AN ENHANCED PRINCIPAL COMPONENT ANALYSIS APPROACH FOR EXTRACTING FEATURES IN PALMPRINT IMAGES

Akande Noah Oluwatobi[#1], Oyediran Mayowa Oyedepo[*2], Ogundokun Roseline Oluwaseun[#3], Adeyemo Isiaka Akinkunmi[*4]

[#1,3]Computer Science Department, Landmark University, Omuaran, Kwara State, Nigeria

[*2,4] Computer Science and Engineering Department, Ladoke Akintola Unuiversity of Technology, Ogbomoso, oyo State, Nigeria

***Abstract*— In reducing the dimensions of feature space so as to achieve better performance, minimal features that yet represent the original image maximally need to be extracted. Principal Component Analysis (PCA) which is an appearance based projection method has been widely used in this regards. Literatures have shown that PCA performs well for dimensionality reduction but instead of reducing within class variations, it increases it therefore resulting in its low performance. In contrast, Independent Component Analysis (ICA) performs preferably well in this instead since it linearly transforms an original data into completely independent components thereby reducing within class variations. While both PCA and ICA have been used for the purpose of feature extraction, no single feature extraction algorithm is exclusively flawless in all ramifications. In lieu of this, this paper introduces an Enhanced Principal Component Analysis (EPCA) approach which first extracts principal components from data and further uses the extracted components as input to an ICA algorithm . The EPCA was employed for feature extraction in a palmprint recognition system. The resultant system was validated using 900 palmprint images which were downloaded from three public palmprint databases. Recognition and verification rates arising from the palmprint recognition system showed that the EPCA performed well than PCA algorithm.**

***Index Terms*— Enhanced Principal Component Analysis, palmprint recognition system, Independent Component Analysis, Principal Component Analysis.**

I. Introduction

Feature extraction helps to derive a meaningful representation of images by mapping high dimensional input space into a lower dimensional feature space. Not only does feature extraction reduces the dimension of an image with an aim to get a classifier that runs faster and uses less memory, it also improves the classification in revealing the intrinsic dimension of the observed pattern. [1]. No feature extraction algorithm is exclusively flawless in all ramifications; for some problems a single algorithm is enough to get the desired features while some require two or more algorithms to get the expected results. Every algorithm has its own strength and weakness, hence, to get more accurate results the strength of one algorithm can be used to complement the weakness of the other. This paper introduces how ICA could be used to overcome PCA limitation of increasing the within class variations of computed features. The approach was validated across two Multispectral Palmprint Databases (CASIA and PolyU) and IIT Delhi touchless palmprint image database (Version 1.0). The result of the approach was further compared to the instance where PCA was used alone.

II. FEATURES EXTRACTION USING PCA ALGORITHM

Features extraction using PCA involves the calculation of the eigenvalues, criterion function and composition of a data covariance matrix usually after algorithmic mean centering of the principal features as well as finding the difference of the dimension vector and the algorithmic mean. PCA steps involve:

1) Input a dataset x with k-dimensions such that xi ϵ Rk, where i=1,2,...n
2) Compute the mean vector m such that $m = \frac{1}{n}\sum_{i=1}^{n} x_i$
3) Compute the scatter matrix S such that $S = \sum_{i=1}^{n}(x_i - m)(x_i - m)^T$
4) Compute the eigenvectors ($e_1, e_2, \ldots e_k$) and the corresponding eigenvalues (λ_1, λ_2, ... λ_k) from the scatter matrix.
5) Sort the eigenvalues into decreasing order and discard the corresponding eigenvectors that has small eigenvalues
6) Develop a k ✕ i dimensional matrix w from the topmost i-eigenvectors.
7) Use the k ✕ i dimensional matrix to transform the dataset into a principal subspace y such that y = WT ✕ x.

