 8 sampling points. The local three-valued mode LTP reference LBP feature selection is used. The AW-HOG algorithm adopts 3 × 3 block method. In order to avoid the influence of other factors, the experiment does not do any light pretreatment of the face image, the classifier unified use SVM classifier.

Because the data set is divided by random partition data set method, it is random from the data set to extract part of the data to do the training set, with the rest of the data as a test set to test, so this experiment does not exist random, to confirm the stability of the method.

7. Conclusion

Through the analysis of the results, we can find that this method has a good recognition effect. LBP algorithm feature extraction and PCA algorithm combined with reduced dimension, interlocking, as far as possible to reduce the loss of information on the picture, and this article broke through a single machine learning and training methods, using the sklearn method library to use all machine learning methods, select the classification effect The best way to learn, greatly improve the recognition effect, compared with the literature 7, 8 in the method has a new upgrade.

Observing the results we can find that this method is not perfect to identify the efficiency of the analysis that the training database is not large enough, leading to the model is not perfect, then consider the purchase of some large library library for deep processing.

Table 2. Test results

Data set	recognition methods	Correct rate/%	Accuracy rate/%	Recall rate/%	F1/%
FG-NET	The method of this article	94.2	93.8	95.1	94.5
	Document 5 method	93.4	94.2	93.6	93.9
	Document 6 method	90.3	92.1	93.0	92.6
Yale	The method of this article	96.1	94.5	94.8	94.6
	Document 5 method	89.1	92.4	91.2	91.8
	Document 6 method	90.5	93.1	92.4	91.7

References

- [1]. Neeti Neeti, J. Ronald Eastman. Novel approaches in Extended Principal Component Analysis to compare spatio-temporal patterns among multiple image time series[J]. *Remote Sensing of Environment: An Interdisciplinary Journal*, 2014, 148:84-96.
- [2]. Niu, L., Qian, M., Song, R. et al. A Texture Matching Method Considering Geometric Transformations in Noninvasive Ultrasonic Measurement of Arterial Elasticity[J]. *Ultrasound in Medicine and Biology*, 2012, 38(3):524-533.
- [3]. Akhilesh Prasad, Santanu Manna, Ashutosh Mahato et al. The generalized continuous wavelet transform associated with the fractional Fourier transform[J]. *Journal of Computational and Applied Mathematics*, 2014, 259(Pt.B):660-670.
- [4]. Ergin Kilic, Melik Dolen. Prediction of slip in cable-drum systems using structured neural networks[J]. *Proceedings of the Institution of Mechanical Engineers, Part C. Journal of mechanical engineering science*, 2014, 228(3):441-456.
- [5]. Khan, N.M., Ksantini, R., Ahmad, I.S. et al. A novel SVM+NDA model for classification with an application to face recognition[J]. *Pattern Recognition: The Journal of the Pattern Recognition Society*, 2012, 45
- [6]. Tsukasa Hokimoto, Kunio Shimizu. A non-homogeneous hidden Markov model for predicting the distribution of sea surface elevation[J]. *Journal of applied statistics*, 2014, 41(1/2):294-319.