Intelligent System for Speaker Identification using Lip features with PCA and ICA

Anuj Mehra Anupam Shukla Mahender Kumawat Rajiv Ranjan Ritu Tiwari

Abstract—Biometric authentication techniques are more consistent and efficient than conventional authentication techniques and can be used in monitoring, transaction authentication, information retrieval, access control, forensics, etc. In this paper, we have presented a detailed comparative analysis between Principle Component Analysis (PCA) and Independent Component Analysis (ICA) which are used for feature extraction on the basis of different Artificial Neural Network (ANN) such as Back Propagation (BP), Radial Basis Function (RBF) and Learning Vector Quantization (LVQ). In this paper, we have chosen “TULIPS1 database, (Movellan, 1995)” which is a small audiovisual database of 12 subjects saying the first 4 digits in English for the incorporation of above methods. The six geometric lip features i.e. height of the outer corners of the mouth, width of the outer corners of the mouth, height of the inner corners of the mouth, width of the inner corners of the mouth, height of the upper lip, and height of the lower lip which extracts the identity relevant information are considered for the research work. After the comprehensive analysis and evaluation a maximum of 91.07% accuracy in speaker recognition is achieved using PCA and RBF and 87.36% accuracy is achieved using ICA and RBF. Speaker identification has a wide scope of applications such as access control, monitoring, transaction authentication, information retrieval, forensics, etc.

Keywords—Biometric authentication; Intelligent System; Lip Features; Independent Component Analysis (ICA); Principal Component Analysis (PCA); Back Propagation (BP); Radial Basis Function (RBF); Learning Vector Quantization (LVQ).

1 INTRODUCTION

Speaker Identification can be introduced to make system/services accessed only by the authorized personnel. The possible applications could vary from retrieval of private information to control of financial transactions, access for the authorized persons in a building and so on [2]. User could be authenticated by old conventional techniques (password/personal identification number, token/card) or by its own biometric characteristic (lips, face, iris, fingerprint, voice, etc.). Previous researchers have proposed different algorithms that automatically extract lip features from speaker images [9]. Previous researches have also shown the different proposed techniques using lips feature to authenticate the speaker. Previous work has been done on the incorporation of Active Shape Model (ASM) [7] features and Hidden Markov Model (HMM). The approaches mentioned above are Brown et al. [3] which extracts the identity relevant information by the usage of six geometric lip features and two inner mouth features, Wark and Sridharan [4] that combines the normal from profiles to the contour points to form the grand profile vector (GPV) for the image, Luettin et al. [5] which extracts the Active Shape Model (ASM) features, and uses ASM to model the lip shape and the intensity profile vector along normal to the model points, Matthews et al. [6] which were proposed due to the intensity variation inside the outer lip contour. The main aim of this research is to make a comparative analysis between the techniques PCA and ICA which are used for extraction of lip features on the basis of BP, LVQ, and RBF. In this paper, we have used the six geometric lip features used for the speaker identification. PCA is an approach which uses lower dimensional feature vectors to approximate the original data. PCA attempts to minimize the reconstruction error under the restrictions such as linear reconstruction, and orthogonal factors and maximize the variance [12, 15]. ICA is a technique that seeks statistically independent and non-Gaussian components to find the multi-dimensional data and their underlying factors. It is basically the recognition & separation of mixtures of sources with little prior information. [17]. ICA can be used in a diversified range of applications such as Medical signal processing, Feature extraction, face recognition, Watermarking, Clustering, Time series analysis, and many more applications [16]. Previous researches have been accomplished by integrating Hidden Markov Model (HMM) [8], as compared to our research which uses PCA [12, 13] and ICA [16, 17] for feature extraction and
BP, RBF and LVQ as neural network classifiers. Section 2 explains the methodology used in the research whereas Section 3 explains the experiments and the results, and Section 4 elucidate about Conclusion and the future scope. The block diagram of speaker identification process in this work is shown in Fig. 1.

```
“TULIPS1 database” [1]

Training dataset       Testing dataset

Feature extraction     Feature extraction

Applying Different ANN-BP, RBF, LVQ

Simulating Network and calculating Accuracy

Desired result is obtained-91.07% using PCA and RBF and 87.36% using ICA

Fig. 1. Block Diagram for Speaker Identification.
```

### 2 METHODOLOGY

Both PCA and ICA are used as a preprocessing technique for the lip images. PCA is a technique which produces feature vectors in a reduced dimension whereas ICA is a technique that seeks statistically independent and non-Gaussian components to find the multi-dimensional data and their underlying factors. The feature vectors obtained in both the above mentioned processes are then used as an input data to the neural network classifiers i.e. Back Propagation Multi-layer neural network structure (BP-MLNN), Radial basis function (RBF) and Learning Vector quantization (LVQ).

