Proceeding of the 2nd Informatics Conference 2016 (ICF-2016) June 18, 2016, at Universitas Budi Luhur, Jakarta, Indonesia, Editor: Setyawan Widyarto, ISSN: 2477-5894

Face Recognition with Principal Component Analysis (PCA) Application Using Euclidean Distance Measurement

Charles Victor B. Saragih, Universitas Budi Luhur,
Setyawan Widyarto, Universiti Selangor

Abstract— Face recognition algorithms can be categorized into 5 based methods of linear and non-linear projection, namely: artificial neural network-based method of non-linear, Gabor filters and wavelets based methods, fractal-based methods and methods based on thermal and hyperspectral. PCA is a statistical method that can explain the formulation of artificial neural networks and is designed to process the multidimensional information. With the PCA method to do efficient calculation where multidimensional information can be simplified into a number of variables, dimensions and factors serve as the basic component. Many researchers use PCA method that allows modeling of a human face by using the parameters in limited quantities. One advantage PCA method is the ability to process high-dimensional data modeling that cannot be done by many other methods because it requires a covariant matrix inverse. PCA is a better method than matching pursuit (MP), especially on the use of time, fast and efficient. The purpose of this study was to analyze the images using image recognition algorithms to calculate the distance euclidean PCA. The result of this research is the image recognition can be performed using PCA algorithm to form a basis vector as the basis for calculating the normalization of an image. Euclidean distance calculations will provide clarity regarding the degree of similarity and dissimilarity drawing a picture

Keyword : Principal Component Analysis, Euclidean distance

I. INTRODUCTION

WITH the growing need for non - invasive system the introduction of facial recognition into an area that is very popular among researchers. So many variations of algorithms that can be used for face recognition process is researched and there are several methods of evaluation are used to test the algorithm. However, existing research is still needed to be improved so that it can be implemented practically to solve real problems in life. Face recognition algorithms can be categorized into 5 based methods of linear and non - linear projection , namely : artificial neural network- based method of non - linear, Gabor filters and wavelets based methods , fractal - based methods and methods based on thermal and hyperspectral[1]. Facial pattern recognition has been a concern for two decades. However, there is still a problem unsolved that require further research . Some of the challenges faced by engineering castaways face is among other things a variation of the position , facial expression , age and lack of lighting . In a face pattern recognition with such a large -scale identification of Driver Licenses (SIM), the data training generally consists of only one image of each person . This situation poses a problem because the data are limited testing . Many face recognition techniques assume that there are a lot of image data on the personal and the technique will have difficulty identifying if there is only one sample data only[2]. PCA is a statistical method that can explain the formulation of artificial neural networks and is designed to process the multidimensional information. With the PCA method to do efficient calculation where multidimensional information can be simplified into a number of variables, dimensions and factors that serve as the basic components [5] . By using permutation testing , PCA can form a matrix that sets the value that have relevance to the basic components [6] . Many researchers using PCA method that allows modeling of a human face by using parameters with a limited number [7]. One advantage PCA method is the ability to process high-dimensional data modeling that cannot be done by many other methods because it requires the matrix inverse covariant [8]. PCA is a better method than matching pursuit (MP) , especially on the use of time quickly and efficiently [9]. The purpose of this paper is to analyze the images using image recognition algorithms to calculate the PCA euclidean distance between the training data and experimental data and review the recognition of an image with a limited testing of 15 experimental data.

A. Principal Component Analysis

The most popular method in face recognition techniques are principal component analysis (PCA), fisher linear discriminant analysis (LDA) and locality preserving projections (LPP)[2]. One algorithm commonly used in facial pattern recognition is the Principal Component Analysis (PCA). PCA using a basis vector corresponding to the maximum variance direction of the image data that is supported by a covariance matrix calculation. Covariance matrix will generate a vector basis Eigenface as the projection matrix. Normalization basis vectors is done before the vector used in the next step . Stages of training at the PCA can be seen in Figure 1 [1].

