

FACE RECOGNITION USING PRINCIPAL COMPONENT ANALYSIS WITH MEDIAN FOR NORMALIZATION ON A HETEROGENEOUS DATA SET

G. Shree Devi¹, M. Munir Ahamed Rabbani², A. Jaya³

Department of Computer Applications,
B.S Abdur Rahman University, Tamil Nadu, India

¹shreedevi@bsauniv.ac.in

²marabbani@gmail.com

³jaya@bsauniv.ac.in

ABSTRACT

Recognizing Faces helps to name the various subjects present in the image. This work focuses on labeling faces on an image which includes faces of humans being of various age group (heterogeneous set). Principal component analysis concentrates on finding the mean of the data set and subtracts the mean value from the data set with an intention to normalize that data. Normalization with respect to image is the removal of common features from the data set. This work brings in the novel idea of deploying the median another measure of central tendency for normalization rather than mean. The above work was implemented using matlab. Results show that Median is the best measure for normalization for a heterogeneous data set which gives rise to outliers.

KEYWORDS

Normalization, median, outliers, heterogeneous.

1. INTRODUCTION

1.1. Face recognition

Face recognition is the major biometric measure which has wide range of applications. The various categories of applications include Face ID, Access Controls, Security, Surveillance, Law Enforcement and Human Computer Interactions and so on[7]. Automatic face recognition is existing today, but requires more refinement to meet the existing and growing challenges in the Information era. For example when recognizing faces in an airport or any entry restricted areas, subjects using the system could be from different age groups, different nation and we could even consider animals like cat or dog. So, this paper concentrates on considering the images taken in an unconstrained real time environment where the subjects could be a baby, old man, young man,

Natarajan Meghanathan et al. (Eds) : ITCSE, ICDIP, ICAIT - 2013

pp. 147–153, 2013. © CS & IT-CSCP 2013

DOI : 10.5121/csit.2013.3915

middle aged man or human from different parts(heterogeneous data set) of the world whose face carry few common details compared to human from the same nation(homogeneous data set). Figure 1.1 shows few subjects which are considered for face recognition.



Figure 1.1 Heterogeneous data set

1.2. Principal Component Analysis(PCA):

The Eigenfaces method is a very proven method which was introduced by Turk and Pentland 1991. Eigen faces method is based on Principal Component Analysis technique for data reduction and feature extraction[8]. Principal component analysis (PCA) is probably the most popular multivariate statistical technique and it is used by almost all scientific disciplines[11]. Its goal is to extract the important information from the image and to express this information as a set of new orthogonal variables called principal components[8]. It can be used for feature extraction, compression, classification, and dimension reduction et cetera. The aim of this paper is to introduce a modification in the Principal component analysis technique which is vital in the normalization of the input data set and leads to a construction of well defined eigenfaces which is subsequently used for Face recognition in a heterogeneous data set[9].

The remaining paper is organized as follows: Related work is discussed in Section 2 and Proposed method in Section 3. Experimental results are shown in Section 4 and Conclusion is given in Section 5.

2. RELATED WORK

Principal Component analysis PCA has been widely investigated and has become one of the most successful approaches in face recognition[5]. Principal component analysis (PCA), also known as Karhunen-Loeve expansion or Hotelling Transformation, is a classical feature extraction and data representation technique widely used in the areas of pattern recognition and computer vision[6]. Principal Component Analysis (PCA) is one of the most successful techniques that have been used in image recognition and compression.

The paper titled Face recognition using PCA, eigenface and ANN (Artificial Neural Networks) has shown 97.018% of accuracy on Olivetti and Oracle Research Laboratory (ORL) face database [1]. In A Discriminative Model for Age Invariant Face Recognition by Zhifeng Li et al,

PCA is performed on the training set, after extracting the SIFT (Scale in variant feature transformation) or MLBP (Multi - Scale local binary patterns) and all the eigenvectors with non-zero eigenvalues are used as the candidates for constructing random PCA subspaces [2]. In the paper Face recognition using Pca, Lda and Ica approaches on colored images, experimental results have shown that PCA out performs LDA(Linear Discriminant Analysis and ICA(Independent Component Analysis)[3]. In Combined face and gait recognition using alpha matte preprocessing principal component analysis (PCA) followed by multiple discriminant analysis (MDA) to reduce the size of the feature vector. A combination of PCA and MDA, results in the best recognition performance[4]. Using principal-component analysis (PCA), many face recognition techniques have been developed: namely eigenfaces [Turk and Pentland 1991], which use a nearest- neighbor classifier and feature-line-based methods, which replace the point-to-point distance with the distance between a point and the feature line linking two stored sample points [Li and Lu 1999] are developed and deployed.

