Fuzzy Local ICA for Speaker Recognition Using Voice and Lip Motion

A. Rouigueb, S. Chitroub, and A. Bouridane, Senior, *IEEE*

***Abstract*— Independent Component Analysis (ICA) is a powerful statistical method which can be used for many applications such as source separation, feature extraction or data representation. This paper proposes a new classification scheme spreading over two stages: (i) density estimation using local ICA based on fuzzy clustering and (ii) correction of the estimation bias using SVM classification. The observed data are grouped into fuzzy clusters and linear ICA models are locally applied on each cluster. The classification experiments are carried out over multibiometric feature vectors obtained from the fusion of lip movement and acoustic features. The results show an improvement in classification rate than the standard SVM classifier.**

***Index Terms*—Fuzzy local ICA, multimodal biometrics, lip motion, speaker recognition**

I. INTRODUCTION

SEVERAL applications require reliable verification schemes to confirm the identity of an individual requesting their service. Examples of such applications include secure access to buildings, computer systems and mobile phones. The emergence of biometric technology using physiological or behavioral traits associated with the person has addressed the problems associated with traditional verification techniques such as passwords and ID cards. However, the use of a single biometric modality may be inefficient to guarantee the performance requirements imposed by some particular applications such as high verification rates, universality and anti-spoofing [15].

The use of lip information, in conjunction with speech information, for robust speech/speaker recognition is widely investigated in the state of art [16]. Lip movements are highly correlated with the audio signal. Hence, speech content can be revealed through lip reading; lip movement patterns also contain information about the identity of the speaker. For the representation of the lip movements, lip geometry and lip texture are usually used.

Manuscript submitted February 22, 2012; revised March 21, 2012. This work was part of the PhD thesis of A. Rouigueb.

A. Rouigueb is with Signal and Image Processing Laboratory, Electronics and Computer Science Faculty, University of Science and Technology of Houari Boumedienne (U.S.T.H.B), Algiers, ALGERIA(corresponding author; phone: +447440393882; e-mail: rouigueb.abdenebi@gmail.com).

S. Chitroub is with Signal and Image Processing Laboratory, Electronics and Computer Science Faculty, University of Science and Technology of Houari Boumedienne (U.S.T.H.B), Algiers, ALGERIA(e-mail: s_chitroub@hotmail.com).

A. BOURIDANE is with School of Computing, Engineering and Information Sciences, Northumbria University, Pandon Building Newcastle upon Tyne, UK(e-mail: ahmed.bouridane@northumbria.ac.uk).

It is clear that lip features are statistically dependent on the acoustic features of speech. Therefore, the fusion at the matching score level or at the decision level does not fully take advantage from such type of correlation since the classification is achieved separately on each biometric modality. On the contrary, fusion at the feature extraction level seems to be more suitable to deal with inter-biometrics correlations. Since we have sufficient data for training, we decided to employ the feature fusion in our experiments.

Biometric recognition systems generally consist of two-stage process: feature extraction and classification. A universal ideal classifier does not exist and the choice of the appropriate classifier depends on the statistical characteristics of features. In this paper, we propose a novel classification scheme based on Independent Component Analysis method (ICA) [1]. Our first goal is to compare the classification performance on the voice-lip dataset of some known linear ICA methods and fuzzy clustering algorithms, which are involved in the proposed classifier. The second goal is to improve the classification rate compared to other standard classifiers as Support Vector Machine (SVM).

ICA is an unsupervised technique for multivariate data analysis. In spite of its application for Blind Source Problem (BSS), it can be viewed as a generalization of PCA using higher-order statistics and it can be used for representation or visualization of data for Knowledge Discovery in Databases [3]. Nonlinear ICA models seem to be more appropriate to describe real world applications than linear counterparts. Karhunen et al. proposed local ICA models [6], in which the observed data is grouped into several clusters prior to the preprocessing of linear ICA using some clustering algorithms such as K-means. Honda et al. [14] enhanced the idea to the technique that uses Fuzzy c-Varieties (FCV) clustering method.

Our goal is to develop a classifier based on the local ICA method in order to exploit the separation power of the recently developed linear ICA tools and to test it on a multibiometric dataset. Our second aim is to investigate the eventual connections between some available ICA tools and fuzzy clustering algorithms in the proposed approach.

