ROBUST TALKING FACE VIDEO VERIFICATION USING JOINT FACTOR ANALYSIS AND SPARSE REPRESENTATION ON GMM MEAN SHIFTED SUPERVECTORS

Ming Li, Shrikanth Narayanan

Signal Analysis and Interpretation Laboratory, University of Southern California, USA

mingli@usc.edu, shri@sipi.usc.edu

ABSTRACT

It has been previously demonstrated that systems based on block wise local features and Gaussian mixture models (GMM) are suitable for video based talking face verification due to the best trade-off in terms of complexity, robustness and performance. In this paper, we propose two methods to enhance the robustness and performance of the GMM-ZTnorm baseline system. First, joint factor analysis is performed to compensate the session variabilities due to different recording devices, lighting conditions, facial expressions, etc. Second, the difference between the universal background model (UBM) and the maximum a posteriori (MAP) adapted model is mapped into the GMM mean shifted supervector whose over-complete dictionary becomes more incoherent. Then, for verification purpose, the sparse representation computed by l^1-minimization with quadratic constraints is employed to model these GMM mean shifted supervectors. Experimental results show that the proposed system achieved 8.4% (group 1) and 10.5% (group 2) equal error rate on the Banca talking face video database following the P protocol and outperformed the GMM-ZTnorm baseline by yielding more than 20% relative error reduction.

Index Terms— face video recognition, sparse representation, GMM supervector, joint factor analysis

1. INTRODUCTION

Face recognition using video sequence has recently gained significant attention [1]. With built-in cameras and microphones becoming a standard feature on most personal computing and mobile devices, audio-visual biometrics has become a natural way for user verification and personal secure access. Specifically, face verification based on a video sequence of a talking face rather than just one or a few still images offers possibility for increased robustness. It has been previously demonstrated that systems based on block wise local features and Gaussian mixture models (GMM) are suitable for video based talking face verification as they offer the best trade-off in terms of complexity, robustness and performance [2, 3]. In this paper, we follow this framework and focus on further enhancing the robustness and performance of the GMM modeling, notably by exploring sparse representations of the talking face.

The verification task based on talking face images acquired in an uncontrolled environment, such as with a mobile device, is very challenging. A large variability in facial appearance of the same subject is caused by variations of recording devices, illumination, and facial expression. These variations are further increased by errors in face localization, alignment and normalization. While facial dynamic information (continuity in head/camera movement, facial expression or photometric continuity) has been studied for robust face video recognition [4], algorithms based on the similarity of unordered image sets have also been proposed [5]. Furthermore, in the GMM framework based on block wise local features, selection of the good image frames using quality measurements [3] and score normalization (ZT-norm) [6] have been proposed to compensate for the session variability. However, the GMM modeling is still based on the UBM training and MAP adaptation framework. Recently, joint factor analysis (JFA) [7] has been successfully used in the speaker verification task in which session variability caused by different channels influences the system performance dramatically [7]. In this work, given data from multiple sessions, we divide each face video sequence into several continuous short segments and adopt the JFA approach to reduce the intra-personal variations within all these segments.

A key concept in the JFA approach is to use a GMM supervector consisting of the stacked means of all the mixture components [8, 7]. Support vector machine (SVM) based on this GMM mean supervectors forme the GMM-SVM supervector system which has been successfully applied in the speaker verification task [8]. More recently, a sparse representations computed by l^1-minimization with equality constraints were proposed to replace the SVM in the GMM mean supervector modeling and has been demonstrated to be effective in the closed set speaker identification task on the clean TIMIT database [9]. However, the sparse representation of GMM mean supervectors has not been explored or exploited in detail to handle the robust face video verification task against large session variabilities.

In this work, we exploit the discriminative nature of sparse representations to perform face verification based on GMM supervectors. Given a verification trial with the test supervector and the target identity, we first construct an over-complete dictionary using all the target supervector samples and non-target background supervector samples, then calculate the sparsest linear representation via l^1 norm minimization. The membership of the sparse representation in the over-complete dictionary itself captures the discriminative information given sufficient training samples [10]. If the trial is true, the test sample should have a sparse representation whose nonzero entries concentrate mostly on the target samples whereas the test sample from a false trial should have sparse coefficients spread widely among multiple subjects [10]. For most verification tasks, the number of non-target background subjects/samples are naturally way larger than the number of target subjects/samples, thus the chance nonzero entries on the target training samples for a test sample from a false trial should be arbitrarily small and close to zero. Therefore, for the calculated sparse representation, the l^1 norm ratio between the target samples and all the samples in the over-complete dictionary becomes the verification decision criterion. Based on the overwhelming unbalanced non-target negative training samples and the very limited target positive training samples, in contrast to the SVM system which requires to tune the SVM cost values each time,

This work was supported in part by United States Army and NSF.

the proposed framework utilizes the highly unbalanced nature of the training samples to form a sparse representation problem.

