2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)
21 - 23 Dec 2017, Dhaka, Bangladesh

Enhancing Performance of PCA, ICA through Distribution Transformation

Anu Singha*, Mrinal Kanti Bhowmik
Department of Computer Science and Engineering
Tripura University (A Central University), Agartala, India
*anusingh5012@gmail.com, mrinalkantibhowmik@tripurauniv.in

***Abstract*—Holistic based face recognition methods are generally more effective on normally distributed data matrix of face images. The distribution of data matrix follows the standard Gaussian distribution according to central limit theorem, which has not been seen in practical scenarios. In this context, a simple and effective method called transformation of a data matrix to Gaussian matrix (TDG) is proposed. The TDG transforms an arbitrarily distributed data matrix to a Gaussian distribution. This transformed data matrix is then processed through Principal Component Analysis (PCA) and Independent Component Analysis (ICA). Experiments on a benchmark face database IRIS are demonstrated that the proposed transformation process could notably improve the accuracy rates than state-of-art methods.**

Index Terms*—Gauusian Orthogonal Ensemble (GOE), Gram-Schmidt Orthogonalization, Z-Scores, Q-Q Plot, TDG

I. Introduction

In image analysis, face recognition have been a challenging and quite attractive key area of research which is usually used in security systems and law enforcements. There are number of popular face feature extraction methods has been proposed which is being started with a data matrix that is a concatenation of 1D image column vectors. The most noticeable methods among these are Principle Component Analysis (PCA) [1], and a generalized version of PCA i.e. Independent Component Analysis (ICA) [2][3]. PCA or ICA are more effective when the original image data matrix to be close to Gaussian distribution [2][4][5]. According to central limit theorem (CTL) [6], a several independent random variables are mixed together in an additive fashion, and then resulting distribution is usually normal or Gaussian. But the CTL states that in many situations, the sample mean vary normally if only the sample size is reasonably large. Since normality of data is one key assumption of many statistical issues, transformation of non-normal data to normal data or near normal becomes important.

Some of the holistic feature extraction (HFE) methods data matrix assumes that the data matrix is normally distributed. This is because the normal distribution has some special properties which may have more effective for HFE methods. If the assumption of normality is not satisfied, then the results of the statistical analysis may be incorrect. If the data is close to being normal, we may still assume it is normal and for mild departures from normality, the analysis may still give good results. So, the distribution issue is very valuable. Recently, X. Shi *et al* [7] tried to reduce the influence of grosses like variations in lighting, facial expressions and occlusions to improve the robustness of PCA. They present a simple unsupervised pre-processing method, two-dimensional whitening reconstruction (TWR) where whitening processed has conceded on a 2D face matrix rather than concatenated 1D vector. The resultant face image with TWR pre-processing could be approximate to a Gaussian signal. L.Z. Liao *et al* [8] develop a whitening method where the whitening parameter in the whitening filter is adaptive to the input data so as to make the power spectrum of the whitened results as flat as possible.

There are some transformations that have been used to make the transformed data normal. For example, in some cases, the square or square root or logarithm or reciprocal of the data values may be normally distributed. Such simple transformations of the data to make the data normal can be grouped together under a transformation called the Box-Cox transformation [9]. The inconvenience of Box-Cox like log transformation works with only non-negative numbers. One can always transform the variable to reflect non-negative numbers but we won't able to anticipate new incoming data and perhaps, the data may still contain negative numbers. Another well known transformation called Johnson transformation [10], is very powerful and it can be used with data that include zero and negative values. The computational cost of transforming variables using Johnson transformation is much higher than Box-Cox transformation due to presence of higher number parameters that needs to be optimized. Both Box-Cox and Johnson transformation, a computer can run through several combinations of transformation parameters to determine which set of parameter(s) makes the transformed data as close to normal as possible. So, both of these transformation are basically parameter based. Another linearly transformation data values i.e. z-scores [11] giving scores a common standard of zero mean and unity standard deviation facilitates their interpretation. The people relate that standardizing scores (z-scores) change the shape of their distribution in any way become more or less normal.

The paper is structured as follows: Section II mainly illustrates and analyzes the proposed method TDG, which transform the arbitrarily distributed image data matrix to a Gaussian distribution. Section III discusses the experimental results on face databases. Finally, Section IV gives the conclusion.

978-1-5386-2175-2/17/$31.00 ©2017 IEEE

II. Transformation of a Data Matrix to Gaussian (TDG) Process and Analysis

When we generate a sample data from face images, sometimes we want to use that data to make predictions about the process and the images. If the acquisition comes from a controlled environment that follow some physical nature, then the data from that process may follow a precise distribution otherwise the distribution may be random. The shape of a distribution provides clues to the process that generates the data, and can be used to draw some interference about its application. Motivated by these findings, we propose a transforming method which transforms the image data matrix to a Gaussian distribution.