The structure of this paper is as follows: Section 2 introduces the Fuzzy local ICA model and presents a brief review of the tested fuzzy clustering algorithms. In Section 3, we present the scheme of the proposed classifier. Extraction process of lip motion and voice features and preparation of the classification dataset are described in Section 4. Several experiments and results discussion are presented in Section 5.

II. FUZZY LOCAL ICA

A. Linear ICA Model

The ICA technique [1], which consists of recovering a set of unknown sources from their instantaneous mixtures, is an important technique in signal processing. Assuming that the sources are independent of one another, ICA algorithm tries to find a transform of the mixtures such that the recovered signals are as independent as possible.

Suppose that there exist M independent source signals $\mathbf{s}(t) = (s_1(t), \ldots, s_M(t))^T$ and N observed mixtures $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))^T$ of the sources signals (where $N \geq M$). A linear instantaneous ICA model is given by:

$$\mathbf{x}(t) = \boldsymbol{A}\mathbf{s}(t) \tag{1}$$

where $\boldsymbol{A}$ is an unknown constant full-rank $N \times M$ mixing matrix. Classical ICA aims to estimate a $M \times N$ demixing matrix $\boldsymbol{B}$ such that the M output signals

$$\mathbf{u}(t) = \boldsymbol{B}\mathbf{x}(t) = \boldsymbol{BA}\mathbf{s}(t) = \boldsymbol{PD}\mathbf{s}(t) \tag{2}$$

Here $\boldsymbol{P} \in \mathrm{R}^{M\times M}$ is a permutation matrix, $\boldsymbol{D} \in \mathrm{R}^{M\times M}$ is a diagonal scaling matrix, and $\mathbf{u}(t) = (u_1(t), \ldots, u_M(t))^T$. Consequently, the source signals are recovered up to scaling and permutation.

B. Nonlinear ICA Model

In the conventional ICA method, a linear mixture model is used for the representation of data. Although, it produces significant results in many applications, it might be unsuitable for more general data distributions. Usually, the observed variable nonlinearly depends on their sources

$$\mathbf{x}(t) = F(\mathbf{s}(t)) \tag{3}$$

where F is a non linear mixing function.

It is shown in previous works that the nonlinear ICA problem (3) is highly non-unique [2]. Consequently, additional assumptions such as external criteria or prior information should be incorporated within the optimized nonlinear ICA model in order that it converges to unique solutions [3-4]. The non-uniqueness of the nonlinear ICA solution makes the BSS problem unsolvable but does not represent an inconvenience for us; we are not interested to recover the original sources $\mathbf{s}$, but rather we are interested to find a representation of the observed signals $\mathbf{x}$ as a function of a set of Independent Components (IC) variables $\mathbf{c}(t) = (c_1(t), \ldots, c_M(t))^T$

$$\mathbf{x}(t) = G(\mathbf{c}(t)) \tag{4}$$

Hence, any solution that maximizes sufficiently the statistical independence among the ICs variables is acceptable.

The most commonly nonlinear ICA methods treat the nonlinearity in two different ways: neuronal networks approach [5] or a combination of a set of linear ICA models [3, 4, 6]. In this work, we have adopted the second approach since it is distinguished by its simplicity of implementation and its robustness for data representation.

C. Fuzzy Local ICA

Local ICA method is a nonlinear ICA method which attempts to group the observed data vectors into several clusters each having a high likelihood to be approximated by a linear ICA mixture. The clustering part is responsible for an overall nonlinear representation of data and linear ICA model are used to describe local features [6]. Karhunen et al. introduced the local ICA technique to yield a representation basis of the multivariate data where the k-means clustering algorithm is experimented. In practice, the fuzzy clustering where a vector can belong to all clusters with different membership's degrees is known to be more appropriate than crisp clustering. Moreover, spherical clusters are not often suitable for the approximation of the linear ICA models, but linear clusters can be frequently approximated by linear ICA models as c-varieties clusters in [14]. For the mentioned reasons, we propose to experiment three different fuzzy clustering algorithms: (i) Fuzzy c-means, (ii) Gustafson-Kessel and (iii) Gath Geva.