Furthermore, we proposed three methods to enhance the robustness and performance against variabilities. First, the sparse representation is computed by l^1-minimization with quadratic constraints rather than equality constraints. Second, by adding a reductant identity matrix at the end of the original over-complete dictionary, the sparse representation is more robust to the variability and noise [10]. Third, the difference between the UBM and the MAP adapted model is mapped into the GMM mean shifted supervector which not only preserves the distance of the associated GMM but also makes the supervector sparse. Compared to the conventional mean supervector, the correlation of the constructed over-complete dictionary becomes smaller and therefore helps to achieve robust sparse representation.

The paper organization is as follows: Section 2 describes the proposed methods, Section 3 provides the experimental results and Section 4 summarizes the conclusions.

2. METHODS

2.1. Face localization and feature extraction

Given a sequence of face images, face detection was performed for each image frame by using the Viola-Jones face detector in the opencv library, and then the Biosecure talking face reference system [5] with the MPT library [11] was used to find the location of two eyes. Given the detected face image and eye location, a geometric normalization tool [12] was applied to crop the detected face image (200×240) into a normalized face image (51×55). The histogram was globally equalized for each cropped and normalized image.

Furthermore, DCTmod2xy [2] feature vectors were extracted for each block of every normalized face image. The 20 dimensional DCTmod2xy feature is the standard DCTmod2 feature [13] plus (x,y) coordinates of each block. Due to the included space domain information, it performs better than the DCTmod2 feature [2, 3]. Thus given the block size of 8×8 pixels, each normalized face generated $11 \times 12 = 132$ frames of DCTmod2xy feature vectors which were assumed to be independent and modeled by GMM. Compared to the holistic feature based systems, this block wise local feature based system was reported to be robust against face localization and normalization errors [2, 3]. Finally, the feature vectors were normalized to mean zero and unit variance on a per-video basis.

2.2. GMM-ZTnorm baseline

The Gaussian Mixture model (GMM) is used to model the DCTmod2xy features. Each target subject is represented by a N components GMM model λ: $\lambda = \{p_i, \boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}\}, i = 1, \cdots, N$. In the proposed work, since the amount of training data for each subject is too limited to train a good GMM, a universal background model (UBM) in conjunction with a MAP model adaptation approach [2, 14] is used to model different subjects' faces in a supervised manner. In order to enhance the robustness, Z-norm and T-norm [15] were used to normalize the GMM log-likelihood scores.

2.3. Joint factor analysis

JFA has been widely used in the audio speaker verification task in which session variability caused by different channels influences the system performance dramatically [7]. In the proposed method, we use the same word “channel” to describe both intra-personal and recording variabilities. Thus, given a multi-sessions development data set, we have multiple realizations from different channels for each subject. Furthermore, for the talking face video data, there are inherent session variabilities (facial expression, face orientation, lip shapes, etc) within each video sequence. Thus, we divided each video sequence of the development data into several 50 frames segments, and use the JFA approach to model the intra-personal variabilities within all the segments from the same subject. A model that has been trained under one channel condition may be adapted towards a different channel condition of new test data to reduce mismatches when the subject is the same. Importantly, the adaptation must be constrained so that adaptation between different subjects is suppressed. This constraint is effected by perform adaptation in a very small subspace of the GMM mean supervector space [7].

2.4. GMM supervector modeling base on sparse representation

2.4.1. GMM mean shifted supervector

For each segment of video sequence, a GMM was adapted from the UBM by MAP adaptation; the GMMs were modeled with diagonal covariance matrices and only the means of the GMMs were adapted [2, 14]. The KL divergence $D(\lambda, \hat{\lambda})$ between two GMM models $(\lambda, \hat{\lambda})$ is approximated by the upper bound distance $d(\lambda, \hat{\lambda})$ which satisfy the Mercer condition [8]:

$$0 \le D(\lambda, \hat{\lambda}) \le d(\lambda, \hat{\lambda}) = \frac{1}{2} \sum_{i=1}^{N} p_i (\boldsymbol{\mu_i} - \boldsymbol{\hat{\mu}_i})^t \boldsymbol{\Sigma_i^{-1}} (\boldsymbol{\mu_i} - \boldsymbol{\hat{\mu}_i}) \quad (1)$$