Proceeding of the 2nd Informatics Conference 2016 (ICF-2016) June 18, 2016, at Universitas Budi Luhur, Jakarta, Indonesia, Editor: Setyawan Widyarto, ISSN: 2477-5894

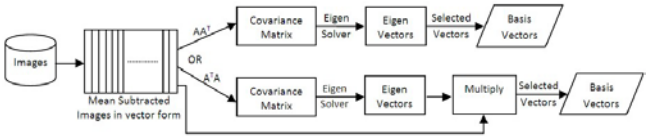


Figure 1. Phase training at PCA

At the stage of projection, the image that has been in the form of two dimensions simultaneously multiplied by the projection matrix that has been previously established so as to form a matrix vector with a lower dimension. There are three databases that can be used for the comparison is FERET, YALE and ORL. FERET databases are generally used for testing FERET evaluation, testing facial pattern (FRVT) and many are also used by researchers with different algorithms. ORL database is one of the oldest most popular database. The reason why many researchers use these databases is therefore provided a fairly complete picture, there are 10 different images for each of 40 subjects so total there are 400 pictures. YALE database consists of 165 images of 15 subjects consisting of 11 variations. Variation consists of three lighting conditions and facial expressions 6 variations of the object. There are 4 measurement matrix within which the calculation of the Euclidean distance, Euclidean Cosine, Mahalanobis and Mahalanobis Cosine [1].

B. Euclidean distance

Euclidean distance is the calculation of the distance of two dots in Euclidean space (Euclidean space). Euclidean space to study the relationship between angles and distances was introduced by Euclid, a mathematician from Greece around 300 B.C. This relates to the Euclidean theorem of Pythagoras and is usually applied to the dimensions of 1, 2 and 3 and it also simple when applied to a higher dimension. In the calculation of one-dimensional exemplified first point is 4 and the second point is -10. How to calculate by subtracting the distance is -10 to 4, resulting in a negative 14. The absolute value of the value of -14 can be obtained with the power two of the number that gets the value 196, then rooted so get a value of 14. So the euclidean distance of 2 points is 14.

The 2-dimensional calculation is almost the same way. For example, the first point has the coordinate (1, 2). The second point is on the coordinate (5, 5). To calculate the distance, it is subtracting the coordinates of the second point with the first point. Namely, (5-1,5-2) in order to obtain the coordinates (4,3). Then rank each so as to obtain (16,9). Then add all of them so as to obtain a value of 16 + 9 = 25. These results are then rooted to 5. So its euclidean distance is 5 as in Figure 2.

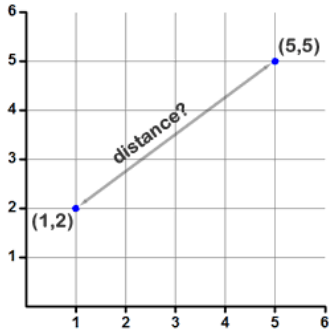


Figure 2. Euclidean distance

The calculation of the Euclidean distance can use the formula as in Figure 3.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Figure 3. Euclidean equation

C. Implementation

In conducting the study used a database of YALE by the number of samples is 15. Pictures were taken at random from several variations of lighting and facial expressions. The images used have a grayscale color patterns with a file size 320x243. Examples of imagery used can be seen in Figure 4 [3].

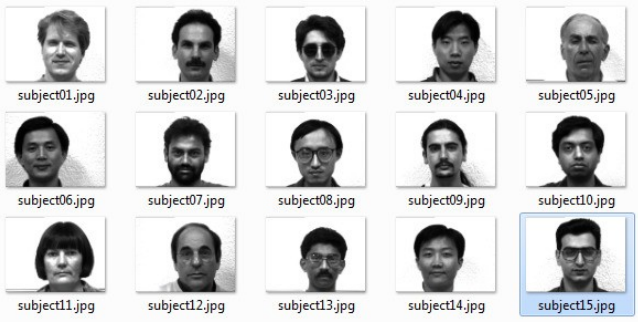


Figure 4. YALE database

In the present study used the PCA method with the following steps:

- Making the matrix to accommodate 15 pictures of YALE database.
- Do reshape against each matrix, forming a row vector 15 vectors.
- Do the calculation of average or mean of all vector lines.
- Do reduction of 15 vector line with the average value has been calculated in the previous step.
- Calculates covariance matrix of all the row vector whose value has been reduced by a mean value.
- Calculating the Eigenvalue and eigenvector in each row vector called Eigenface .