Definition: Fuzzy partitioning space *let $X = [x^1, x^2, \ldots, x^N]$ be a finite set of N vectors in K-dimension space, and let $2 < L < N$ be an integer (cluster's number), the fuzzy partitioning space for X is the set*

$$M_{fc} = \{\boldsymbol{U} \in \mathrm{R}^{N\times L} / u_{ij} \in [0,1];\ \sum_{j=1}^{L} u_{ij} = 1, \forall i;\ 0 < \sum_{i=1}^{N} u_{ij}, \forall j\} \tag{5}$$

The i^{th} row of $\boldsymbol{U}$ contains values of the membership function of the vector x^{th} which constrains the sum of each row to 1.

Fuzzy c-means clustering algorithm

One of the most widely used fuzzy clustering algorithms is Fuzzy c-means (FCM) algorithm, developed by Dunn [7] and improved by Bezdek [8]. FCM is based on the minimization of the following objective function

$$\mathrm{J}(\mathrm{X}; \boldsymbol{U}, \boldsymbol{V}) = \sum_{i=1}^{N} \sum_{j=1}^{L} (u_{ij})^m \|x_i - v_j\|_{\boldsymbol{A}}^2 \tag{6}$$

where m is a real number greater than 1 that defines the fuzziness degree, $\boldsymbol{V} = [v_1, v_2, \ldots, v_L], v_j \in \mathrm{R}^K$ is a vector of the clusters centers, which have to be determined, and $\|*\|$ is any norm expressing the similarity between data vectors and the computed centers. Distance based on the A matrix norm $\boldsymbol{A}$ is computed as follows:

$$D_{ij\boldsymbol{A}}^2 = \|x_j - v_i\|_{\boldsymbol{A}}^2 = (x_j - v_i)^T \boldsymbol{A} (x_j - v_i) \tag{7}$$

Gustafson-Kessel clustering algorithm

The Gustafson-Kessel (GK) algorithm is an extension of the FCM algorithm using an adaptive distance norm in order to find clusters of different geometrical shapes [9]. Therefore, each cluster has its own matrix norm $\mathrm{A_i}$ which yields the following inner product norm

$$D_{ij\boldsymbol{A}}^2 = (x_j - v_i)^T \boldsymbol{A}_i (x_j - v_i) \quad i = 1 \ldots L, j = 1 \ldots N \tag{8}$$

The Gath Geva algorithm

The Gath Geva (GG) [11] clustering algorithm employs a distance norm based on the fuzzy maximum likelihood estimates, proposed by Bezdek and Dunn in [10]:

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$D_{ij} = \frac{\sqrt{det(F_{wi})}}{\alpha_i} exp(\tfrac{1}{2}(x_j - v_i)^T {F_{wi}}^{-1}(x_j - v_i)) \quad (9)$$

Unlike the GK, the exponential term makes the distance decreases faster than the inner product. F_{wi} denotes the fuzzy covariance matrix of the i^{th}cluster, given by:

$$F_{wi} = \frac{\sum_{j=1}^{N}(u_{ij})^{w}(x_j - v_i)(x_j - v_i)^T}{\sum_{j=1}^{N}(uij)^{w}} \quad (10)$$

α_i is the prior probability of selecting cluster i and w is a constant.

The local ICA performances highly depend to the clustering task which is an ill-posed problem; it is related to the adopted distance and to the parameters initialization. Indeed, a linear ICA model is assigned to each cluster and therefore the distribution of the cluster vectors (cluster shape) influences the performance of the associated linear ICA, see [6, 14] for more details. Until now, the optimal shape of clusters in the local ICA context is not known and one of our objectives is to investigate the introduced clustering algorithms. Fig. 1, Fig. 2 and Fig. 3 show a two-dimensional clustering example illustrating the behavior of the proposed clustering algorithms.