A. Transformation of a Data Matrix to Gaussian (TDG)

Algorithm 1 briefly illustrates the steps of TDG method. Give a set of 2D training images $y^1, y^2, ..., y^n \in \Re^{p \times q}$, and then convert each 2D training image into column vector, say, $x_i = [y^i_{11}, y^i_{21}, ..., y^i_{pq}]^T \in \Re^{m \times 1}$ where $m = p \times q$, let data matrix $X = [x_1, x_2, ..., x_n] \in \Re^{m \times n}$.

The Gaussian orthogonal matrix distribution or Gaussian Orthogonal Ensemble (GOE) was one of the three Gaussian matrix ensembles originally suggested by Eugene Winger as a tool to study fluctuations in nuclear physics [12][13]. Matrix ensembles like the GOE are of considerable importance in the study of random matrix theory, as well as in various branches of physics and mathematics. From GOE, $Distributed[A, GaussianOrthogonalMatrixDistribution(\sigma, n)]$, written more concisely as $A \approx GaussianOrthogonalMatrixDistribution(\sigma, n)$ can be used to assert that a random squared symmetric matrices $A = \{a_{ij}\}$ are distributed according to a Gaussian Orthogonal Matrix Distribution have probability densities proportional to Normal Distribution $(0, \sigma)$. We know, the expression form of a symmetric matrix A is $A = QDQ^T$ where Q is an orthogonal matrix, and D is a diagonal matrix of eigenvalue elements of A. According to GOE, each element of symmetric matrix follows a distribution of Gaussian where each element is coming from an orthogonal basis space of data matrix X.

In this regards, with the help of Gram-Schmidt process [14], if $\beta = \{x_1, x_2, \ldots, x_n\} \in \Re^{m \times n}$ is a basis for the vector space Q, then the vectors

$$q_1 = x_1$$
$$q_2 = x_2 - (\frac{< x_2, q_1 >}{\|q_1\|}).q_1$$
$$q_3 = x_3 - (\frac{< x_3, q_1 >}{\|q_1\|}).q_1 - (\frac{< x_3, q_2 >}{\|q_2\|}).q_2$$
$$\ldots$$
$$q_n = x_n - (\frac{< x_n, q_1 >}{\|q_1\|}).q_1 - (\frac{< x_n, q_2 >}{\|q_2\|}).q_2 - \cdots - (\frac{< x_n, q_{n-1} >}{\|q_{n-1}\|}).q_{n-1} \quad (1)$$

Algorithm 1. Transformation of Data Matrix to a Gaussian (TDG)

Input: An image data matrix $X = [x_1, x_2, \ldots, x_n] \in \Re^{m \times n}$. There are total of *n* images consist *m* number of real valued pixels.

1. Let, a basis vector space for Q is $\beta = \{x_1, x_2, ..., x_n\} \in \Re^{m \times n}$ where each x_i is a image column vector of length m.
2. Form an orthogonal basis $Q = \{q_1, q_2, q_3, \ldots, q_n\}$ through Gram-Schmidt Orthogonalization which satisfy the following:
   (a) $\|q_i\| = 1$ for $1 \le i \le n$
   (b) $< q_i, q_j >= 0$ for $1 \le i, j \le n, i \ne j$
   (c) $Q^T Q = I$
3. Calculate the z-scores of data matrix of orthonormal vectors space Q as follows:
   $$\tilde{Q} = \frac{Q(1:m,:) - \mu(1:m)}{\sigma(1:m)}$$
4. Finally, the feature extraction methods like PCA, ICA are applied on this newly transformed data matrix $\tilde{Q}$.

Output: Transformed Data Matrix $\tilde{Q} \in \Re^{m \times n}$.

---

form an orthogonal basis $Q = \{q_1, q_2, q_3, ..., q_n\}$ satisfying the following (a) $\|q_i\| = 1$ for $1 \le i \le n$ (b) $< q_i, q_j >= 0$ for $1 \le i, j \le n, i \ne j$ (c) $Q^T Q = I$. In each sequence, the basis vector will be subtracted from the normalized relationships i.e. inner product between current basis vectors and previously formed orthogonal vectors.