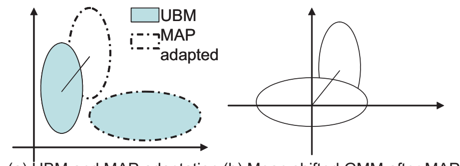
The linear kernel is defined as the corresponding inner product of the GMM mean supervectors $\boldsymbol{m}$ which is a concatenation of the weighted GMM mean vectors $\boldsymbol{m} = [\boldsymbol{m_1^t}, \cdots, \boldsymbol{m_i^t}, \cdots, \boldsymbol{m_N^t}]^t$ [8]:

$$K(\lambda, \hat{\lambda}) = \sum_{i=1}^{N} \boldsymbol{m_i^t \hat{m}_i} = \sum_{i=1}^{N} (\sqrt{p_i} \boldsymbol{\Sigma_i^{-\frac{1}{2}}} \boldsymbol{\mu_i})^t (\sqrt{p_i} \boldsymbol{\Sigma_i^{-\frac{1}{2}}} \boldsymbol{\hat{\mu}_i}) \quad (2)$$

where p_i and $\boldsymbol{\Sigma_i}$ are the i^{th} UBM mixture weight and diagonal covariance matrix, $\boldsymbol{\mu_i}$ corresponds to the mean of the i^{th} Gaussian component in this GMM. Let the UBM model be $\widetilde{\lambda} = \{p_j, \boldsymbol{\widetilde{\mu}_j}, \boldsymbol{\Sigma_j}\}$, then we can construct mean shifted GMM model $\lambda^\star$ by subtracting the UBM mean vector from the MAP adapted model: $\lambda^\star = \{p_i, \boldsymbol{\mu_i^\star}, \boldsymbol{\Sigma_i}\} = \{p_i, \boldsymbol{\mu_i} - \boldsymbol{\widetilde{\mu}_i}, \boldsymbol{\Sigma_i}\}, i = 1, \cdots, N$. It is clear that the distance between the mean shifted models is exactly the same as the distance between the original models. Thus, the GMM mean shifted supervector $\boldsymbol{s}$ is generated from the mean shifted GMM using (2). In MAP adaptation [14], the mean vector is updated as follows:

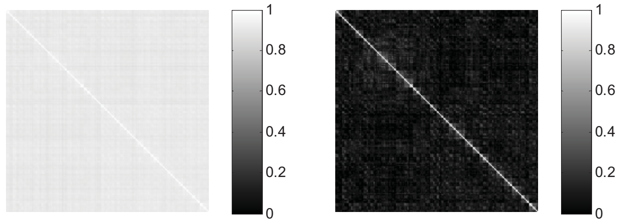
$$\boldsymbol{\mu_i} = \alpha_i \boldsymbol{E_i(x)} + (1 - \alpha_i) \boldsymbol{\widetilde{\mu}_i}, \qquad \alpha_i = \frac{\sum_{t=1}^{T} Pr(i|\boldsymbol{x_t})}{\gamma + \sum_{t=1}^{T} Pr(i|\boldsymbol{x_t})} \quad (3)$$

where $Pr(i|\boldsymbol{x_t})$ denotes the occupancy probability of feature frame t belonging to the i_{th} gaussian component and γ is the constant relevance factor. Therefore, $\boldsymbol{\mu_i^\star} = \alpha_i(\boldsymbol{E_i(x)} - \boldsymbol{\widetilde{\mu}_i})$. Given a segment of feature vectors and a large sized UBM, α_i can be arbitrarily small on certain gaussian components due to the small occupancy probability and lack of enough data to update [14]. Thus the entries of the corresponding dimensions on mean shifted supervector are close to zero. It is shown in Fig.2 that the over-complete dictionary constructed using mean shifted supervectors significantly reduced the correlation between atoms. Since an incoherent over-complete dictionary can provide better performance in l^1-minimization sparse representation [16, 10], the proposed mean shifted supervector is more suitable in this framework compared to the traditional mean supervector. From Fig.1, we can see that the GMM mean shifted supervector models



(a) UBM and MAP adaptation (b) Mean shifted GMM after MAP

Fig. 1. MAP adaptation and mean shifted GMM model



(a) using mean supervectors (b) using mean shifted supervectors

Fig. 2. The correlation matrix of the over-complete dictionary $\boldsymbol{A_2}$ (4), $N_2 = 4601$. The coherence values $max_{j\neq k}\langle \boldsymbol{A_{2j}}, \boldsymbol{A_{2k}}\rangle)$ are 0.996 (left) and 0.963 (right), respectively.