Proceeding of the 2nd Informatics Conference 2016 (ICF-2016) June 18, 2016, at Universitas Budi Luhur, Jakarta, Indonesia, Editor: Setyawan Widyarto, ISSN: 2477-5894

- The final result is obtained by multiplying the basis vectors vector line with Eigenface .
- For phase matching the images to be compared to the calculation such as testing image obtained final result vector row that is multiplied by Eigenface (normalization) .
- Perform Euclidean distance calculations to determine the outcome.

Subsequent calculations are implementations using MatLab to calculate the PCA to the 15 sample testing images that were discussed in the previous section . The algorithm in matlab can be shown in Figure 5.

```
% --- Executes during object creation, after setting all properties.
function citraasli_CreateFcn(hObject, eventdata, handles)
% hObject    handle to citraasli (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: place code in OpeningFcn to populate citraasli

[b1,k1]=size(c1);
c1r=reshape(c1,1,b1*k1);
c_all=[c1r c2r c3r c4r c5r c6r c7r c8r c9r c10r c11r c12r c13r c14ar c15ar];
c_mean=mean(c_all);
c1n=c1-c_mean;
cov1=cov(c1n);
[V1,D1] = eig(cov1);
D1=diag(D1);
m1=V1(:,find(D1==max(D1)));
f1=m1'*c1n';
```

Figure 5. PCA algorithm

The results of calculation of the mean of 15 basis vectors is 0.723. Table 1. is the result of the row vector of the picture number 1 that has been multiplied by Eigenface or also known as a vector normalization as a comparison image testing.

1.6522	2.9179	2.9179	2.9179	2.9179	2.9179	2.9179	2.9179	2.8979	2.8731	2.8121
2.7808	2.681	2.6345	2.4378	2.347	2.2008	2.1873	2.0832	2.0578	2.0341	2.005
1.9296	1.9031	1.7931	1.7566	1.6578	1.6042	1.4247	1.3936	1.2699	1.2088	0.9569
0.8228	0.2974	0.1355	-0.2119	-0.3502	-0.8229	-1.0523	-1.5022	-1.622	-1.6634	-1.7085
-2.0292	-2.1341	-2.3468	-2.4635	-2.8047	-2.887	-3.0886	-3.1235	-3.1885	-3.2643	-3.5096
-3.5429	-3.5	-3.4926	-3.6028	-3.6275	-3.5391	-3.5184	-3.5593	-3.5831	-3.5674	-3.5654
-3.673	-3.7035	-3.7623	-3.7843	-3.818	-3.8499	-3.9221	-3.9309	-3.962	-3.9611	-3.933
-3.924	-4.0201	-4.0485	-4.1957	-4.2171	-4.3455	-4.3962	-4.4453	-4.4603	-4.4811	-4.4819
-4.4493	-4.4432	-4.3316	-4.32	-4.3882	-4.3986	-4.434	-4.4658	-4.4883	-4.5104	-4.5244
-4.5227	-4.5066	-4.4991	-4.4647	-4.4618	-4.4628	-4.4741	-4.4921	-4.4815	-4.5204	-4.5592
-4.59	-4.5696	-4.591	-4.5907	-4.5245	-4.4953	-4.3397	-4.2905	-4.1911	-4.1507	-3.9737
-3.9302	-3.6746	-3.6397	-3.3504	-3.2636	-3.0418	-2.9629	-2.6716	-2.5379	-2.0754	-1.9616
-1.6442	-1.5616	-1.2102	-1.1015	-0.8358	-0.7973	-0.5774	-0.5106	-0.2976	-0.2413	-0.1304
-0.1395	-0.1835	-0.2223	-0.2517	-0.279	-0.3773	-0.3664	-0.4407	-0.4889	-0.5066	-0.5535
-0.5379	-0.5705	-0.5439	-0.5326	-0.5341	-0.525	-0.4637	-0.4399	-0.3396	-0.3033	-0.0957
-0.0161	0.1031	0.1159	0.0522	0.0425	0.037	0.0543	0.1251	0.1337	0.1312	0.1333
0.1969	0.1861	0.1108	0.0645	0.3095	0.4136	0.6464	0.7262	0.8251	0.8618	1.0636
1.0724	1.1161	1.1111	1.0718	1.0749	1.1085	1.0902	1.1131	1.1549	1.261	1.2406
1.3278	1.3914	1.5363	1.5924	1.6605	1.7232	1.911	1.8962	1.9548	1.9478	1.9493
1.9504	1.8371	1.7629	1.7148	1.7239	1.6274	1.5574	1.4971	1.5176	1.4583	1.4082
0.903	0.8071	0.6192	0.5653	0.4385	0.4921	0.6637	0.6606	0.6681	0.6484	0.6403
0.5961	0.4929	0.4766	0.2729	0.2244	-0.0938	-0.2056	-0.5036	-0.502	-0.5866	-0.7127
-4.806										