III. THE PREPROCESSING STAGE OF ICA

For the purpose of reducing the dimension space, the two ICA pre-processing stage proposed in [2] were adopted. This involves:

A. Centering the image

This was achieved by subtracting the sample mean vector m=E(x) from the input vector data y such that the centered vector $\vec{v} = y - m$ with m = 0. Centering enables ICA to deal with only zero mean data. Whitening ensures that the dimensions of the data are treated equally a piori before running the ICA algorithm. After estimating the mixing matrix A with the centered data, the estimation was completed by adding the mean vector of s back to the centered estimates of s such that

$E(s) = A^{-1}E\{x\}$

(1)

Note: s is the vector of unknown source images and since the principal component extracted from the source image serves as the input to the ICA, then s = y.

B. Whitening the Image

This means that after centering, the observed vector x needs to be transformed linearly so as to obtain a new vector which is white (i.e. its components have variances which equals unity and are uncorrelated. The covariance matrix equals the identity matrix as illustrated in equation 2.

$E(\tilde{x}\tilde{x}^T) = I$

(2)

The whitening transformation now involves finding the Eigen Value Decomposition (EVD) of the covariance matrix such that $E(\tilde{x}\tilde{x}^T)$ equals EDET where E is the orthogonal matrix of the eigenvectors of $E(\tilde{x}\tilde{x}^T)$ and D is the diagonal matrix of the eigenvalues of E. The diagonal matrix d is given as diag (d_1, d_2,…d_n). Whitening can now be achieved using equation 3.

$\tilde{X} = ED^{-\frac{1}{2}}E^T As = \tilde{A}$ s

(3)

The beauty of the whitening transformation is seen in the new mixing matrix $\tilde{A}$ which is orthogonal. With this, whitening helps to reduce the number of parameters to be estimated. Instead of having to estimate the n2 elements of the original matrix A, we only need to estimate the new orthogonal mixing matrix $\tilde{A}$. The new matrix has n (n-1)/2 degrees of freedom.

IV. FEATURES EXTRACTION USING ICA ALGORITHM

As proposed in [2], ICA considers an original dataset as a mixture of an unknown set of statistically independent source images and an unknown mixing matrix. A separating matrix is then used to recover a set of statistically independent basis images as shown in Fig. 1.

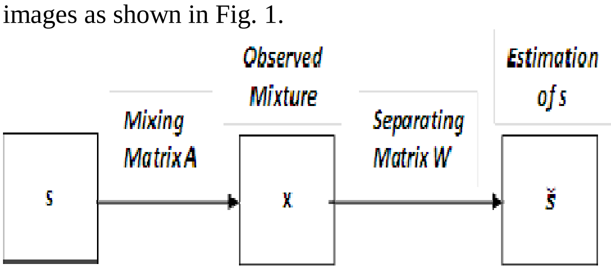

Fig. 1: ICA Implementation

Where s is the vector of unknown source images, v is the vector of observed mixtures and A is the unknown mixing matrix. The mixing process is written as:

$v = A\check{s}$ (4)

The goal of ICA is to find the separating matrix W such that:

$\check{s} = Wy$

(5)

The steps taken in computing independent components are the following:

1) Input matrix y of m x n dimension
2) Compute the transpose of the input matrix (yT)
3) Compute the independent basis vector Bˇ such that $\check{B}$ such that $\check{B} = \check{W} * y^{-1}$
4) Compute the independent coefficient R such that R = Z * y, where Z is the zero mean images.
5) Compute the coefficient matrix C such that $C = R * \check{B}$
6) Finally, reconstruct the image I such that $I = C * \check{B}$

V. FEATURES EXTRACTION USING THE ENHANCED PRINCIPAL COMPONENT ANALYSIS APPROACH

Both PCA and ICA attempts to transform a data with dimension d into a k-dimensional subspace such that k<d. In order to find the appropriate k-dimensional subspace that represents the data well, PCA constructs a principal subspace by computing the eigenvectors from the principal subspace. The eigenvectors which are also called principal components are grouped to form a scatter-matrix. Each eigenvectors is associated with an eigenvalue which determines the length or magnitude of the eigenvectors. If the eigenvalues are similar, this is an indication that a good principal subspace has been constructed or if there are some eigenvalues which are higher than the others, the corresponding eigenvectors can be selected to form the desired principal subspace while the eigenvectors with small eigenvalues will be discarded. The block diagram of the enhanced principal component analysis is shown in Figure 2:

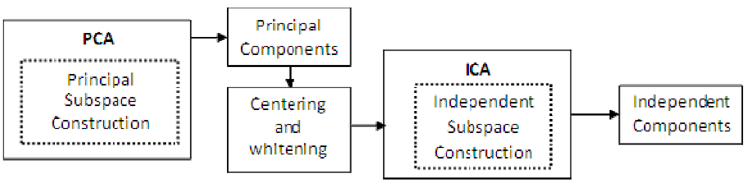

Figure 2: Block Diagram of the Enhanced Principal Component Analysis Approach

The steps taken in constructing the principal subspace are enumerated below:

1) dataset x with k-dimensions were selected such that $x_i \in R^k$, where i=1,2,...n
2) the class label of the selected dataset were ignored since they are not needed for PCA analysis
3) the k-dimensional mean vector m was computed such that m $m = \frac{1}{n}\sum_{i=1}^{n} x_i$
4) the scatter matrix S was computed such that $S = \sum_{i=1}^{n}(x_i - m)(x_i - m)^T$
5) the eigenvectors (e_1, e_2... e_k) and the corresponding eigenvalues (λ_1, λ_2, ...λ_k) were computed from the scatter matrix.

6) To ensure the accuracy of the eigenvectors and eigenvalues $Sv = \lambda v$, $1 \le i \le n$ where $\lambda 1 \ge \lambda 2 \ge ... \lambda n \ge 0$.
7) The eigenvalues were sorted into decreasing order and the corresponding eigenvectors that has small eigenvalues were discarded because they contain least information about the distribution of the data.
8) The top i-eigenvectors were used to form k x i dimensional matrix w (where every column represents an eigenvector)
9) The k x i dimensional matrix was used to transform the dataset into the new principal subspace y from which the principal features computed.
10) The new principal subspace y is given as WT ✕ x

The pre-application of PCA can discard small trailing eigenvalues and reduce computational complexity by minimizing pair-wise dependency [3]. The next stage after extracting the principal component is to preprocess this principal component by centering and whitening. The number of independent components found by the ICA algorithm corresponds to the dimensionality of the input dataset [4] and since PCA has been initially used to extract the principal components from the original dataset, ICA was applied on the subspace y created by the PCA algorithm using these steps:

11) the transpose of subspace y was computed
12) the independent basis vector $\breve{B}$ was computed such that $\breve{B} = \breve{W} * y^{-1}$
13) the independent coefficient R was computed such that R = Z * y, where Z is the zero mean images.
14) The coefficient matrix of ICA is given as $C = R * \breve{B}$
15) Finally the reconstructed hybrid image is given as $I = C * \breve{B}$

VI. EVALUATING THE ENHANCED PRINCIPAL COMPONENT ANALYSIS APPROACH

To validate the enhanced algorithm, it was employed in the development of a palmprint recognition system. This involves the following steps:

A. *Palmprint Image Acquisition*

Palmprint images were acquired from CASIA, PolyU and IIT Delhi palmprint databases. Three hundred palmprint images were downloaded from each database; this was done with an intention to verify the performance of the enhanced algorithm across different popular databases. The first hundred images from each database were used as training data while the remaining two hundred from each database were deployed as testing data.

B. *Preprocessing the Palmprint Images*

Preprocessing was carried out so as to enhance the quality of the acquired palmprint images, to improve its visual appearance and to further convert the image to a form better suited for analysis by a human or a machine. The preprocessing stage involves enhancement techniques such as brightness and contrast stretching, normalization, image de-noising and integration, gray scale conversion, segmentation, histogram equalization, thresholding and binarization.