the distance between the MAP adapted model and the UBM model rather than the mean of the MAP adapted model itself. By taking out the common and dominant component (UBM) from each adapted model, the over-complete dictionary composed of all the training supervectors becomes more incoherent while the discriminative distance measure is still preserved. The reason that the coherence was not reduced significantly in Fig.2 might due to some close to duplicate samples from the same subject. Thus other dictionary design approaches [17], such as duplicate samples removal, can also be adopted on the mean shifted supervectors to further enhance the robustness and performance of sparse representations.

2.4.2. Sparse representation based on mean shifted supervectors

Given N_1 target training samples $\boldsymbol{A_1}$ and N_2 non-target background training samples $\boldsymbol{A_2}$, we construct the over-complete dictionary $\boldsymbol{A}$:

$$\boldsymbol{A} = [\boldsymbol{A_1}\boldsymbol{A_2}] = [\boldsymbol{s_{11}}, \boldsymbol{s_{12}}, \cdots, \boldsymbol{s_{1N_1}}, \boldsymbol{s_{21}}, \boldsymbol{s_{22}}, \cdots, \boldsymbol{s_{2N_2}}]. \quad (4)$$

Each sample $\boldsymbol{s_{ij}}$ is an M dimensional GMM mean shifted supervector and has been normalized to unit l^2 norm. Throughout the entire testing progress, the background samples $\boldsymbol{A_2}$ were fixed; and only the target samples $\boldsymbol{A_1}$ were replaced according to the claimed target identity in the test trial. Let us denote $N = N_1+N_2$, then $N_1 \ll N_2$ and $M < N$ need to be satisfied for sparse representation. For any test sample $\boldsymbol{y} \in \mathbb{R}^M$ with unit l^2 norm, we want to use the over-complete dictionary $\boldsymbol{A}$ to linearly represent $\boldsymbol{y}$ in a sparse way. If $\boldsymbol{y}$ is from the target, then $\boldsymbol{y}$ will approximately lie in the linear span of training samples in $\boldsymbol{A_1}$ [10]. Since the equality constraint $\boldsymbol{Ax} = \boldsymbol{y}$ is not robust against large session variabilities [10], we constrain the distance (1) between the GMM associated with the test sample $\boldsymbol{y}$ and the GMM associated with the linear combination of training samples to be smaller than ϵ which resulted in a standard convex optimization problem (l^1-minimization with quadratic constraints):

$$\text{Problem A}: \quad \min\|\boldsymbol{x}\|_1 \qquad \text{subject to}\|\boldsymbol{Ax}-\boldsymbol{y}\|_2 \leq \epsilon \quad (5)$$

For each class i, ($i = 1, 2$), let $\delta_i : \mathbb{R}^N \rightarrow \mathbb{R}^N$ be the characteristic function which selects the coefficients only associated with the i_{th} class [10]. For $\boldsymbol{x} \in \mathbb{R}^N$, $\delta_1(\boldsymbol{x}) \in \mathbb{R}^N$ is a new vector whose nonzero entries are the only entries in the first N_1 elements of $\boldsymbol{x}$. Now based on the sparse representation $\boldsymbol{x}$, we propose two decision criteria for verification purpose, namely l^1 norm ratio and l^2 residual ratio. Larger score represents higher likelihood for the testing sample to be from the target subject.

$$l^1\text{norm ratio} \quad = \quad \|\delta_1(\boldsymbol{x})\|_1/\|\boldsymbol{x}\|_1 \quad (6)$$
$$l^2\text{residual ratio} \quad = \quad \|\boldsymbol{y}-\boldsymbol{A}\delta_2(\boldsymbol{x})\|_2/\|\boldsymbol{y}-\boldsymbol{A}\delta_1(\boldsymbol{x})\|_2 \quad (7)$$

Due to large session variabilities, the test sample $\boldsymbol{y}$ could be partially corrupted. Thus an error vector $\boldsymbol{e}$ was introduced to explain the variability [10]:

$$\boldsymbol{y} = \boldsymbol{y_0} + \boldsymbol{e} = \boldsymbol{Ax_0} + \boldsymbol{e} \quad (8)$$

