Face reconstruction based on PCA algorithm

*sunbao 1023040918

Abstract—**In the process of face recognition, the picture is generally regarded as a vector for processing, advanced mathematics we contact are generally two-dimensional or three-dimensional vectors, the dimension of the vector is determined according to the number of variables that constitute the vector. In the process of abstracting an image into a vector, we set each pixel of the image as one dimension. For an ordinary image of 92*112, it is finally abstracted into a high-dimensional vector of 10304 dimension. Such a huge dimension is quite difficult for the subsequent image calculation formula. Therefore, it is necessary to reduce the image dimension without losing important information as much as possible. PCA is a method to reduce the image dimension. After PCA transformation, the image can retain any number of dimensional components that contribute more to the image features, that is, you can choose to reduce the dimension to 30 or 90 or other dimensions. Of course, the more dimensions are retained at the end, the less information is lost in the image, but the more complex the calculation. This experiment selected 40 groups of photos, each group of photos 10 faces, a total of 400 photos. In this paper, 80 % of the 400 photos are taken as the training set, and finally, photos are selected from the test set to predict the faces in the test set, which can get a high accuracy.**

Index Terms—**face recognition,dimension,PCA**

I. Introduction

The face recognition is widely used in visual surveillance of various departments in China, which is mainly used in the security system, identification of criminal field, proof identification and other important situations [1].Face recognition is one of the most popular problems in the field of image analysis and understanding. Identifying a person from an unknown face is usually done by comparing the unknown face with the known faces from a face database. The interest of researchers and engineers in face recognition problem has grown rapidly in the recent years since there is a wide range of commercial and law enforcement applications on face recognition [2]. The increasing need for surveillance-related applications, especially due to drug traffic and terrorist activities, has a great impact on the growth of interest in the field of face recognition. Some of the application areas of face recognition includes personnel identification of credit cards, driver's licence, passport checks, entrance control, computer access control, criminal investigations, crowd surveillance, witness face reconstruction, and ATM machines. The people interested in face recognition problem are not only the engineers who work in the area of machine learning, but also psychophysicists and neuroscientists are studying on this problem to understand human recognition mechanisms. The studies and findings of psychophysicists and neuroscientists [3] help the engineers who are designing and/or implementing algorithms or systems for machine recognition of faces. Principal Component Analysis (PCA) is one of the most popular appearance-based methods used mainly for dimensionality reduction in compression and recognition problems. PCA is known as Eigenspace Projection which is based on linearly projecting the image space to a low dimensional feature space that is called eigenspace. It tries to find the eigenvectors of the covariance matrix that correspond to the directions of the principal components of the original data. Another powerful dimensionality reduction technique is the Linear Discriminant Analysis (LDA) which is also known as Fisher's Discriminant Analysis. LDA searches for a linear transformation such that the feature clusters are most separable after the transformation which can be achieved through scatter matrix analysis. Linear Discriminant Analysis deals directly with discrimination between classes, whereas PCA deals with the data in its entirety for the principal components analysis without paying any particular attention to the underlying class structure. The third appearance-based statistical method is Independent Component Analysis (ICA) which is a special case of redundancy reduction technique and it represents the data in terms of statistically independent variables. ICA is a method for transforming an observed multidimensional random vector into components that are statistically as independent from each other as possible. In this paper, we mainly analyze the application of PCA algorithm in face reconstruction.

II. RELATED WORK

A. Introduction to PCA

PCA, or Principal component analysis, is one of the most widely used data dimensionality reduction algorithms. The main idea of PCA is to map N-dimensional features to K-dimensional features, which are new orthogonal features, also known as principal components, and reconstructed K-dimensional features on the basis of the original N-dimensional features. The work of PCA is to find a set of mutually orthogonal coordinate axes sequentially from the original space, and the choice of new coordinate axes is closely related to the data itself. Among them, the first new axis is selected as the direction with the largest difference in the original data, the second new axis is selected as the plane with the largest variance in the plane orthogonal to the first axis, and the third axis is the plane with the largest difference in the plane orthogonal to the first and second axes. And so on, you get n of these axes. With the new axes obtained in this way, we find that most of the variance is contained in the first k axes, and the variance in the latter axes is almost zero. Thus, we can ignore the rest of the axes and keep only the first k axes that contain most of the variance. In fact, this is equivalent to retaining only the dimensional features that contain most of the variance, while ignoring the feature dimensions that contain almost zero variance, so as to achieve dimensionality reduction processing of data features. Here is a simple example, first to consider a special case, there are four points in a coordinate axis, then their data is stored in an array, is a two-dimensional array, x and y together a total of 8 data. In order to reduce the dimension of the data, PCA is to find a new coordinate system, whose origin is the center of the data, so that the y coordinates of the four data fall on the x axis as far as possible, so that only the one-dimensional array can save the original two-dimensional data. Therefore, PCA is to find a coordinate system so that the loss of information is minimal when the data is kept in only one dimension.