Tabel 1. Row vector for picture number 1

Euclidean distance calculation algorithm in MatLab can be shown as in Figure 6.

```
function Euclidean = vektor(citra_s, citra_t)
global Pointer_rec

clc;
Pointer_rec = Pointer_rec+1;

A = citra_s;
B = citra_t;

Eu=0;

MA=mean2(A);
MB=mean2(B);

EA=entropy(A);
EB=entropy(B);

SA=std2(A);
SB=std2(B);
Eu=sqrt((MA-MB)^2+(EA-EB)^2+(SA-SB)^2);
Euclidean = Eu;
```

Figure 6. Euclidean algorithm

By using MatLab GUIDE display design is shown in Figure 7. In the design of the display there are 15 boxes that can be used to accommodate 15 images according to user needs. Once the images are visible on the display, the user can make the process of normalizing the vector row by pressing Mean located in the middle menu .

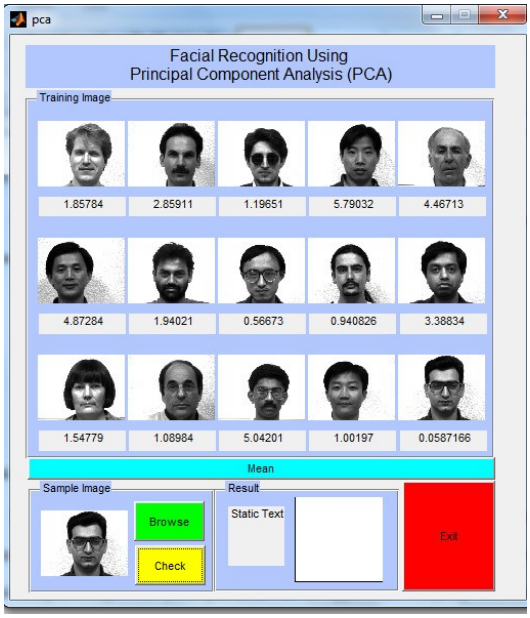


Figure 7. Display design

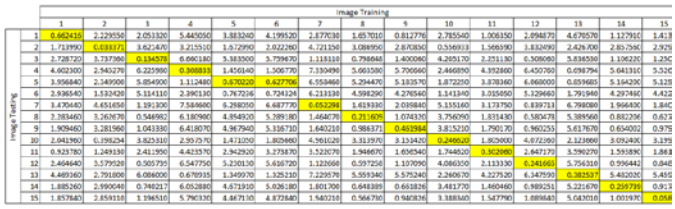
In order to do a comparison between the testing image and the training image, it can make the selection of files to be compared so that images can be shown in the box on the bottom left . After that to determine the Euclidean distance can be selected menu "Check" so that in each image seen figures showing the Euclidean distance between the testing image with training images.