III. Classification Using Fuzzy Local ICA

A. Density Estimation using Fuzzy Local ICA

Let $X = (X_1, X_2, \dots, X_K)^T$ be a K-dimensional random vector variable where $X_i \in X$ are continuous and at most one of them is allowed to be Gaussian. First, we consider the simple case where the multivariate variable X can be approximated by the linear ICA model. As the transformation $X = \boldsymbol{A} \times S$ is continuous, it has continuous partial derivatives and defines a one-to-one mapping, the joint density distribution of X can be simplified using the Jacobian formulation

$$P_X(X) = (\frac{1}{|det\ (Jac(X))|})P_S(\boldsymbol{A}^{-1} \times X) \quad (11)$$

This yields the following expression

$$P_X(X) = (\frac{1}{|det\ \boldsymbol{A})|})P_S(S) \quad (12)$$

Using the independence assumption of the independent components S, the joint density $P_S(S)$ can be factorized into a product

$$P_X(X) = \left(\frac{1}{|det\ \boldsymbol{A})|}\right) * \prod_{i=1}^{K} P_{S_i}(S_i) \quad (13)$$

Note that in (13), the density estimation problem is furthermore simplified so that we need only to estimate the density of each component S_i separately using known estimators. In this work, we have employed the Parzen window for the density estimation of the ICs.

For the general case, where the observed signals X can be nonlinearly dependent, we propose to apply local ICA models where fuzzy clustering is carried out prior to the ICA. The estimation algorithm includes the following steps:

1) Group the observed data vectors into L more homogenous clusters using fuzzy clustering methods;
2) Construct the fuzzy clusters where the defuzzification process is performed using a duplication process;
3) Subtract the cluster mean vector from vectors belonging to each cluster;
4) Estimate linear ICA separating matrix $\boldsymbol{W}$ for each zero-mean cluster using a suitable ICA algorithm;
5) Estimate the univariate density distribution for each IC using the standard methods such as the Parzen-Window estimator.

Therefore, the density estimation of variable X can be carried out by marginalisation over the constructed fuzzy clusters:

$$P_X(X) = \sum_{j=1}^{L} P_X(X, C_j)$$

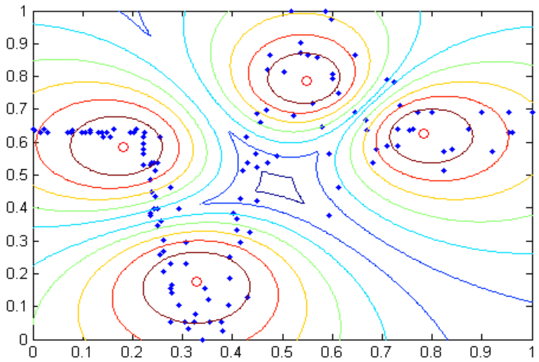

Fig. 1. FCM clustering example, 04 clusters in 2D space.

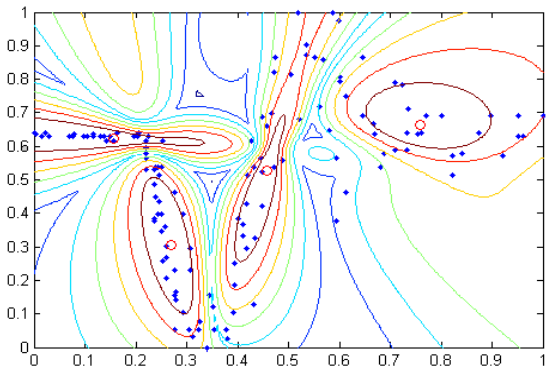

Fig. 2. GK clustering example, 04 clusters in 2D space.

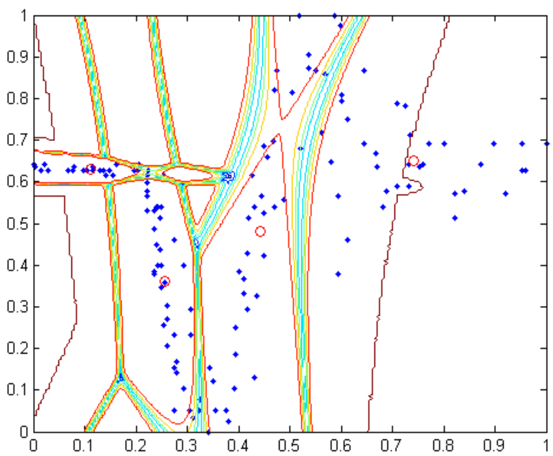

Fig. 3. GG clustering example, 04 clusters in 2D space.

By using Baye's rule, we obtain

$$P_X(X) = \sum_{j=1}^{L} P_c(C_j) * P_X(X/C_j) \quad (14)$$

where $P_c(C_j)$ is the prior probability of the cluster C_j and $P_X(X/C_j)$ is computed by using a linear ICA model (13) associated to the cluster C_j.