#### 2.1. PCA PREPROCESSING

PCA provides an effective technique for dimensionality reduction [12, 15]. The procedure is to first compute the eigen values and eigenvectors from the covariance matrix of the original input data vector. It then computes the orthonormal vectors which are the basis of the computed data. These are the unit vectors and are referred to as the Principle Components. These principle components serve as the new set of axes for the input data.

Let \( X = (x_1, x_2, ..., x_n) \) represents the \( n \times N \) data matrix, where each \( x_i \) is a lip vector of dimension \( n \), concatenated from a \( p \times q \) lip image.

\[
Y = W^T x
\]

(1)

Where \( Y \) is the \( m \times N \) feature vector matrix, \( m \) is the dimension of the feature vector and transformation matrix \( W \) is an \( n \times m \) transformation matrix whose columns are the eigenvectors corresponding to the \( m \) largest eigen values computed according to the formula

The principle component \( w_i \) of the data set \( X \) is as follows:

\[
w_i = \arg \max_{|w| = 1} \text{var}(W^T x)
\]

(2)

#### 2.2. ICA PREPROCESSING

Independent component analysis (ICA) [16, 17] is a method that seeks statistically independent and non-Gaussian components which makes it different from the other feature extraction methods. It is a method to find the multivariate or multi-dimensional data and their underlying factors. A broad idea of ICA is shown in Fig. 2.

```
ICA — Non-Gaussian

No (noise) model

Projection

Gaussian data

PCA

Fig. 2. ICA model and related methods.
```

ICA model can be explained in the given below equation,

\[
X = AS
\]

(3)
where, $x$ and $s$ be two random vectors representing $x_1, x_2, x_3, ..., x_n$ and $s_1, s_2, s_3, ..., s_m$ respectively and $A$ represents the linear static transformation with elements $a_{ij}$. Thus, $x^T$ is a row vector. The above equation can be rewritten as

$$x = \sum_{i=1}^{n} a_i s_i$$  \hspace{1cm} (4)

It is also known as blind source separation. For simplicity, we have used FastICA algorithm and have converted the noisy database into noise-free by the whitening process as it is useful for our research work.

### 2.3. BACK PROPAGATION MULTI-LAYER NEURAL NETWORKS STRUCTURE

In back propagation the partial derivatives of the cost function with respect to the free parameters of the network are being determined by back propagation error signals through the network, layer by layer. The feed forward multi-layer neural networks are used as a classifier for the PCA output data. The employed neural network is a feed forward multilayer neural network hidden layer as given in the following Fig.3.

![Fig. 3. Back Propagation Multi-layer neural network structure.](image)

In Fig. 3, we have considered a training data which consists of the pairs $(x, y)$ where vector $x$ represents a pattern of input to the network, and the vector $y$ the corresponding desired output (target). According to the error back propagation algorithm, the gradient $\Delta W_{ij}$ can be computed by,

$$\Delta W_{ij} = -\frac{\partial E}{\partial net_i} \frac{\partial net_i}{\partial W_{ij}}$$  \hspace{1cm} (5)

where, weight from unit $j$ to unit $i$ is represented by $w_{ij}$, $E$ is the output error, first factor represents the error of unit $i$, and the second factor represents the activity of the output unit $y$. The activity of the input units $y_i$ is calculated by the network’s external input $x$. The equation of error back propagation

$$\delta_j = -\sum_{i \in P_j} \frac{\partial E}{\partial net_i} \frac{\partial net_i}{\partial y_j} \frac{\partial net_j}{\partial y_j}$$  \hspace{1cm} (7)

where, first factor represents the error of node $i$, second factor is the weight from unit $j$ to $i$, and third factor is the derivative of node $j$’s activation function.

Back Propagation networks frequently use one or more hidden layers of sigmoid neurons and output layer of linear neurons [12].

### 2.4. RADIAL BASIS FUNCTION NETWORKS

These networks find the input to output map using local approximators and are extremely fast and require fewer training samples. It has a static Gaussian function as the nonlinearity for the hidden layer processing elements which strictly responds to a small region of the input space where the Gaussian is centered. An unsupervised approach to find the centre of Gaussian function instead of the supervised approach leads us to the better results.

The simulation begins with the training of the unsupervised layer which derives the Gaussian centres and the width from the input data. The output layer is being derived from the input data weighted by a given Gaussian mixture. As the training is completed the supervised segments sets the centres of the Gaussian function and determines the width of each Gaussian. A typical RBF [13, 14] neural network structure is shown in Fig. 4.

