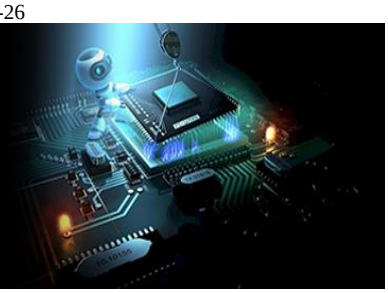
International Journal of Engineering in Computer Science



E-ISSN: 2663-3590
P-ISSN: 2663-3582
IJECS 2023; 5(2): 21-26
Received: 08-06-2023
Accepted: 17-07-2023

Mahesha Y
Department of CSE, Mysuru Royal Institute of Technology, Mandya, Karnataka, India

Nagaraju C
Department of ECE, The National Institute of Engineering, Mysore, Karnataka, India

Corresponding Author:
Mahesha Y
Department of CSE, Mysuru Royal Institute of Technology, Mandya, Karnataka, India

Principal component analysis and local binary patterns: A comparative study using different databases

Ayodele Oloyede, Abubakar Dauda, Basiru Saka, Enem Theophilus and Ibrahim Akanbi

DOI: https://doi.org/10.33545/26633582.2023.v5.i2a.95

Abstract
This paper compares the efficiency of two popular feature extraction methods Principal Component Analysis and Local Binary Pattern using two different iris databases CASIA and UBIRIS. For classification, Support Vector Machine has been used. The models were tested using 200 iris images. The Receiver Operating Characteristic Curve has been drawn and AUC was calculated. The result shows that LBP achieves better performance with both CASIA and UBIRIS databases compared to PCA. The experiment has been extended by varying the dataset sizes. The result has shown that LBP outperforms PCA with both CASIA and UBIRIS.

Keywords: Local binary, component analysis, databases

1. Introduction
The authentication of as person is of high importance in modern days [1]. Other than passwords and magnetic cards, authentication using biometric system is based on bodily or behavioural characteristics of a person. The physical characteristics such as palm print, finger print, face and iris recognition are proved to be accurate and fast. These characteristics are unique to an individual and remain stable through life, and the pattern variability is high among different persons make iris very fascinating for use as biometric system.
A biometric system enables the identification of a person using distinct feature or characteristic exhibited individual person. The Biometric systems are designed for recognition by taking various physical traits as input such as palm prints, finger prints, face, and the iris [2].

Related Work
The iris is the coloured circle located around the pupil which includes number of randomly distributed structures which are not mutable and hence iris are distinct from another [3]. The past of iris biometrics recognition system is considered as roughly from 2001 [4]. The original algorithm for iris identification works with analysis of texture [5]. The algorithm uses codes which are generated using two dimensional Gabor wavelet. The accuracy achieved is more compared to other methods. The important work done by Wildes [6] has adopted algorithm which uses isotropic band-pass decomposition generated by the application of Laplacian Gaussian filters to the image. This approach has modelled the eyelids using parabolic arcs. Boashash and Boles [7] have proposed a method which has adopted zero-crossings. The zero crossings of the wavelet transform are calculated at different resolutions of circles of the iris which concentric. The single dimensional signals are compared with the features with the help of various dissimilarity functions. Same kind of approach has been presented which has adopted discrete dyadic wavelet transform and showed better efficiency [8]. The Multi-resolution Independent Component Identification (M-ICA) has good ability to use signals along with time frequency to generate the iris features. The accuracy obtained is low since the M-ICA does not work efficiently with class-separability.
Chen and Yuan have proposed a new approach for generating the features of iris using fractal dimension [9].

The iris part has been divided into small parts and the local fractal dimension features are generated as the iris code. The generated patterns are then matched using the neural networks. Robert *et al*. have proposed an algorithm for achieving localization and extraction. For localization, integrodifferential operators along with a Hough Transform has been adopted and for extraction, emergent frequency has been adopted [10]. Code for iris has been generated using both emergent frequency and instantaneous phase.