So the original optimization problem becomes the following form:

$$\text{Problem B}: \quad \min\|\boldsymbol{z}\|_1 \qquad \text{subject to}\|\boldsymbol{Bz}-\boldsymbol{y}\|_2 \leq \epsilon \quad (9)$$
$$\boldsymbol{B} = [\boldsymbol{A} \quad \boldsymbol{I}] \in \mathbb{R}^{M\times(N+M)}, \boldsymbol{z} = [\boldsymbol{x}^t \quad \boldsymbol{e}^t]^t \in \mathbb{R}^{(N+M)} \quad (10)$$

If the error vector $\boldsymbol{e}$ is sparse and has no more than $(M + N_1)/2$ nonzero entries, the new sparse solution $\boldsymbol{z}$ is the true generator according to (8) [10]. Finally, we redefine the two decision criteria based on the new sparse solution $\hat{\boldsymbol{z}} = [\hat{\boldsymbol{x}}^t \quad \hat{\boldsymbol{e}}^t]^t$.

$$l^1\text{norm ratio} \quad = \quad \|\delta_1(\hat{\boldsymbol{x}})\|_1/\|\hat{\boldsymbol{x}}\|_1 \quad (11)$$
$$l^2\text{residual ratio} \quad = \quad \|\boldsymbol{y}-\hat{\boldsymbol{e}}-\boldsymbol{A}\delta_2(\hat{\boldsymbol{x}})\|_2/\|\boldsymbol{y}-\hat{\boldsymbol{e}}-\boldsymbol{A}\delta_1(\hat{\boldsymbol{x}})\|_2$$

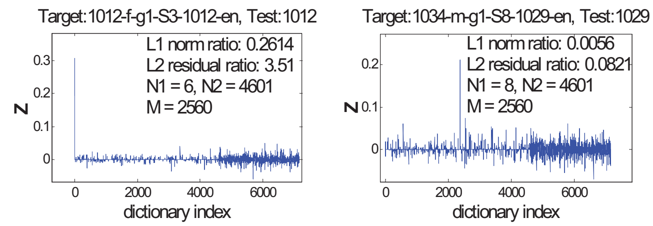


Fig. 3. The sparse solution of two trials using problem B (9)

Fig.3 demonstrates the sparse solutions and ratio scores of two trials (left: true, right: false) in the evaluation using problem B (9).

3. EXPERIMENTAL RESULTS

3.1. The evaluation database and protocols

The Banca talking face video database [18] was used for evaluation. The Banca English database has 2 groups. Each group has 26 target subjects (13 female, 13 male) and for each subject, 12 sessions were recorded in 3 different scenarios. The P protocol [18] was employed for evaluation. There are another 60 world model video recordings from different subjects (not in the 52 subjects set) for background UBM training. The evaluation method is the Equal Error Rate(EER). Since data set g1 and g2 are separated, g2 is the development set for g1, and vise versa. For the GMM modeling, the mixture number was 128 (tuned by system 2) while Z-norm, T-norm, JFA and non-target background data sets were all from the development data set. A relevance factor of 12 was used for the MAP adaptation. N_2 and mean value of N_1 are 4601 and 8.62 for g1 and 3843 and 11.34 for g2. The P protocol has 544 trials (232 true and 312 false) and 545 trials (234 true and 311 false) for g1 and g2, respectively. The sparse representation was achieved by the SPGL tool[19]. Rather than using

Table 1. EER (%) performance of the proposed systems

Methods / System ID	1	2	3	4	5
GMM baseline	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
ZT norm		$\surd$	$\surd$		
JFA			$\surd$	$\surd$	$\surd$
GMM-SVM				$\surd$	
GMM-Sparse					$\surd$
P protocol G1	27.9	14.2	12.8	10.2	**8.4**
P protocol G2	29.6	13.3	12.0	12.3	**10.5**

Table 2. EER (%) of the sparse representation system (P protocol)

criterion / problem settings	A eq(5)	B eq(9)	B eq(9)
supervectors	mean	mean	mean shifted
g1: l^2 residual ratio	**15.5**	11.5	9.9
g2: l^2 residual ratio	15.0	13.3	11.8
g1: l^1 norm ratio	12.5	10.9	**8.4**
g2: l^1 norm ratio	13.6	12.4	**10.5**

the provided pre-selected 5 images set to perform testing, we used all the images in the testing video sequence without selection by face quality measurements which will be considered in the further work.