3. PROPOSED METHOD

Input to the Face recognition system is set of faces. The data set considered comprises of faces belonging to various regions in the globe and subjects of young age, babies and old aged people. This heterogeneous data set is taken with an assumption that all the images considered are of the same size 140X175. The figure 1.1 shows the sample images considered for recognition.

The original PCA algorithm subtract the mean value from each input data. Mean is a single value which gives the central tendency of measure. Measures of central tendency are sometimes called measures of central location. The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others. The mean is only representative if the distribution of the data is symmetric, otherwise it may be heavily influenced by outlying (outliers) measurements. In such conditions median is always best representative of the centre of the data. Mathematically the median is preferred over the mean (or mode) when our data is skewed (i.e., the frequency distribution for our data is skewed)[7] . The measure asymmetric of a distribution is called skewness.

3.1 Steps in Principal Component Analysis using Median for Normalization:

1. The First step is to construct a training data set with M images of same size.
2. Convert the RGB image to a Gray Scale image.
3. Convert each face 2-Dimensional image data into a face vector of 1 Dimension by concatenating each row to the 1st row in a chronological order.

The Figure 3.1 shows the intensity value of the pixels for the images considered where each row represents face vector of one person in the data set.

Columns 24481 through 24497

69	72	74	73	74	80	84	86	89	92	92	89	86	84	80	75	74
16	16	16	16	17	18	19	19	18	19	19	19	19	18	19	19	19
241	231	238	181	87	107	111	107	100	97	100	97	108	94	100	93	89
22	38	49	50	44	51	72	59	68	69	67	69	70	68	73	76	75
52	53	55	56	56	57	58	59	59	61	65	68	71	73	75	77	80
49	48	67	97	105	101	103	101	103	103	103	103	103	103	103	103	103
33	33	34	34	34	35	35	35	37	36	35	34	34	35	36	37	34
221	220	220	222	226	227	227	226	229	229	230	231	231	231	230	230	236
49	50	41	30	35	53	66	67	68	67	66	66	67	67	66	65	66

Figure 3.1- Face vectors representing the pixel intensity values.

4. Steps for Normalization

4.1 Construct a Matrix T with each row containing the face information of one person in the data set.

4.2 Calculate the Median vector Md, for T column wise.

The following Screen Shot shows that the values are not symmetric and calculation of the skewness. Skewness being the measure for the degree of asymmetry of a distribution is calculated to determine whether mean or median is to be used. The Screen shot reveal that Skewness value corresponding to the column considered is +ve (1.5132). Similarly the skewness of each column is considered and show either a negative or positive value hence median is considered over the mean.

```

a1=rgb2gray(a);
b1=rgb2gray(b);
c1=rgb2gray(c);
d1=rgb2gray(d);
e1=rgb2gray(e);
f1=rgb2gray(f);
g1=rgb2gray(g);
h1=rgb2gray(h);
i1=rgb2gray(i);
a2=reshape(a1,1,[]);
b2=reshape(b1,1,[]);
c2=reshape(c1,1,[]);
d2=reshape(d1,1,[]);
e2=reshape(e1,1,[]);
f2=reshape(f1,1,[]);
h2=reshape(h1,1,[]);
i2=reshape(i1,1,[]);
mat=[a2;b2;c2;d2;e2;f2;g2;h2;
g2=reshape(g1,1,[]);
mat=[a2;b2;c2;d2;e2;f2;g2;h2;
skewness(double(mat));
ans
c24492=mat(:,24492);
ans
c24492=mat(:,24492);
c24492
skewness(double(c24492));
ans
Columns 24497 through 24500
1.6219 1.5275 1.4351 1.4810
>> c24492=mat(:,24492);
>> c24492
c24492 =
    89
    19
    97
    69
    68
    103
    34
    231
    66
>> skewness(double(c24492));
>> ans
ans =
    1.5131

```

5. Subtract the Median vector Md for the corresponding column with each column value. The result obtained is normalized matrix N which is represented by \emptyset . Normalization in image is to remove the common features from each face and each face is left with unique features.

$$\text{i.e., } N_i = T_i - Md$$

6. Next proceed with the Calculation of the EigenVector and EigenValues, EigenValues are obtained from the Covariance matrix C the normalized matrix N_i . The Covariance matrix C is calculated using the formula

$$C = (N_i) (N_i)^T$$

7. Now proceed with the calculation of the eigen values and the eigenvectors using $[V,D] = \text{eig}(C)$.

8. Select $K < M$ eigen vectors corresponding to the greatest eigen values in the matrix D. This K eigen vectors are used to reconstruct all the M faces in the Training set.