Li Ma. *et al*. [11] have adopted Haar-Wavelet transform for features extraction in iris. The transformation has been used on the image to generate a feature vector. To classify the vectors, two approaches namely weight vector initialization and winner selection methods are adopted [12]. The recognition capability of classification methods are dependent on feature quality and the size data used for training. The features are generated from shape and texture of segmented parts. Karu *et al*. [13] presented an approach to achieve automatic identification of textured parts and the categorization uses statistical measures. The Co-occurrence approach is claimed to be the best for classification of texture by Conners and Harlow [14]. The iris structure can be classified with the help of coherent Fourier spectra using optical transmission [15]. The iris biometric designed by Hamed Ranjzad [16] has adopted Independent Component Analysis (ICA) for feature extraction. Image acquisition is performed at different illumination and noise levels.

Material and Methods

CASIA and UBIRIS datasets have been used in the experiment. Fig. 1 shows the overall architecture of the work carried out. For feature extraction, two methods have been used and compared. One is Principal Component Analysis (PCA) [17, 18] and another one is Local Binary Pattern (LBP). Support Vector Machine has been used for classification.

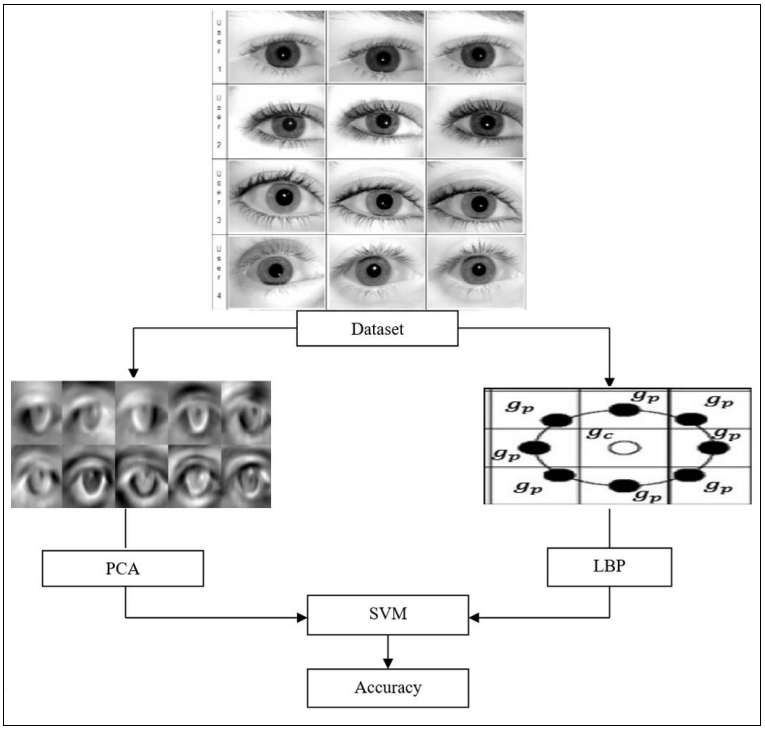


Fig 1: Overall architecture

3.1.1 Principal Component Analysis (PCA)

PCA is one of the most import aspects of linear algebra and it is widely being adopted for analysis purpose. Since it is simple and non-metric method it can be used in neuroscience as well as in computer graphics. PCA guides how to reduce a high dimension data to a lower dimensional data to explore only the sufficient and important features.

The important aspect of PCA is to minimize the dimensionality of a data consisting of a large number of correlated variables and retaining only the important features of data. This is achieved by converting the data to a set of variables called the principal components (PCs). These PCs are independent and are arranged so that the first few components represent the most important features present in all of the original variables.

3.1.2 Local Binary pattern

The number of methods has been developed to extract the useful features from iris images to perform iris biometric recognition. Local Binary Pattern (LBP) [19, 20] is one among them. LBP make it is possible to represent the image texture and shape.

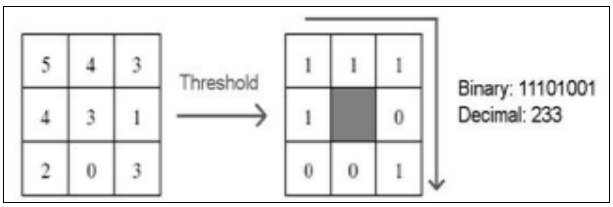


Fig 2: Local Binary Pattern

The LBP operator was originally proposed by Ojala *et al*. [21]. The LBP operator uses pixel's neighbours and takes the centre pixel value as a threshold as shown in Fig. 2. If a neighbour pixel has a larger gray value than the centre pixel then a one is assigned to that pixel otherwise it takes a zero. The LBP code for the centre pixel is evaluated by