B. PCA algorithm

PCA is a dimensionality reduction technique which is used for compression and recognition problems. It is also known as Eigenspace Projection or Karhunen-Loeve Transformation [4] [5] [6]. PCA projects images into a subspace such that the first orthogonal dimension of this subspace captures the greatest amount of variance among the images and the last dimension of this subspace captures the least amount of variance among the images [7]. The main goal of PCA is the dimensionality reduction, therefore the eigenvectors of the covariance matrix should be found in order to reach the solution. The eigenvectors correspond to the directions of the principal components of the original data, their statistical significance is given by their corresponding eigenvalues [6]. The steps in finding the principal components can be summarized as follows [8]:

- Collect xi of an n dimensional data set x, i=1,2, ..., m
- Mean correct (center) all the points: Calculate mean mx and subtract it from each data point, xi – mx
- Calculate the covarance matrix: C = (xi - mx)(xi - mx)T
- Determine eigenvalues and eigenvectors of the matrix C.
- Sort the eigenvalues (and corresponding eigenvectors) in decreasing order.
- Select the first $d \leq n$ eigenvectors and generate the data set in the new representation.
- The projected test image is compared to every projected training image by using a similarity measure. The result is the training image which is the closest to the test image.

III. EVALUATION

A. Data analysis

Instead of using the data set in sklearn library given by the teacher, this paper finds a data set from the Internet, which contains 40 groups of photos, each group of photos consists of 10 faces. These 40 groups of photos are saved in the attfaces folder. In the experiment, faces array is used to store 400 photos. Use the cv2 library read to read the photos in the faces array and generate a label array where the labels in each set of photos are the same. Then draw the photos of the last two sets of data, as shown in Fig. 1.

Fig. 1. The last two sets of data faces

5 rows × 10304 columns pd.DataFrame

	0	1	2	3	4	5	6	7	8	9	...	10294	10295	10296	10297	10298	10299	10300	10301	1030
0	48	49	45	47	49	57	39	42	53	49	...	39	44	40	41	49	42	44	47	
1	34	34	33	32	38	40	39	49	54	57	...	42	44	38	30	37	30	36	37	
2	60	60	62	53	48	51	61	60	71	68	...	27	35	28	33	31	31	37	32	
3	39	44	53	37	61	48	61	45	35	40	...	23	30	36	32	28	32	31	29	
4	63	53	35	36	33	34	31	35	39	43	...	173	169	166	161	158	169	137	41	

Fig. 2. Features of the first 5 pictures

B. Viewing data dimensions

Then the image array is converted to numpy array, since the image is 92*112, that is, 10304 dimensional, a total of 400 photos, the numpy array size is 400*10304. Then import pandas library and use DataFrame function to draw the feature matrix, as shown in Fig. 2, showing the features of the first five images.

C. Generated feature face

Next, the data set is divided. In this paper, 80 % of the data is selected as the training set, 20 % of the data is used as the test set, and the data of 10304 dimension is reduced to 100 dimension. The size of the training set after dimensionality reduction is 320*100, and the size of the test set is 80*100. Finally, the feature faces are drawn according to the data in the training set, as shown in Fig. 3.

D. Dimension selection

After that, we test the most appropriate dimension for face reconstruction. First, we use the attribute explained variance ratio to check the percentage of information accounted by each new feature vector in the total information of the original data after dimensionality reduction, as shown in Fig. 4.It can be seen that the more we think about it, the smaller the value, their sum is 0.9023, not equal to 1, because some data is lost in the process of dimensionality reduction. Then draw a graph of the number of features and the number of information carried, as shown in the figure Fig. 5. From the figure below, we can see that when the dimension is reduced to 140, the reconstructed image retains the most information, but it will also bring huge computational burden, so this paper chooses to reduce the dimension to 100.

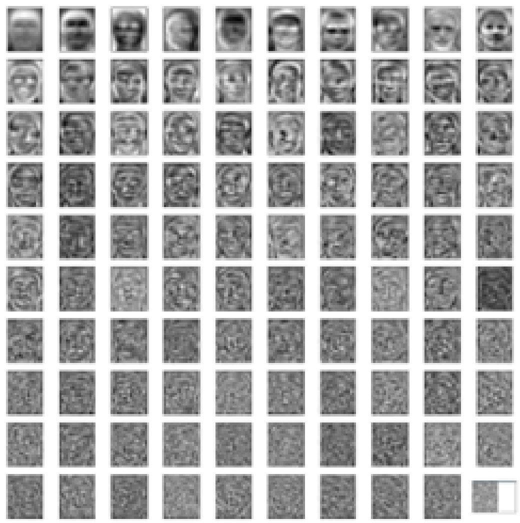

Fig. 3. Eigen face

```
array([0.16754406, 0.11712118, 0.08050592, 0.05800583, 0.04899411,
       0.03236304, 0.02552568, 0.02246334, 0.02105942, 0.01869678,
       0.01492577, 0.01452819, 0.01195689, 0.01106418, 0.01061136,
       0.00920361, 0.00893044, 0.00841665, 0.00815548, 0.00745415,
       0.00684847, 0.00674609, 0.00641437, 0.00555017, 0.00533678,
       0.00511044, 0.00498169, 0.00493545, 0.00477643, 0.0046901 ,
       0.00452947, 0.00443995, 0.00424948, 0.00415627, 0.00402244,
       0.00391703, 0.00380438, 0.00365518, 0.00347555, 0.00338822,
       0.00325   , 0.00306806, 0.00305956, 0.00297671, 0.00286721,
       0.00281228, 0.00272433, 0.00266031, 0.00257338, 0.00251557,
       0.00247235, 0.00243605, 0.00236254, 0.00232992, 0.00225821,
```