C. *Image Enhancement*

The quality of the downloaded palmprint image was improved through image enhancement; this was done so as to ensure that the vein patterns, palm lines and the Region of Interest (ROI) can be more easily detected for the segmentation and feature extraction process. Also, the background of the acquired palmprint images has a very low intensity values hence, there is a need to increase the quality of the palmprint images. Image enhancement involves brightness stretching, normalization, image de-noising and integration

D. *Gray Scale Conversion*

The acquired palmprint images were coloured Red (R) Green (G) and Blue (B) palmprint images which was not suitable for image analysis, hence, the coloured RGB palmprint images were converted to a gray scale format using equation (6)

$$\text{Gray value} = 0.2989X*R + 0.7870X*G + 0.1140X*B \quad (6)$$

X is the weight value of the coloured compound and R, G, B indicates the colour compound of the palmprint images.

E. *Segmentation*

Effective segmentation can reduce image noise and the difficulty of matching the palmprint features. It was also carried out so as to separate the palmprint image from its background which will further ensure an accurate and effective feature extraction process. Due to the possibility of having some variations in contrast and illumination gradient, thresholding and binarization were also carried out. Thresholding involves calculating a threshold value for each palmprint images while binarization involves converting the gray scaled palmprint images to binary palmprint. The optimum threshold value and binary image conversion was carried out using Otsu's method as proposed by authors in [5].

$$P_{gray}(x,y) = \begin{cases} 255 & \text{if } P_{gray}(x,y) > Th \\ 0 & \text{if } P_{gray}(x,y) < Th \end{cases} \quad (7)$$

$$P_{bin}(x,y) = \begin{cases} 1 & \text{if } P_{gray}(x,y) > Th \\ 0 & otherwise \end{cases} \quad (8)$$

where Pbin is binary palmprint, Pgray is the original gray-scale palmprint, and Th is the optimum threshold value.

F. *Extracting Palmprint Region of Interest*

Extracting the Region of Interest (ROI) during preprocessing stage is very important as not all the features in a palmprint image are needed. It is the region that contains the palmprint features such as principle lines, wrinkles and ridges needed for recognition and it involves cropping the palmprint image from the hand image. Extracting ROI is needed so as to reduce redundancy from palmprint image [6]. ROI can either be circular, half elliptical or square region with the square region being the easiest and widely used [7]. According to [8], the

ROI extraction methods include the following main steps, which are:

1) To separate the fingers and palms.
2) To find the two valley points of the index finger and middle finger, ring finger and little finger.
3) To rotate image based on the two valley points and correct image position.
4) To create coordinate system according to valley points and determine ROI.

The square shaped ROI used in the experiment is of size 150 x 150.

G. Histogram Equalization

Histogram equalization involves stretching the contrast of the high histogram regions of the palmprint images and compressing the contrast of the low histogram regions. This is done so as to achieve a wider and more uniform distribution of intensity values and to ensure that the intensity values are spread over the ROI of the palmprint images. In instances where the ROI occupies only a small portion of the image, the palmprint image will not be successfully enhanced by histogram equalization but the thresholded version of the histogram equalization methods tries to overcome this problem by restricting the enhancement rate.

VII. EXPERIMENTAL RESULTS

Recognition rates of PCA and the EPCA across three public palmprint databases were measured. In determining the recognition rate, Probabilistic Neural Network (PNN) and Euclidean Distance (EU) classifiers were employed. Furthermore, the verification rate of the palmprint recognition system was computed using False Acceptance Rate (FAR) and False Rejection Rate (FRR).

A. Recognition Rate

The recognition rates of PCA and EPCA are shown in Table 4.1. It can be observed from the results of PNN and EU used as classifiers that the EPCA performs better than PCA across the three public databases. It was also observed that PNN classifier performs better than EU classifier across all the three databases for both PCA and EPCA. This is due to the fact that PNN can generalize well on data that it has not seen before and can also account for subtle differences that exist between data while EU only take into account the differences between the data. It was also observed that both PCA and EPCA have a higher recognition rate with IIT Delhi palmprint database when compared to CASIA and PolyU palmprint Databases. Conclusively, it is obvious that the EPCA outperform PCA from both classifiers used.