![Fig. 4. RBF network structure with one output.](image)
In the network as shown in Fig. 4, \( n \) -dimensional input feature vector are accepted to the input layer which is a set of \( n \) units consisting of input vectors \( x_1, x_2, \ldots, x_n \) and are input to the \( l \) hidden functions. The output of the hidden function is then multiplied by the weighting factor \( w(i, j) \) which is the input to the output layer of the network \( y(x) \).

\[
y(x) = \sum_{i=1}^{N} w_i \Phi(\|x - c_i\|)
\]

where the approximating function \( y(x) \) is represented as a sum of \( N \) radial basis functions, each associated with a different centre \( c_i \) and weighted by an appropriate coefficient \( w_i \) and \( \| \| \) indicates the Euclidean norm on the input space.

### 2.5. LEARNING VECTOR QUANTIZATION NETWORKS

LVQ is a prototype-based supervised classification algorithm. LVQ [10] is a supervised learning technique that uses class information to move Voronoi vectors [13] to some extent, so as to improve the quality of classifier decision regions.

\[
w_c(n+1) = w_c(n) + \alpha \cdot [x - w_c(n)]
\]

where, \( \alpha = 1 \).

\[
w_c(n+1) = w_c(n) - \alpha \cdot [x - w_c(n)]
\]

where, \( \alpha \) is the learning constant with the number of iterations \( n \).

(b) The other Voronoi vectors are not modified.

As shown in Fig. 6, an LVQ network has a first competitive layer and a second linear layer. We have considered a set of \( n \) input vectors \( x_1, x_2, \ldots, x_n \) which are multiplied to the weights \( w_i \) to compute the output. The weight vector of the competitive layer is shown by the equation (10) where \( w_c(n+1) \) the adjusted weight vector is after adjustment.

### 3 EXPERIMENT AND RESULTS

In this research work, we study and explore the use of neural networks in the speaker using lip features. First of all, the database named “TULIPS1 database, (Movellan, 1995)” (shown in Fig. 7) is selected. The database consists of 12 subjects saying the first 4 digits in English and the digits are repeated twice by each speaker. Each digit spoken constitues of 6 instances and the size of each image is 100 by 75 pixels. In the second step, the database is split into two datasets i.e. training and testing dataset. In this experiment, 7 subjects are chosen for the research work constituting in the formulation of 7 classes with 24 images present in each class i.e. 168 images are selected for training and 168 images are selected for testing dataset. In the next step,
PCA and ICA are applied on training dataset for feature extraction followed by the neural network classifiers (BP, RBF, and LVQ). An output matrix is achieved which consists of the reduced set of images. A target matrix is formed and testing images are trained using PCA and ICA followed by the simulation of the neural networks. In the last step, accuracy is calculated by the number of images of training and testing dataset matched.

As shown in Fig. 8, H1 represents height of the outer corners of the mouth, W1 represents width of the outer corners of the mouth, H2 represents height of the inner corners of the mouth, W2 represents width of the inner corners of the mouth, H3 represents height of the upper lip, and H4 represents height of the lower lip.

### 3.1. CLASSIFICATION BY THE BACK PROPAGATION NETWORK

PCA and ICA along with BP is applied to these lip images for training and then it is tested with rest of the images. Here, two hidden layers are used in BP.

### 3.2. CLASSIFICATION BY THE RADIAL BASIS FUNCTION NETWORK

PCA and ICA along with RBF is applied to these lip images for training and then it is tested with rest of the images. In RBF a hidden radial basis layer & output linear layer is used.
3.3. CLASSIFICATION BY THE LEARNING VECTOR QUANTIZATION NETWORK

PCA and ICA along with LVQ is applied to these lip images for training and then it is tested with rest of the images. In LVQ a hidden competitive layer and an output linear layer is used.

3.4. EXPERIMENTAL RESULTS

The results for the recognition test with BP, RBF & LVQ with PCA and ICA are shown in Table IV:

| Methods          | BP     | RBF    | LVQ    |
|------------------|--------|--------|--------|
| Recognition Rate with PCA | 89.88\% | 91.07\% | 87.5\% |
| Recognition Rate with ICA  | 85.8\%  | 87.36\% | 84.52\% |
Result shows that the recognition performance for PCA with RBF is better than the other methods used.

4 CONCLUSIONS

It can be concluded from the experimental result section that RBF overrules BP and LVQ when used with PCA and ICA. It can also be concluded that BP network when used either with PCA or ICA achieved better accuracy than LVQ due to its back propagation mechanism. When PCA and ICA are compared with each other it can be seen from the result section that PCA overrules ICA in all the neural network classifiers used. The second conclusion can be made with respect to computational time; PCA took much lesser time than ICA as the former one does not incorporates whitening process as used in the latter one. Also, RBF took lesser computational time than BP and LVQ in both the cases due to the fact that it employs local approximators and hence require fewer training samples. Hence, RBF in lip recognition is fast and robust than the above methods. In future, the work can be done on an identification system that incorporates the features of both lips and speech. Accuracy can be improved by the use of other feature extraction techniques. Accuracy can also be improvised by incorporating several feature extraction techniques to form a unique one.