The distribution of this orthonormal vectors space $Q \in \Re^{m \times n}$ is approximately proportional to less or more Gaussian distribution. We know, if a variable is roughly normally distributed, z-scores will follow near to a standard normal distribution. So, we are then calculated the z-scores of data matrix of orthonormal vectors space Q as follows:

$$\tilde{Q} = \frac{Q(1:m,:) - \mu(1:m)}{\sigma(1:m)} \quad (2)$$

i.e. subtracting the mean $\mu(1:m) = \sum_{j=1}^{n} Q_j \Big/ n$ over all scores from each individual score, and then dividing by the standard deviation $\sigma(1:m) = \sqrt{\sum_{j=1}^{n} (Q_j - \mu)^2 \Big/ n}$ over all scores. Z-scores are linearly transformed data values having a mean of zero and standard deviation of 1 i.e. they are scores (or data values) that follow a standard Gaussian distribution.

B. Analysis of Data Matrix

In order to better display the influence of TDG on face data matrix distribution, we present a statistical graphical test called Quantile-Quantile (Q-Q) plot [15] which is to help us assess if a set of data plausibly came from some theoretical distribution such as Gaussian. If we run a statistical analysis that assumes our data matrix is normally distributed, we can use standard normal Q-Q plot to check that assumption. It is just a visual check, allows us to see at-a-glance if our assumption is plausible. A Q-Q plot is a scatter plot created by plotting two sets quantiles against one another. If both sets of

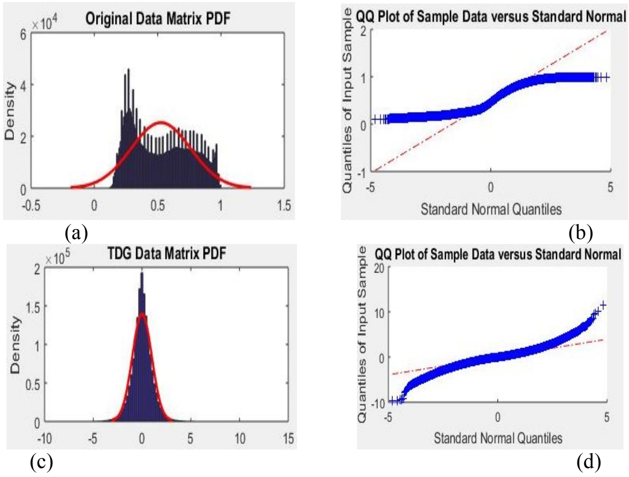

Fig. 1. A sample of pixel value distribution of the original data matrix and its transformed data matrix.

quantiles came from the same distribution, we should see the points forming a line that's roughly straight.

We have conducted an analysis from Imaging, Robotics, and Intelligent Systems (IRIS) [16] visual expression dataset. From the dataset, a data matrix is generated by combining 660 face images of 30 individuals as each column of data matrix represents a face image. In Fig. 1a, the curve indicates density of a normal distribution and that histogram area is not fitted properly to normal curve. It becomes clearer in Fig. 1b, when we plot out the Q-Q of original sample data vs. standard normal. From the points of both sample data matrix and standard normal, we see that the points are not forming a straight line. That means original data matrix is not a normally distributed on which holistic feature extraction methods like PCA, ICA are not more effective. In Fig. 1c, we have shown the Probability Density of transformed (TDG) data values where noticed that histogram area is almost laid down with classic symmetrical bell curve standard normal distribution with a mean of 0. Fig. 1d Q-Q plot also shows that the points of TDG data and standard normal forming a line that's roughly straight. So, the TDG can transfer original data matrix to be a signal close to normal distribution.

III. Performance Evaluation and Analysis

A well-known benchmark face database, IRIS [16] is used to test the proposed method TDG on PCA and ICA. From the database, experiments are considered on visual faces of expression and illumination datasets which are captured under various expressions, illuminations, and poses. For our experiment, we have collected 3 different expressions Exp1 (Surprised), Exp2 (Laughing), Exp3 (Anger), and 4 different illumination conditions Lon (left light on), Ron (right light on), 2on (both lights on), off (left and right lights off) with varying poses. The expression dataset reposes of total 660 images of 20 individual persons and there are 33 different images with 3 different expressions. Similarly, the illumination dataset reposes of 748 images of 17 individual persons and there are 44 different images with 4 illuminations. To arrange the database according to our experiment, we have first created 3 sets for expression and 10 sets for illumination as shown in second, third and fourth columns of Table I. The 660 images of expression is arranged as for instance, in Set 1, *Exp 1* and *Exp 2* images i.e. 20x22=440 images are separated for training purpose and the rest of the images in *Exp 3* i.e. 20x11=220 images are kept as testing images. In case of illumination, the 748 images of four (4) different illuminations are organized in ten (10) different sets for training and testing purpose. For example, for Set 1, *Lon*, *Ron* and *2on* images i.e. 17x33=561 images are separated as training images and the rest of the images in *off* i.e. 17x 11 =187 images are kept as testing images. In this way up to Set 4 but alternative illuminations for training and testing images has been organized. Again, for Set 5, *Lon* and *Ron* images i.e. 17x22=374 are grouped as training set and *2on*, *off* images i.e. 17x22=374 images are grouped as testing set, and the same way will continued for another 5 sets i.e. up to Set 10.