B. Classification System

First, for each class ω_i, $i = 1, \dots Nc$, a density estimator π_i is built using training data following the stages described in the previous Section (§III.A) where, in our case, a class represents a speaker. The parameters set of the model π_i includes the clustering parameters, the ICA mixture matrix of each cluster and the univariate density estimation parameters of each independent component.

The classifier is said to assign a feature vector $\mathbf{x}$ to class ω_c if $\mathbf{x}$ maximizes the a posteriori likelihood of the class ω_c, $c = argmax_i(P(\omega_i/\mathbf{x}))$. This yields the following decision rule:

$$c = argmax\ (P(\mathbf{x}/\omega_i)P(\omega_i)) \quad (15)$$

where $P(\omega_i)$ is the prior probability of class ω_i. However, the comparison in (15) is not quite rigorous. Indeed, each estimator is built using a different sample of data (training vectors of its associated class). Consequently, the estimators are slightly biased and then a post-processing step as normalization is indispensable.

C. Bias Correction using SVM Post-Classification

In order to address the bias problem due to the density estimation, we propose to make a classification over the space of the obtained likelihoods. Therefore, the overall scheme of the proposed classifier shown in Fig. 4 is as follows:

1) Construction of a density estimator for each class.
2) Validation by computing the likelihoods of the training vectors using the obtained estimators.
3) Training a SVM classifier over the likelihoods space where the feature k of an example i corresponds to the likelihood $P(x_i/\omega_k)$ of the training vector i in the class ω_k.

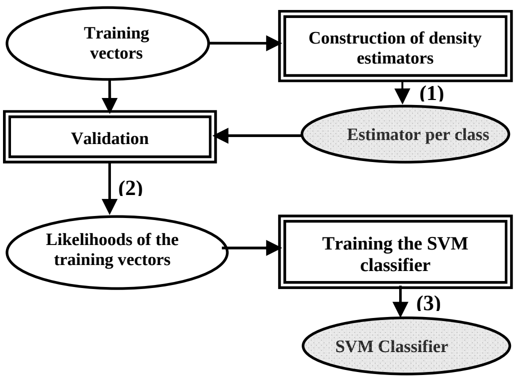

Fig. 4. Classification system design (the training phase).

During the prediction phase, we compute the likelihoods of the test vector within the different classes. Then, the final class is predicted by making SVM classification of the obtained vector of likelihoods.

IV. Data Preparation

A. MOBIO Biometric Database

The classification tests are performed on a sample of the MOBIO corpus [13]. MOBIO database consists of bi-modal (audio and video) data taken from 152 people with 12 sessions each and it was collected in six different sites from five different countries. This had led to a diverse bi-modal database with both native and non-native English speakers. The database was recorded using two mobile devices: laptop and mobile phone. The mobile phone used to capture the database was a NOKIA N93i mobile.

B. Lip Motion Traking

The images are converted from RGB into YCbCr color space. YCbCr color space splits the RGB into luminance component, Y and chrominance components, Cb and Cr. As the chrominance components in YCbCr color space are almost independent of luminance component and present a good discrimination between skin regions [20], we considered only Cb and Cr components in our model. We are interested to compute the Smallest Rectangular Window (SRW) containing the whole lips.

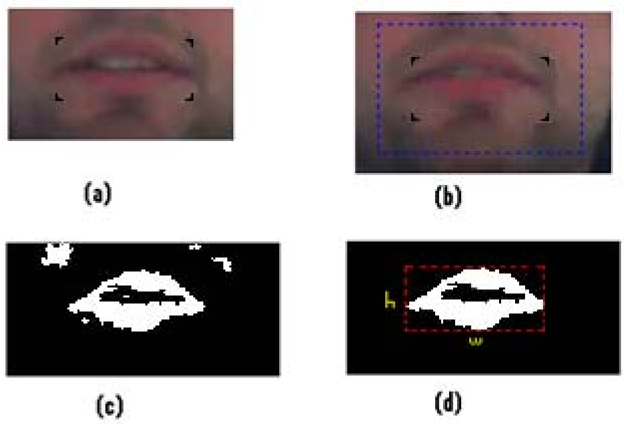

Fig. 5. Lip motion tracking. (a) SRW of the current image. (b) Region of interest on the successive image bounded by dashed rectangle. (c) Segmentation. (d) The new SRW on the successive image.