3.2. The evaluation results

As shown in Table 1, ZT-norm dramatically improved the results which matches the conclusion of [6]. Furthermore, the use of JFA reduced the EER by 1.3% absolutely which demonstrates the good performance of the JFA method in terms of variability compensation. Comparing the results of system 4 with 5, we can observe that sparse representation based on mean shifted supervectors consistently outperformed the SVM mean supervector system by 1.8% absolute EER reduction in both groups. By only using the proposed JFA and sparse representation approaches, system 5 achieved the best performance which significantly improved the performance of GMM ZT-norm baseline. Moreover, compared to the results in [2, 3, 6], the proposed method also achieved highly competitive results.

In Table 2, the performance of different sparse representation problem settings and decision criteria are shown. First, the sparse solution computed by the problem B achieved better results compared to the problem A which validates the assumption that adding an error vector can enhance the robustness against variabilities. Second, the systems using mean shifted supervectors performed consistently better than the ones employed the traditional mean supervectors. This can be attributed to the more incoherent over-complete dictionary constructed by mean shifted supervectors. Third, the l^1 norm ratio criterion outperformed the l^2 residual ratio criterion which matches the results of sparsity concentration index (SCI) criterion in the open set rejection task [10]. Finally, adopting the sparse representation method in [9] (15.5% EER) did not improve the results compared to system 4. However, in the proposed framework, our system 5 achieved significantly improvement against system 4.

4. CONCLUSIONS

In this work, a robust video based face verification approach using block wise local features and GMM modeling was proposed. The main novelties are as follows. First, by employing joint factor analysis to mitigate the intra-personal variabilities, the system performance was enhanced. Second, the over-complete dictionary based on GMM mean shifted supervectors is more incoherent while the discriminative GMM distance measurement is still preserved which improved the performance dramatically. Third, we proposed a general verification framework based on sparse representation on GMM supervectors which outperformed the conventional SVM method by more than 15% relative error reduction.

5. REFERENCES

[1] S. Zhou, V. Krueger, and R. Chellappa, "Probabilistic recognition of human faces from video," *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 214–245, 2003.

[2] F. Cardinaux, C. Sanderson, and S. Bengio, "User authentication via adapted statistical models of face images," *IEEE Trans. Signal Processing*, vol. 54, no. 1, pp. 361–373, 2005.

[3] N. Poh, C. Chan, J. Kittler, S. Marcel, C. Cool, E. Rúa, J. Castro, M. Villegas, R. Paredes, V. Štruc, et al., "Face video competition," *Lecture Notes in Computer Science*, vol. 5558, pp. 715–724, 2009.

[4] K.C. Lee, J. Ho, M.H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *CVPR*, 2003, vol. 1, p. 313.

[5] H. Bredin, G. Aversano, C. Mokbel, and G. Chollet, "The biosecure talking-face reference system," in *2nd Workshop on Multimodal User Authentication*, 2006.

[6] N. Poh, J. Kittler, S. Marcel, D. Matrouf, and J.F. Bonastre, "Model and Score Adaptation for Biometric Systems: Coping With Device Interoperability and Changing Acquisition Conditions," in *ICPR*, 2010, pp. 1229–1232.

[7] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1435–1447, 2007.

[8] W.M. Campbell, D.E. Sturim, and D.A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308–311, 2006.

[9] I. Naseem, R. Togneri, and M. Bennamoun, "Sparse Representation for Speaker Identification," in *ICPR*, 2010, p. 4460.

[10] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal Mach Intell*, vol. 31, no. 2, pp. 210–227, 2008.

[11] I. Fasel, R. Dahl, J. Hershey, and B. Fortenberry, *The Machine Perception Toolbox*, http://sourceforge.net/projects/mptbox/.

[12] J.R. Beveridge, D. Bolme, B.A. Draper, and M. Teixeira, "The CSU face identification evaluation system," *Machine vision and applications*, vol. 16, no. 2, pp. 128–138, 2005.

[13] C. Sanderson and KK Paliwal, "Fast feature extraction method for robust face verification," *Electronics Letters*, vol. 38, no. 25, pp. 1648–1650, 2002.

[14] D.A. Reynolds, T.F. Quatieri, and R.B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital signal processing*, vol. 10, no. 1-3, pp. 19–41, 2000.

[15] C. Barras and J.L. Gauvain, "Feature and score normalization for speaker verification of cellular data," in *ICASSP*, 2003, vol. 2.

[16] D.L. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[17] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[18] E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, V. Popovici, F. Poree, et al., "The BANCA Database and Evaluation Protocol," *Lecture Notes in Computer Science*, vol. 2688, pp. 625–638, 2009.

[19] E. Van Den Berg and M.P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.