TABLE I. Classification Results on IRIS Visual Database

Dataset	Training Sample	Number of Training Images	Number of Testing Images	ICA	TDG-ICA	PCA	TDG-PCA
Expression	**Set 1**	Surprised, Laughing (20 x 22 = 440)	Anger (20 x 11 = 220)	75.45	76.36	78.63	80
	Set 2	Laughing, Anger (20 x 22 = 440)	Surprised (20 x 11 = 220)	75.45	83.18	75.45	84.54
	Set 3	Surprised, Anger (20 x 22 = 440)	Laughing (20 x 11 = 220)	77.72	79.54	85.45	87.27
			Average	**76.21**	**79.69**	**79.84**	**83.93**
Illumination	**Set 1**	Lon, Ron, 2on (17 x 33 = 561)	Off (17 x 11 = 187)	72.72	76.47	85.02	86.09
	Set 2	Lon, Ron, Off (17 x 33 = 561)	2on (17 x 11 = 187)	89.83	89.30	93.58	94.11
	Set 3	Lon, 2on, Off (17 x 33 = 561)	Ron (17 x 11 = 187)	88.23	91.44	95.72	97.32
	Set 4	Ron, 2on, Off (17 x 33 = 561)	Lon (17 x 11 = 187)	89.30	89.30	95.18	94.65
	Set 5	Lon, Ron (17 x 22 = 374)	2on, Off (17 x 22 = 374)	79.67	82.08	87.16	87.43
	Set 6	Lon, 2on (17 x 22 = 374)	Ron, Off (17 x 22 = 374)	75.93	78.07	81.55	83.68
	Set 7	Lon, Off (17 x 22 = 374)	Ron, 2on (17 x 22 = 374)	77.80	79.14	79.41	87.43
	Set 8	Ron, 2on (17 x 22 = 374)	Lon, Off (17 x 22 = 374)	72.72	74.59	83.42	85.02
	Set 9	Ron, Off (17 x 22 = 374)	Lon, 2on (17 x 22 = 374)	84.49	83.95	85.82	90.37
	Set 10	2on, Off (17 x 22 = 374)	Lon, Ron (17 x 22 = 374)	83.24	87.5	94.14	90.42
			Average	**81.39**	**83.18**	**88.10**	**89.65**

A. Components Selection

Liu [17] has shown that the selection of number of principal components (PCs) has a significant effect on the performance of PCA/ICA based face recognition. In the selection of PCs approach, it is not always that large number of PCs will always give better performances. Sometimes, only few numbers has effected more whereas increasing number of PCs is decreasing the recognition performances. In the experiments, an optimal number of 60 and 80 PCs/ICs for PCA/ICA on expression and illumination datasets are chosen respectively. Here, FastICA [18] with a contrast function was $G(x) = -\exp(x^2/2)$ used in case of ICA.

B. Performance Evaluation and Analysis

To clearly show the performance of TDG on state-of-art methods, we present two popular face recognition methods Principal Component Analysis (PCA) and Independent Component Analysis (ICA). To demonstrate that TDG is effective for both state-of-art methods, we have evaluated the effect of recognition performance through a classification technique i.e. Random Forest [19] is configured to have 50 decision trees.

Table I lists the average recognition rates across several training sample sets for each method. There are few points to be taken from the table, in case of expression dataset. Most of the sets of transformed data are getting approximately 1% to 2% better accuracy but in set 2, about 8% more of both PCA and ICA. In terms of average recognition rate, a total of 3% and 4% differences are showing between ICA and TDG-ICA, and PCA and TDG-PCA respectively. On the other hand i.e. illumination dataset, all sets performance differences vary up to 4% respectively. In few sets, there is no significance performances, even slightly less than in sets 2, 9 for ICA, TDG-ICA and sets 4, 10 for PCA, TDG-PCA. But at total average, we are getting satisfactory results.

To show the comparative performances of TDG, we also present three Gaussian transformation methods Box-Cox [9], johnson [10] and Z-Score [11] in our experiments. Table II. shows the comparative results on IRIS database over number of PCs/ICs 100, 120 and 140 respectively. TDG method gets more improvement than Box-Cox, Johnson, or Z-Score, but obtains few exceptions in case of ICA on expression dataset.