REFERENCES

[1] “TULIPS1 database, (Movellan J. R. (1995)” Movellan J. R. (1995) Visual Speech Recognition with Stochastic Networks. in G. Tesauro, D. Toruńczyk, & T. Leen (eds.) Advances in Neural Information Processing Systems, Vol 7, MIT Press, Cambridge.

[2] D.A. Reynolds and L.P. Heck, “Automatic Speaker Recognition Recent Progress, Current Application, and Future Trends”, AAAS 2000 Meeting Humans, Computers and Speech Symposium, 19 February 2000.

[3] C. C. Brown, X. Zhang, R. M. Mersereau and M. Clements, “Automatic speech reading with application to speaker verification”, Proc. 2002 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’02), vol.1, pp. 685-688, Orlando, USA, May 2002.

[4] T. Wark and S. Sridharan, “A Syntactic Approach to Automatic Lip Feature Extraction for Speaker Identification”, Proc. 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’98), vol.6, pp. 3693-3696, Seattle, USA, May 1998.

[5] J. Luettin, N. A. Thacker and S. W. Beet, “Learning to recognise talking faces”, Proc. 13th International Conference on Pattern Recognition, vol.4, pp. 55-59, Vienna, Austria, Aug. 1996.

[6] I. Matthews, T. F. Cootes, J. A. Bangham, S. Cox, R. Harvey, “Extraction of visual features for lip reading”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.24, issue 2, pp. 198-213, Feb. 2002.

[7] L. L. Lau, W. H. Lau, S. H. Leung, S. L. Wang and H. Yan, “Person Authentication Using ASM Based Lip Shape and Intensity Information”, Proc. 2004 IEEE International Conference on Image Processing (ICIP 2004), pp 561-564, Singapore, Oct. 2004.

[8] S.L. Wang and A. W. C. Liew, “ICA-Based Lip Feature Representation For Speaker Authentication”, Proc. 2008 Third International IEEE Conference on Signal-Image technologies and Internet-Based System Pages(s): 763-767, 16-18 Dec. 2007.

[9] K.L. Sum, WH. Lau, S.H. Leung, Alan WC. Liew and K. W Tse , “A New Optimization Procedure for extracting the point -based lip contour using active shape model”, Proc.2001 IEEE International Conference on Acoustics, Speech, and Signal Processing-Volume 03, Pages 1485-1488, 2001.

[10] Guangming Dong, Jin Chen, Xuanxyet Li, Zuguoi Ning, Dongsheng Wang, and Xiongxian Wang, “Global-Based Structure Damage Detection Using LVQ Neural Network and Bispectrum Analysis”, Proc.2005 Springer-Verlag pp. 531 537, Berlin Heidelberg 2005.

[11] Jaakko Ho I lmen , Volker Tresp and Olli Simula, “A Learning Vector Quantization Algorithm For Probabilistic Models”, Proc.2000 European Signal Processing Conference, Volume II, pp.721-724, 2000.

[12] Hui Kong, Xuchun Li, Lei Wang, Earm Khwang Teoh, Jian-Gang Wang, Vengateswarlu, R “Generalized 2D principal component analysis”, Proc. 2005 IEEE International Joint Conference on Volume 1, Aug. 2005.

[13] Haykin, S. (1994), Neural Networks: A Comprehensive Foundation , Upper Saddle River, NJ : Prentice Hall.

[14] Harry Wechsler, Vishal Kakad, Jeffrey Huang, Srinivas Gutta, V. Chen, “ Automatic Video-based Person Authentication Using the RBF Network” First International Conference on Audio- and Video-Based Biometric Person Authentication, pages 85-92, 1997.

[15] Xiaopeng Hong, Hongxun Yao, Yuqi Wan, Rong Chen , “A PCA Based Visual DCT Feature Extraction Method for Lip-Reading” Proc. 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, (IH-MSP 06)., pp. 321-326, Pasadena,CA,USA2006.
[16] A. Hyvärinen, “Independent component analysis: algorithms and applications”, Neural Networks, vol. 13, no. 4-5, pp. 411-430, June 2000.

[17] A. Hyvärinen, “Fast and Robust Fixed-Point Algorithms for Independent Component Analysis”, IEEE Transactions on Neural Networks 10(3):626-634, 1999.