combining the eight binary digits to a code.
The value of the LBP code of a pixel (Xc, Yc) is given by Eq. (1) and Eq. (2)

$$LBP_{M,N} = \sum_{m=0}^{M-1} l(g_m - g_c)2^m \quad (1)$$

$$l(x) = \begin{cases} 1, x \geq 0; \\ 0, otherwise \end{cases} \quad (2)$$

Support Vector Machine (SVM)
A SVM [22] which is used to classify data which are linearly separable is called linear SVM which is depicted in Fig. 3. Linear SVM searches for a hyper-plane with the maximum margin. This is why a linear SVM and hence called as a maximal margin classifier (MMC). Steps involved in finding the Maximum Margin Hyper-plane are discussed below:

1. Consider a problem of binary classification consisting of n training data.
2. Each tuple is represented by (Xi, Yi) where Xi = (xi1,xi2.....xim) corresponds to the attribute set for the i^{th} tuple (data in m-dimensional space) and Y_i ε [+,-] denotes its class label.
3. Given {(X_i, Y_i)}, a hyper-plane is generated which separates all X_i into two sides of it.
4. Consider a two dimensional training data with attributes A_1 and A_2 as X = (x_1, x_2), where x_1 and x_2 are values of attributes A_1 and A_2, respectively for X.
5. Equation of a plane in 2-D space can be written as

$w_0 + w_1x_1 + w_2x_2 = 0$ [e.g., ax + by + c = 0] where w_0, w_1, and w_2 are some constants defining the slope and intercept of the line.
Any point lying above such a hyper-plane satisfies

$w_0 + w_1x_1 + w_2x_2 > 0$

Similarly, any point lying below the hyper-plane satisfies

$w_0 + w_1x_1 + w_2x_2 < 0$

6. An SVM hyper-plane is an n-dimensional generalization of a straight line in two dimensions.
7. Euclidean equation of a hyper-plane in R_m is $w_1x_1 + w_2x_2 + ... + w_mx_m = b$ (3) where w_i's are the real numbers and b is a real constant.
8. In matrix form, a hyper-plane thus can be expressed as W.X + b = 0 (4) where W = [w_1,w_2.....w_m] and X = [x_1,x_2.....x_m] and b is a real constant.

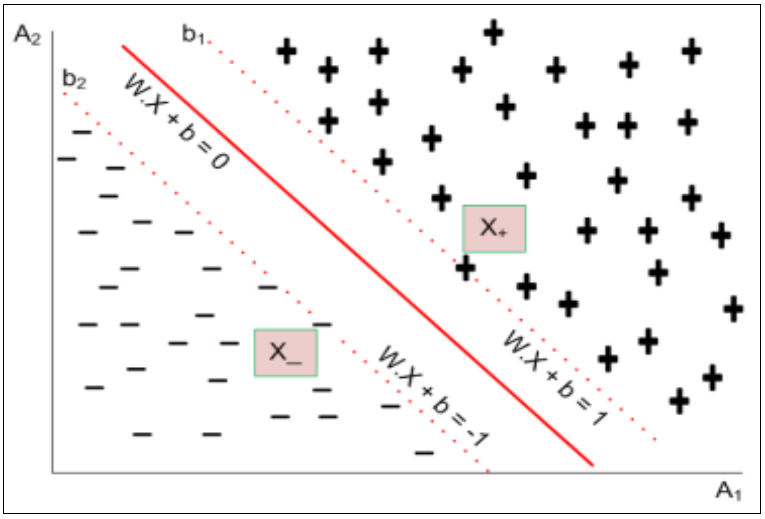


Fig 3: Shows the computation of MMH

Results and Discussion
100 intruder and 100 genuine iris images have been used to conduct the experiment. The True Positive Rate (TPR), False Positive Rate and Accuracy have been presented in Table 1. The accuracy has been evaluated over different probability threshold values.