IV. CONCLUSION

TABLE II. COMPARATIVE RESULTS ON IRIS VISUAL DATABASE

Method	Expression			Illumination		
	PCs/ICs = 100	PCs/ICs = 120	PCs/ICs = 140	PCs/ICs = 100	PCs/ICs = 120	PCs/ICs = 140
TDG + PCA	85.15	84.24	83.48	88.29	89.38	89.20
BoxCox + PCA	78.33	79.84	78.63	87.67	88.21	85.54
Johnson + PCA	77.27	79.84	77.42	78.91	79.71	81.37
Z-Score + PCA	81.06	79.24	77.87	88.15	86.26	86.82
TDG + ICA	72.15	71.81	73.33	87.17	85.08	86.34
BoxCox + ICA	**72.87**	65.15	63.93	82.62	81.95	78.51
Johnson + ICA	**73.63**	**71.84**	72.54	77.39	82.46	81.90
Z-Score + ICA	67.72	66.51	62.42	81.96	80.30	78.96

In this paper, we present a transformation method called TDG. To draw the influence of TDG, we present a Q-Q statistical test. Finally, classification experiments on visual IRIS database demonstrate that TDG based state-of-art methods could greatly improve the performances than other related methods.

Acknowledgment

The work presented here is being conducted in the Biometrics Laboratory of Department of Computer Science and Engineering of Tripura University (A Central University), Tripura, Suryamaninagar-799022.

References

[1] L. Sirovich, and M. Kirby, "Low-dimensional procedure for characterization of human faces," Jour. Opt. Soc. Am., Vol. 4, No. 3, pp. 519–524, 1987.

[2] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent component representation for face recognition," in Proc. SPIE Symp. Electron. Imaging: Sci. Technol., pp. 528–539, 1998.

[3] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," IEEE Trans. Neural Netw., Vol. 13, No. 6, pp. 1450–1464, 2002.

[4] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," Jour. ACM, Vol. 58, No. 3, Art. 11, 2011.

[5] I. Joliffe, "Principal Component Analysis," New York, NY, USA: Springer-Verlag, 1986.

[6] J. A. Rice, "Mathematical Statistics and Data Analysis (2nd ed.)," Duxbury Press., 1995.

[7] X. Shi, Z. Guo, F. Nie, L. Yang, J. You, and D. Tao, "Two-Dimentional Whitening Reconstruction for Enhancing Robustness of Principal Component Analysis," IEEE Trans. on Patt. Analy. and Mach. Intell., Vol. 38, No. 10, pp. 2130-2135, 2016.

[8] L.Z. Liao, S.W. Luo, and M. Tian, " 'Whitenedfaces' Recognition With PCA and ICA," IEEE Signal Processing Letters, Vol. 14, No. 12, 2007.

[9] G. E. P. Box, and D. R. Cox, "An analysis of transformations," Jour. Royal Statistical Society, Series B," Vol. 26, No. 2, pp. 211–252, 1964.

[10] N. L. Johnson, "Systems of Frequency Curves Generated by Methods of Translation," Biometrika, Vol. 36, No. 1/2, pp. 149–176, 1949.

[11] E. Kreyszig, "Advanced Engineering Mathematics (Fourth ed.)," Wiley. pp. 880, eq. 5. 1979.

[12] E. Wigner, "On the distribution of the roots of certain symmetric matrices, " The Annals of Mathematics Vol. 67, pp. 325-327, 1958.

[13] M. Chiani, "Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy–Widom distribution," Jour. of Multivariate Analysis, Vol. 129, pp. 69-81, 2014.

[14] W. Cheney, and D. Kincaid, "Linear Algebra: Theory and Applications," Sudbury, Ma: Jones and Bartlett. pp. 544, 558, 2009.

[15] M. B. Wilk, and R. Gnanadesikan, "Probability plotting methods for the analysis of data," Biometrika, Biometrika Trust, Vol. 55, No. 1, pp. 1–17, 1968.

[16] http://www.cse.ohio-state.edu/otcbvs-bench/Data/02/download.html.

[17] C. Liu, "Enhanced independent component analysis and its application to content based face image retrieval," IEEE Trans. Syst. Man Cybern. B Cybern. Vol. 34, pp. 1117–1127, 2004.

[18] A. Hyvarinen, E. Oja, "Independent component analysis: algorithms and applications," Neural Netw. Vol. 13, pp. 411–430, 2000.

[19] T. K. Ho, " The Random Subspace Method for Constructing Decision Forests," IEEE Trans. on Patt. Analy. and Mach. Intell., Vol. 20, No.8, pp. 832–844, 1998.