9. The eigenvectors are also called as eigen faces must be of the original dimensionality of the face vectors. So map the eigen vector in lower dimensional space to the corresponding eigenvector in the higher dimensional space. Suppose U_i, V_i are the eigen vector in higher dimensional space and lower dimensional space then

$$U_i = A \cdot V_i$$

10. Steps for reconstruction of the faces in the heterogeneous data set.

$$\text{Final Data} = U_i \times N_i$$

$$N_i = (\text{Final Data})^T \times U_i$$

$$\text{Original data set} = (\text{Final Data})^T \times U_i + Md$$

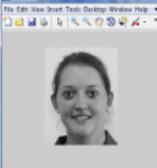
11. Recognition

For recognition express each image in the training set as a linear combination of the eigenfaces plus the median image. When considering the linear combination of the eigenfaces calculate the weight associated with each eigenface. The weight determine the proportion of contribution of the specific eigenface towards the reconstruction of the original face. Once the weight vector for all the images in the data set is determine, to recognize the unknown face, we have repeat the step1 to step5 of the above specified algorithm. Then try the represent the unknown face as a combination of the already constructed eigenfaces and determine the weight vector. If the weight vector for the unknown face is one among the weight vector of the data set already considered then we declare the face as a known face other vice versa.

4. EXPERIMENTAL RESULTS

Table 1 shows the heterogeneous images considered for applying pca using median for normalization, The median image obtained and the reconstructed image are shown

Table 1. Input images, median image and reconstructed image

Input Images	
Median Image	
Reconstructed image from the eigen faces	

5. CONCLUSION

From the experimental results it is clear that face reconstruction with the eigenfaces after normalizing the data set using Median as a Central Measure is efficient on a Heterogeneous data set. Recognition follows effectively as subsequent step after the reconstruction. Future work will be to concentrate on methodology for determining the weight vector which is impact apply PCA for Face recognition.

REFERENCES

- [1] Agrawal, H. ; Jain, N. ; Kumar. M. (2010) "Face Recognition using Principle Component Analysis, Eigenface and Neural Network", International Conference on Signal Acquisition and Processing published in IEEE Computer Society, pp 310 -314.
- [2] Zhifeng Li, Member, Unsang Park, Member, and Anil K. Jain, (2011) "A Discriminative Model for Age Invariant Face Recognition", IEEE Transactions on information forensics and security, Vol. 6, No. 3, September pp 1028 -1037.
- [3] Önsen TOYGAR1, Adnan ACAN2, (2003) " Face recognition using PCA, LDA and ICA approaches on colored images" Journal of electrical & electronics engineering Vol. No. :: 3 : 1 pp (735-743).
- [4] Hofmann, M.Schmidt, S.M., Rajagopalan, A.N. Rigoll. G (2012), "Combined face and gait recognition using alpha matte preprocessing" , Proceedings 5th IAPR International Conference on Biometrics (ICB).
- [5] A. Pentland, (2000), "Looking at People: Sensing for Ubiquitous and Wearable Computing," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, pp. 107-119.
- [6] Jian Yang, David Zhang, Senior Member , Alejandro F. Frangi, and Jing-yu Yang, (2004), "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition" , Transactions on pattern analysis and machine intelligence, Vol. 26, No. 1, pp 131 –137.
- [7] Handbook of Face Recognition edited by Stan Z. Li, Anil K. Jain.
- [8] Matthew Turk and Alex Pentland, (1991) " Eigen faces method", Journal of Cognitive NeuroScience Vol 3, No. 1 .

- [9] Patrik Kamencay, Martina Zachariasova, Robert Hudec, Roman Jarina, Miroslav Benco, Jan Hlubik, (2013) " A Novel Approach to Face Recognition using Image Segmentation Based on SPCA-KNN Method", Radio Engineering Vol 22, No1.
- [10] M A Rabbani and C. Chellappan, (2007) "A Different Approach to Appearance –based Statistical Method for Face Recognition Using Median", IJCSNS International Journal of Computer Science and Network Security Vol 7, No. 4, pp : 262-267
- [11] A tutorial on Principal Components Analysis Lindsay I Smith February 26, 2002
www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.

AUTHORS

G. Shreedevi presents working as Assistant Professor(Sr.Grade) in the Department of Computer Applications, B. S Abdur Rahman University. Pursuing Ph. D in Image Processing under the able guidance of Dr. Munir Ahamed Rabbani.



Dr.M.Munir Ahamed Rabbani Professor Department of Computer Applications, B. S Abdur Rahman . Completed his Ph. D in Area of Image Processing. Posses rich International Experience in the Field of Teaching and Research.



Dr. Jaya Professor, Department of Computer Application, B. S. Abdur Rahman University. Completed Ph. D in area of Artificial Intelligence.