Fig. 4. information accounted

E. Accuracy analysis

After that, we use OpenCV's EigenFace algorithm for recognition. His principle is to project both the training set image and the test set image into the feature vector space, and then use clustering methods (nearest neighbor or K-nearest neighbor, etc.) to get the nearest image of each image in the test set and classify it. The first to use cv2 Face EigenFaceRecognizercreate () to create model of face recognition by image array model and the corresponding tag array to training. The Predict function will return a label for the test data, and after viewing it, we can see a picture of the successful prediction. This paper tests the accuracy of the data set, and the final test set recognition accuracy is 93 %. Finally, we generated the average face of these 400 data as shown in the Fig. 6.

CONCLUSION

In this paper, PCA algorithm is used for face reconstruction, and PCA is used to change high-dimensional data into low-dimensional data. Principal component analysis (PCA) is a dimensionality reduction technique for compression and

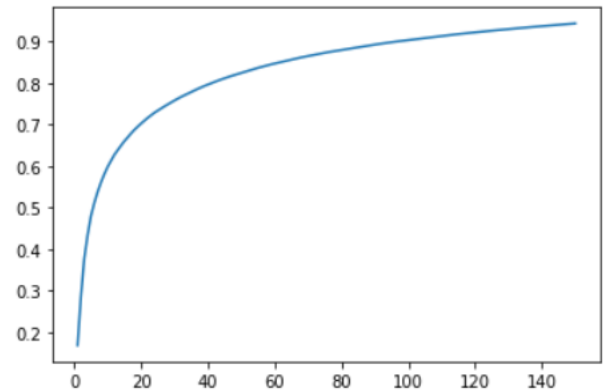

Fig. 5. Dimension and retention of information

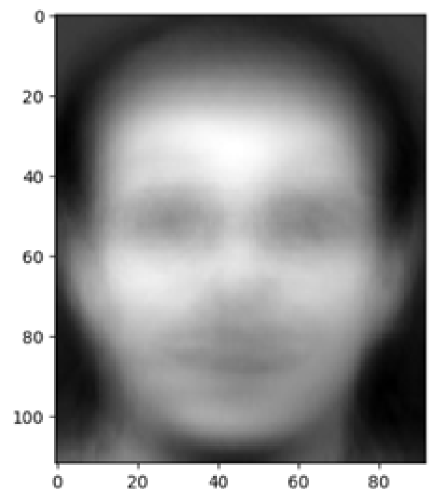

Fig. 6. Average face

identification problems. It is also known as the feature space projection or Karhunen-Loeve transformation. PCA projects images into a subspace such that the first orthogonal dimension of that subspace captures the greatest variance between the images, while the last dimension of that subspace captures the smallest variance between the images. There are still many shortcomings in this experiment, no reconstructed photo is given to find his previous photo, and this paper has not been compared with other face reconstruction algorithms, and there is still a lot of work to be done.

REFERENCES

[1] Lei Songze. 2006. MATLAB Realization of Face Feature Extraction Based on PCA. Computer Development and Application.19 (11): 20-21.

[2] Zhao W., Chellappa R., Rosenfeld A., and Phillips P. J.. "Face recognition: A literature survey" Technical Report CAR-TR-948, CS-TR-4167, N00014-95-1-0521, October 2000.

[3] Chellappa R., Sirohey S., Wilson C.L., Barnes C.S., "Human and Machine Recognition of Faces: A Survey", in Proceedings of IEEE, May 95, Vol. 83, pp 705-740.

[4] Turk M. and Pentland A., "Face recognition using eigenfaces", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, 1991.
[5] Belhumeur N., Hespanha J. and Kriegman D., "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", Proc. ECCV, 45–58, 1996.
[6] Calvo R., Partridge M., and Jabri M., "A Comparative Study of Principal Components Analysis Techniques", In Proc. Ninth Australian Conf. on Neural Networks, Brisbane, QLD., pp. 276-281, 1998.
[7] Yambor W.S., "Analysis of PCA-Based and Fisher Discriminant-Based Image Recognition Algorithms", Technical Report CS-00-103, Computer Science Department, Colorado State University, July 2000.
[8] Önsen TOYGAR,Adnan ACAN.FACE RECOGNITION USING PCA, LDA AND ICA APPROACHES ON COLORED IMAGES[J].ELECTRICA,2003,3(1):735-743.