In the sequence video, given that the SRW is well detected on an input image, the detection of the SRW of the successive image is performed as follows:

1) Finding a concentric rectangle with SRW in the successive images such that the size of the new rectangle window should be chosen in a way that it sufficiently covers all the lip motion only excluding the surrounding parts as noise and ears. The size of the new rectangle in the successive image is taken equal to the double of the SRW of the current image (dashed rectangle in Fig. 5(b)).
2) The lip detection is performed on the skin pixels after segmenting the normalized image into skin and non-skin regions using the histogram thresholding. The use of automatic thresholding is justified by the fact that the SRW histograms contain two clear peaks (lips /skin around lips) and one valley, the peak corresponding to teeth (if the mouth is open) is relatively small. Moreover, we have prior information about the

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

approximate color range and part of pixels constituting lips. Therefore, segmentation is carried out by following a peak-and-valley thresholding [21] variant; the threshold is estimated by finding the deepest valley succeeded the lip peak. (see example of segmentation on Fig. 5(c)).

3) Extraction of the largest white connected surface and determination of the smallest rectangle window containing lips on the successive image (dashed rectangle on Fig. 5 (d)).

The region of interest containing the mouth is fixed manually in the first image in the sequence video.

In order to well characterize lip motion while speaking, five global features describing the geometry and the texture of lip motion are considered. These features are respectively the wide w and the high h of the SRW, the lips surface (numbers of white pixels on Fig.4. (d)), the mean and the standard deviation of the SRW.

It is worth recalling that the selected features vary from a person to another and for the same person they vary according to the spoken phonemes.

C. Acoustic Features

Cepstral coefficient (MFCC) features are one of the widely used features for speech and speaker recognition [12]. In this work, speech data are acquired with a sampling rate equals to 16Khz. We use MFCC as the main feature vector where 12 MFCCs are computed for every 6.25ms using a 25ms Hamming window.

D. Fusion

The multibiometric fusion is carried out at the level of features. The 05 feature lip vector is appended with the 12 MFCC vector to give 17 features in all. The frame rate of the voice signal is equal to 160 frames/sec, whereas it is 30 frames/sec for the lip video. Consequently, acoustic and visual frames need to be synchronized. Hence visual features are interpolated from 30 frames/sec to 160 frames/sec; for each MFCC vector, the closest lip vector in terms of chronology order is appended to it as illustrated in Fig. 6.

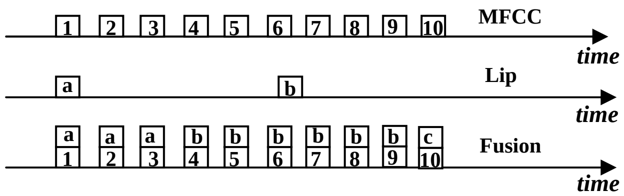

Fig. 6. Vectors features fusion process.

For each speaker, we have used 05 phrases for training the classifier and 05 phrases for testing. Following the described extraction process, the set vectors number of the different classes is shown in the Table I.

TABLE I
NUMBER OF CLASSIFICATION VECTORS

Class	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Train	3413	1884	1505	1709	1010	1987	1876	1628	1676	1384
Test	1279	1463	1495	1184	944	1338	1200	963	1112	848

V. RESULTS AND DISCUSSION

Traditionally, the Receiver Operating Characteristics (ROC) curve and the Equal Rate Error (EER) are often used for biometric systems performance evaluation. In our experiments, we have only employed a sample of 10 speakers where 05 phrases per person are used for testing. Due to the size of the sample, the EER or ROC measures are not reliable for the evaluation. Indeed, the recognition rate of phrases (containing at least 200 of feature vectors each) obtained by vote over the classification results of vectors is 100%. Therefore, we propose to evaluate the performance by means of the classification rate of features vectors.