Table 1: Accuracy for different probability threshold

Probability threshold	FP	FN	TP	TN	Recall	Precision	TPR	FPR	Accuracy
PCA + SVM + CASIA									
>=0.0	100	0	100	0	1	0.50	1	1	50%
>=0.1	88	2	98	12	0.98	0.53	0.98	0.88	55%
>=0.2	85	2	98	15	0.98	0.54	0.98	0.85	57%
>=0.3	75	3	97	25	0.97	0.56	0.97	0.75	61%
>=0.4	65	5	95	35	0.95	0.59	0.95	0.65	65%
>=0.5	18	5	95	82	0.95	0.84	0.95	0.18	89%
>=0.6	10	15	85	90	0.85	0.89	0.85	0.10	88%
>=0.7	8	30	70	92	0.70	0.90	0.70	0.08	81%
>=0.8	5	42	58	95	0.58	0.92	0.58	0.05	77%
>=0.9	1	70	30	99	0.30	0.97	0.30	0.01	65%

PCA + SVM + UBIRIS									
>=0.0	100	0	100	0	1	0.50	1	1	50
>=0.1	82	4	96	18	0.96	0.54	0.96	0.82	57
>=0.2	80	8	92	20	0.92	0.53	0.92	0.80	56
>=0.3	70	9	91	30	0.91	0.57	0.91	0.70	61
>=0.4	63	10	90	37	0.90	0.59	0.90	0.63	64
>=0.5	60	10	90	40	0.90	0.60	0.90	0.60	65
>=0.6	58	16	84	42	0.84	0.59	0.84	0.58	63
>=0.7	54	20	80	46	0.80	0.60	0.80	0.54	63
>=0.8	46	40	60	54	0.60	0.57	0.60	0.46	57
>=0.9	32	62	38	68	0.38	0.54	0.38	0.32	53
LBP + SVM + CASIA									
>=0.0	100	0	100	0	1	0.50	1	1	50
>=0.1	88	0	100	12	1	0.53	1	0.88	56
>=0.2	72	1	99	28	0.99	0.58	0.99	0.72	64
>=0.3	41	1	99	59	0.99	0.71	0.99	0.41	79
>=0.4	18	2	98	82	0.98	0.84	0.98	0.18	90
>=0.5	9	3	97	91	0.97	0.92	0.97	0.09	94
>=0.6	8	3	97	92	0.97	0.92	0.97	0.08	95
>=0.7	8	28	72	92	0.72	0.90	0.72	0.08	82
>=0.8	5	48	52	95	0.52	0.91	0.52	0.05	74
>=0.9	2	88	22	98	0.20	0.92	0.20	0.02	57
LBP + SVM + UBIRIS									
>=0.0	100	0	100	0	1	0.50	1	1	50
>=0.1	84	0	100	16	1	0.54	1	0.84	58
>=0.2	78	2	98	22	0.98	0.56	0.98	0.78	60
>=0.3	54	3	97	46	0.97	0.64	0.97	0.54	72
>=0.4	24	3	97	76	0.97	0.80	0.97	0.24	87
>=0.5	13	4	96	87	0.96	0.88	0.96	0.13	92
>=0.6	11	16	84	89	0.84	0.88	0.84	0.11	87
>=0.7	11	25	75	89	0.75	0.87	0.75	0.11	82
>=0.8	10	55	45	90	0.45	0.82	0.45	0.10	68
>=0.9	5	70	30	95	0.30	0.86	0.30	0.05	63

The ROC-Curve has been drawn for different combinations of methods and databases. Fig. 4 shows the ROC curves.