Table 4.1: Recognition Rates of the algorithms across CASIA, PolyU, IIT Delhi Databases

Algorithms	Recognition Rate		
	Palmprint Databases	Probabilistic Neural Network (%)	Euclidean Distance (%)
PCA	CASIA	94.3600	92.3100
	PolyU	95.4511	92.5000
	IIT Delhi	95.5300	93.5130
EPCA	CASIA	98.1000	95.1500
	PolyU	97.8150	95.0450
	IIT Delhi	98.5250	96.5105

B. Verification Rate

To find the verification rate of the system, standard error rates (False Acceptance Rate (FAR) and False Rejection Rate (FRR)) were computed as illustrated in equation 9 and 10. Both rates must be as low as possible for the biometric system to work effectively.

$$FAR = \frac{\text{Number of accepted imposter palmprint}}{\text{Total number of imposter palmprint}} \; X\ 100\% \tag{9}$$

$$FRR = \frac{\text{Number of rejected genuine palmprint}}{\text{Total number of genuine palmprint accesses}} \; X\ 100\% \tag{10}$$

Using the FAR and FRR values, the Total Success Rate (TSR) of the system was also computed using equation 11. The results of the computed verification rates are as shown in Table 4.2.

$$TSR = 1 - \frac{FAR + FRR}{Total\ Number\ of\ Genuine\ Palmprint\ Accesses} \; X\ 100\% \tag{11}$$

Table 4.2: Verification Rate of the algorithms across CASIA, PolyU and IIT Delhi Databases

Algorithms	Verification Rate			
	Palmprint Databases	FAR (%)	FRR (%)	TSR (%)
PCA	CASIA	1.4280	1.2150	97.5000
	PolyU	1.3500	1.2200	96.6000
	IIT Delhi	1.3250	1.1800	97.5500
EPCA	CASIA	1.3255	1.1800	98.8250
	PolyU	1.3150	1.1500	98.5000
	IIT Delhi	1.3025	1.1000	99.0000

The verification result shows that the EPCA outperforms PCA across the three public databases used. It recorded a 99% verification rate with IITT Delhi touchless palmprint database and approximately 98% across CASIA and PolyU palmprint databases. The system also recorded a low FAR and FRR which is good for a biometric system.

VIII. CONCLUSION

This work has showed the possibility of improving the performance of PCA algorithm for features extraction in palmprint images. The EPCA introduced outperforms PCA in terms of recognition and verification rate. This shows that the output of PCA algorithm if used as an input to ICA for further features extraction will yield a biometric system with a better recognition and verification rate.

REFERENCES

[1] Trupti M. Kodinariya and Prashant R. Makwana, "Hybrid Approach to Face Recognition System using Principle component and Independent component with score based fusion process", CoRR abs, pp 1-11, 2014

[2] Tee Connie, Andrew Teoh, Michael Goh, David Ngo, "Palmprint Recognition with PCA and ICA", Faculty of Information Sciences and Technology, Multimedia University, Melaka, Malaysia, 2003.

[3] Baek, K., Draper, B.A., Beveridge, J.R., She, K., "PCA vs ICA: A Comparison on the FERET data set", International Conference on Computer Vision, Pattern Recognition and Image Processing in conjunction with the 6th JCIS, 2001.

[4] 4. Hyvärinen, A. and E. Oja, Independent Component Analysis: Algorithms and Applications. Neural Networks, vol. 13(4-5), pp.411-430, 2000.

[5] 5. Gonzalez R. C. and Woods R. E., "Digital Image Processing", Prentice Hall, 3rd Edition, 2008.

[6] 6. Ali Younesi and Mehdi Chehel Amirani, "Efficient Algorithms for Palmprint Identification Based on Non-linear Feature Extraction and Contourlet Transform", International Journal of Natural and Engineering Sciences, Vol. 8, Issue 1, Pp. 53-60, 2014.

[7] 7. Jaspreet Kour, Shreyash Vashishtha, Nikhil Mishra, Gaurav Dwivedi and Prateek Arora, "Palmprint Recognition System ", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 2, Issue 4, Pp. 1006-1009, 2013

[8] 8. Shuang Xu and Jifeng Ding), "Palmprint Image Processing and Linear Discriminant Analysis Method ", Journal of Multimedia, Vol. 7, No. 3, Pp. 269-275, 2012.