D. Research result

Sampling was done by random that way provide an opportunity or equal chance of each element in the population to be selected as a sample. Measurements in this study using accidental sampling , where sampling can be done by chance

Proceeding of the 2nd Informatics Conference 2016 (ICF-2016) June 18, 2016, at Universitas Budi Luhur, Jakarta, Indonesia, Editor: Setyawan Widyarto, ISSN: 2477-5894

or represent all the information needed[4]. The result of this research is the conclusion of the testing that has been done. The test results based on the specifications of the application indicates that the application is successful for the calculation Eigenface and normalizing the vector of each image.



Tabel 2. Result

Factors tested in calculating the Euclidean distance factors include the average , standard deviation and entropy. The test results can be seen in Table 2. From the results of these tests can be drawn some conclusions :

- The average value of the fit between training images and test images is 69 %.
- Highest suitability value is 97% contained in the image number 2.
- The highest discrepancy was found in the picture number 5 where only scored 28%. It becomes a record because the value of the image number 5 in accordance with the picture number 6 even though from a visual, not seen the similarity between image number 5 and image number 6.
- When do facial recognition with images that are not registered in the training image recognition errors will occur at 74 %

II. Conclusion and Suggestion

Image recognition can be performed using PCA algorithm to form a basis vector as the basis for calculating the normalization of an image. Euclidean distance calculations will provide clarity regarding the degree of similarity and dissimilarity drawing a picture. The number of images of testing as many as 15 images no possibility of error image recognition. This can be seen in the picture number 6 which is similar to the picture number 5 instead of the actual image.

Future studies could use the algorithm LCC and LDA as a comparison result of the PCA as well as using other methods of calculation as the Mahalanobis distance on the database FERET or ORL.

References

[1] Bajwa UI, Taj IA, Anwar MW, Wang X (2013) A Multifaceted Independent Performance Analysis of Facial Subspace Recognition Algorithms. PLoS ONE 8(2): e56510. doi:10.1371/journal.pone.0056510

[2] Li, Qin, Hua Jing Wang, Jane You, Zhao Ming Li, Jin Xue Li. (2013). Enlarge the Training Set Based on Inter-ClassRelationship for Face Recognition from One Image per Person. Plos One.

[3] http://vision.ucsd.edu/content/yale-face-database.

[4] Diposumarto, Ngadino Surip. (2012). Metodologi Penelitian Teori dan Terapani. Mitra Wacana Media.

[5] Rakover, Sam S (2002), Featural vs. configurational information in faces: A conceptual and empirical analysis. British Journal of Psychology, ProQuest.

[6] Wagner, Florian (2015). GO-PCA: An Unsupervised Method to Explore Gene Expression Data Using Prior Knowledge. Plos One.

[7] Galantucci, Luigi, Maria, Eliana Di Gioia, Fulvio Lavecchia, Gianluca Percoco (2014). Is principal component analysis an effective tool to predict face attractiveness? A contribution based on real 3D faces of highly selected attractive women, scanned with stereophotogrammetry. International Federation for Medical and Biological Engineering.

[8] Bart, De Ketelaere, Mia Hubert, Eric Schmitt (2015). Overview of PCA-Based Statistical Process-Monitoring Methods for Time-Dependent, High-Dimensional. Journal of Quality Technology.

[9] Xie, Shengkun, Feng Jin, Sridhar Krishnan, Farook Sattar (2012). Signal feature extraction by multi-scale PCA and its application to respiratory sound classification. International Federation for Medical and Biological Engineering, 2012.

Charles Victor B. Saragih is a postgraduate student at Universitas Budi Luhur

Setyawan Widyarto was born in Purwokerto, Central Java, Indonesia in 1963. He received first degree from IPB Bogor in 1986 and M.Sc. degrees in Manufacturing System Engineering and Management from University of Bradford (UK) under Chevening Awards (British Council) in 1998. In 2008 he finished his Ph.D. degree in Computer Science from University Teknologi Malaysia. in 2008, he joined University Selangor (Unisel). Since 2012, he has been an Associate Professor with the Computer Science Department (Industrial Computing), Unisel Bestari Jaya, Kuala Selangor. His research interest includes (agile) software engineering, virtual environment/reality and image processing. He is an editor member of some journals.