TABLE II
SVM CLASSIFICATION RATE

Dataset	SVM kernel RBF	Linear	Polynomial
Lips	**0.6783** (c^=1000, σ^=0.0021)	0.6675 (c^=10)	0.6775 (c^=100, d^=2)
Voice	**0.2471** (c^=1000, σ^=0.0351)	0.1859 (c^=1000)	0.2371 (c^=10, d^=3)
Lip+Voice	0.7195 (c^=1000, σ^=0.0011)	0.7233 (c^=10)	**0.7245** (c^=10, d^=3)

TABLE II shows the classification rates obtained by the standard SVM classifier using RBF, linear and polynomial kernels, respectively. The optimal parameters: the cost (c), sigma (σ) for RBF kernel and degree (d) for polynomial kernel are determined empirically. From the results, it can be noticed that the fusion improves the classification rates.

In the proposed classification framework, the following linear ICA algorithms are involved in the experiments:

- FastICA [1] based on the maximization of the non gaussianty;
- JADE [17] based on the Joint Approximate Diagonalization of 4th order cumulant matrices;
- KDICA [18], which makes use of a novel super-fast kernel density estimation algorithm that is designed using a Laplacian kernel for solving the ICA problem;
- FastKernelICA [19], which uses kernel measures of statistical independence. It employs an approximate Newton method to perform the optimization efficiently for larger-scale problems.

All the possible combinations of the introduced ICA algorithms and fuzzy clustering algorithms GK, FCM and GG are tested. The classification rates of lip, voice and fusion (lip&voice) are respectively displayed on TABLE III, IV and V.

TABLE III
LIP CLASSIFICATION RATE

ICA tool	Fuzzy clustering algorithm GK	FCM	GG
FastICA	0.7068	0.7160	0.7160
JADE1	0.7029	0.7118	0.7086
FastKernelICA	0.7100	0.6988	0.7157
KDICA	0.7146	**0.7170**	0.7129

TABLE IV
VOICE CLASSIFICATION RATE

ICA tool	Fuzzy clustering algorithm GK	FCM	GG
FastICA	0.2431	0.2391	0.2429
JADE1	0.2407	0.2389	0.2438
FastKernelICA	0.2195	0.2244	0.2272
KDICA	0.2418	0.2380	**0.2456**

TABLE V
LIP & VOICE CLASSIFICATION RATE

ICA tool	Fuzzy clustering algorithm GK	FCM	GG
FastICA	0.7623	0.7648	0.7464
JADE1	0.7650	0.7658	0.7475
FastKernelICA	0.7435	0.7344	0.7496
KDICA	**0.7726**	0.7710	0.7275

While comparing the performance of the different combinations of ICA and fuzzy clustering algorithms, the obtained classification rates are generally close one another. However, we can reveal some interesting remarks. In the most cases, KDICA provides slightly better performance than the other ICA tools.

For the voice dataset results, shown in Table V, GG outperform slightly the GK and FCM clustering algorithms for all ICA tools. In the three datasets, FastKernelICA fits well with GG clustering algorithm than FCM and GK.

Fig. 7 shows the impact of the post-processing stage in order to adjust the estimation bias using SVM classification. We note that the improvement gain of the bias adjustment is about 10%.

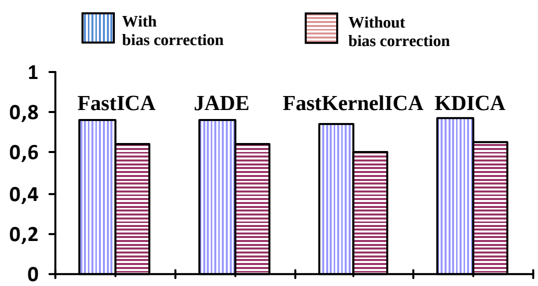

Fig. 7. Improvement gain of the classification rate using SVM post-classification (where GK clustering is employed).

VI. CONCLUSION

We have proposed a novel scheme of classification on two stages. Density estimation for each class is first achieved via a technique using local ICA models based on fuzzy clustering. While testing phase, computed probabilities of vectors are often slightly biased differently among class's estimators. This estimation bias is due to the fact each estimator is built separately using the vectors of its class. Hence, we propose to make a classification over the probabilities space in order to adjust the bias. The proposed scheme may be seen as a hybridization of a generative and a discriminative classifiers [22], the first one (density estimator in our case) relates to evaluating the conditional probability P(y/x) of a class y given a vector x using Bayes rule, whereas the discriminative one (SVM in our case) tries to predict the potential class by regarding the posterior probabilities of all the classes. The proposed technique is used for the vectors classification of an audio-visual biometric dataset. The dataset vectors are obtained by the fusion of the acoustic and lip motion features for the speaker recognition purpose. This paper addresses two issues: (i) make a comparative study of the possible combinations of a set of available ICA tools and fuzzy clustering algorithms in our approach, (ii) improve the classification rate. On the introduced bimodal dataset (lip/voice), the proposed technique shows a significant improvement than the standard SVM classifier.