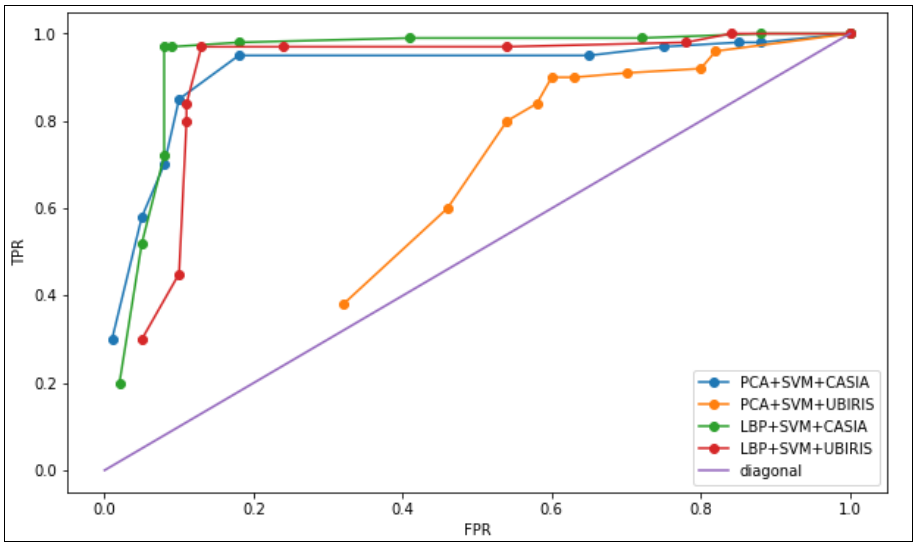


Fig 4: ROC-Curves

The Area under curve (AUC) has been calculated for different combinations of feature extraction methods and classification methods along with two databases listed. The result obtained is presented in Table 2.

Table 2: Area under curve (AUC)

Method	AUC
PCA + SVM + CASIA	0.9125
PCA + SVM + UBIRIS	0.5518
LBP + SVM + CASIA	0.9395
LBP + SVM + UBIRIS	0.8962

The experiment has been conducted using different dataset size. The result obtained is presented in Table 3 and Fig. 5. For different data set size, the accuracy remains high for the combinations PCA + SVM + CASIA, LBP + SVM + CASIA and LBP + SVM + UBIRIS.

Table 3: Accuracy and AUC for different dataset size

Method	Dataset size							
	200		400		600		800	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
PCA + SVM + CASIA	89%	0.9125	88%	0.9025	90%	0.9185	87%	0.8892
PCA + SVM + UBIRIS	65%	0.5518	64%	0.5416	68%	0.5698	58%	0.5123
LBP + SVM + CASIA	95%	0.9395	93%	0.9256	96%	0.9356	93%	0.9256
LBP + SVM + UBIRIS	92%	0.8962	90%	0.8756	91%	0.8562	91%	0.8562

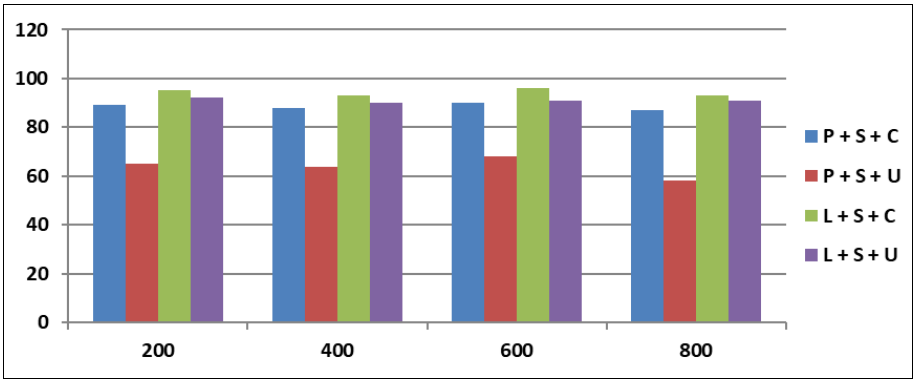


Fig 5: Accuracy graph

Conclusion

Experiment has been conducted to test the performance of PCA and LBP with two popular iris databases CASIA and UBIRIS. The models were tested using 200 iris images. In case of LBP, the AUC values obtained are 0.8962 and 0.9395 for CASIA and UBIRIS datasets respectively. In case of PCA, the AUC values are 0.9125 and 0.5518 for CASIA and UBIRIS datasets. The result shows that LBP achieves better performance with CASIA and UBIRIS compared to PCA. The experiment was extended for different dataset sizes of 400, 600 and 800. The result shows that LBP outperforms PCA with both CASIA and UBIRIS datasets.