REFERENCES

[1] A. Hyvrinen, J. Karhunen and E. Oja, "*Independent Component Analysis*," JOHN WILEY & SONS, INC, 2001.

[2] A. Hyvarinen and P. Pajunen, "Nonlinear independent component analysis: Existence and uniqueness Results," *Neural Networks*, vol. 12, pp. 429-439, 1999.

[3] K. Honda and H. Ichihashi, "Fuzzy local independent component analysis with external criteria and its application to knowledge discovery in databases," *International Journal of Approximate Reasoning*, vol. 42, pp. 159-173, 2006.

[4] K. Honda and H. Ichihashi, "Fuzzy Local ICA for Extracting Independent Components Related to External Criteriato knowledge discovery in databases," *Applied Mathematical Sciences*, vol. 2, no. 6, pp. 275-291, 2008.

[5] E.F. Simas Filho, J.M. de Seixas and L.P. Calba, "Modified postnonlinear ICA model for online neural discrimination," *Neurocomputing*, vol. 73, pp. 2820-2828, 2010.

[6] J. Karhunen and S. Malaroiu, "Local Independent Component Analysis Using Clustering," In Proc. First Int. *Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pp. 43-48, 1999.

[7] J.C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *Journal of Cybernetics*, vol. 3, pp. 32-57, 1973.

[8] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algoritms*. Plenum Press, New York, 1981.

[9] D.E. Gustafson and W.C. Kessel, "Fuzzy clustering with fuzzy covariance matrix," *Proc. IEEE CDC*, San Diego, pp. 761-766, 1979.

[10] J.C. Bezdek and J.C. Dunn, "Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal dustrubutions. IEEE Transactions on Computers," pages 835-838,1975.

[11] I. Gath and A.B. Geva, "Unsupervised optimal fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence," vol. 7, pp. 773-781, 1989.

[12] D.A Reynold, "Speaker identification and verification using Gaussian mixture speaker models," Speech Comm, vol. 17, pp. 91–108, 1995.

[13] http://www.mobioproject.org/, dec, 2011. MOBIO (Mobile Biometry) is an Europpean Funded Project where the objectives include robust face and speaker authentication.

[14] K. Honda, H. Ichihashi, M. Ohue, K. Kitaguchi, "Extraction of local independent components using fuzzy clustering," in: Proceedings of 6th International Conference on Soft Computing (IIZUKA2000), pp. 837–842, 2000.

[15] A. Ross, A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, pp. 2115–2125, 2003.

[16] H. E. Çetingül, Y. Yemez, E. Erzin, and A. M. Tekalp, "Discriminative Analysis of Lip Motion Features for Speaker Identification and Speech-Reading," IEEE *Transactions on Image Processing*, vol. 15, no. 10, pp. 2879-2891, 2006.

[17] J.F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362-370, 1993.

[18] A. CHEN, "Fast kernel density independent component analysis," *In ICA '06; Lecture Notes in Computer Science*, Vol. 3889, pp. 24-31, 2006.

[19] H. Shen, S. Jegelka and A. Gretton, "Fast Kernel-Based Independent Component Analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3498-3511, 2009.

[20] K. Li, M. Wang, M. Liu, A. Zhao, "Improved Level Set Method For Lip Contour Detection," in *Proc. IEEE 17th International Conference on Image Processing*, Hong Kong, Sep. 26-29, 2010.

[21] M. I. Sezan, "A peak detection algorithm and its application to histogram-based image data reduction,'' *Graph. Models Image Processing*, *vol.* 29, pp. 47-59, 1985.

[22] A. Y. Ng and M. I. Jordan, "On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes,'' *NIPS* , pp. 841-848 , 2001.

ISBN: 978-988-19251-3-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)