References

1. Jain AK, Ross A. Prabhakar S. An introduction to biometric recognition. IEEE Transactions on Circuits and Systems for Video Technology, 1 14, January 2004;4(20):1051-8215.
2. Sharma Abhilash. Biometric System: A Review. International Journal of Computer Science and Information Technologies. 2015;6:4616-4619.
3. Sevugan Prabu, Swarnalatha P, Gopu Magesh, Sundararajan Ravee. Iris recognition system. International Research Journal of Engineering and Technology; c2017.
4. Kak Neha, Rishi Gupta, Sanchit Mahajan. Iris Recognition System. International Journal of Advanced Computer Sciences and Applications, 2010, 1. 10.14569/IJACSA.2010.010106.
5. Richard Yew Fatt Ng, Yong Haur Tay, Kai Ming Mok. Iris recognition algorithms based on texture analysis, 2008 International Symposium on Information Technology, 2008, 1-5. DOI: 10.1109/ITSIM.2008.4631667.
6. Wildes RP. Iris recognition: an emerging biometric technology, Proceeding, pp s of the IEEE. 1997;85(9):1348-1363.
7. Boles Wageeh, Boashash Boualem. A Human Identification Technique Using Images of the Iris and Wavelet Transform. Signal Processing, IEEE Transactions on. 1998;46:1185-1188. 10.1109/78.668573.
8. Yankui Sun, Yong Chen, Hao Feng. two-dimensional stationary dyadic wavelet transform, decimated dyadic discrete wavelet transform and the face recognition application. 2011;9(3):397-416.
9. Xiao Zhou Chen, Chang Yin Wu, Liang Lin Xiong, Fan Yang, The Optimal Matching Algorithm for Multi-Scale Iris Recognition, Energy Procedia, 2012;16:Part B.
10. Hassanein Allam S, *et al*. A Survey on Hough Transform, Theory, Techniques and Applications; c2015. ArXiv abs/1502.02160.
11. Ma Li, Tan Tieniu, Zhang Dai. Efficient Iris Recognition by Characterizing Key Local Variations. IEEE transactions on image processing: A publication of the IEEE Signal Processing Society. 2004;13:739-50. 10.1109/TIP.2004.827237.
12. Sousa Celso. An overview on weight initialization methods for feedforward neural networks; c2016. 10.1109/IJCNN.2016.7727180.
13. Daugman J. How Iris Recognition Works, IEEE Trans. Circuits and Systems for Video Technology. Jan. 2004;14(1):21-30.
14. Conners RW, Harlow CA. A theoretical comparison of texture algorithms. IEEE Trans Pattern Anal Mach Intell. 1980 Mar;2(3):204-22. DOI: 10.1109/tpami.1980.4767008. PMID: 21868894.
15. Hoxha Julian, Stoja Endri, Domnori Elton, Cincotti Gabriella. Multicarrier Digital Fractional Fourier Transform for Coherent Optical Communications;

c2017. 10.1109/EUROCON.2017.8011073.
16. Ranjzad Hamed, Ebrahimi Afshin, Ebrahimnezhad Hossein. Improving feature vectors for iris recognition through design and implementation of new filter bank and locally compound using of PCA and ICA; c2008. 10.1109/ISABEL.2008.4712612.
17. Mishra Sidharth, Sarkar Uttam, Taraphder Subhash, Datta Sanjoy, Swain Devi, Saikhom Reshma, *et al*. Principal Component Analysis. International Journal of Livestock Research, 2017, 1. 10.5455/ijlr.20170415115235.
18. Karamizadeh Sasan, Abdullah Shahidan, Manaf Azizah, Zamani Mazdak, Hooman Alireza. An Overview of Principal Component Analysis. Journal of Signal and Information Processing; c2013. 10.4236/jsip.2013.43B031.
19. Song Ke-Chen, YAN Yun-Hui, CHEN Wen-Hui, Zhang Xu. Research and Perspective on Local Binary Pattern. Acta Automatica Sinica. 2013;39:730–744. 10.1016/S1874-1029(13)60051-8.
20. Huang di, Shan Caifeng, Ardabilian Mohsen, Chen Liming. Local Binary Patterns and Its Application to Facial Image Analysis: A Survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C. 2011;41:765-781. 10.1109/TSMCC.2011.2118750.
21. Ojal T, Pietikinen M, Harwood D. A Comparative study of texture measures with classification based on featured distributions, Pattern Recognition. 1996;29(l):51-59.
22. Srivastava Durgesh, Bhambhu Lekha. Data classification using support vector machine. Journal of Theoretical and Applied Information Technology. 2010;12:1-7.