Sparse Representation Based Classification for Face Recognition by *k*-LiMapS Algorithm

Alessandro Adamo[1], Giuliano Grossi[2], and Raffaella Lanzarotti[2]

[1] Dipartimento di Matematica, Università degli Studi di Milano
Via Saldini 50, 20133 Milano, Italy
`alessandro.adamo@unimi.it`

[2] Dipartimento di Scienze dell'Informazione, Università degli Studi di Milano
Via Comelico 39, 20135 Milano, Italy
`{grossi,lanzarotti}@dsi.unimi.it`

Abstract. In this paper, we present a new approach for face recognition that is robust against both poorly defined and poorly aligned training and testing data even with few training samples. Working in the conventional feature space yielded by the Fisher's Linear Discriminant analysis, it uses a recent algorithm for sparse representation, namely *k*-LiMapS, as general classification criterion. Such a technique performs a local ℓ_0 pseudo-norm minimization by iterating suitable parametric nonlinear mappings. Thanks to its particular search strategy, it is very fast and able to discriminate among separated classes lying in the low-dimension Fisherspace. Experiments are carried out on the FRGC version 2.0 database showing good classification capability even when compared with the state-of-the-art ℓ_1 norm-based sparse representation classifier (SRC).

1 Introduction

The face recognition problem has been widely studied in several fields, involving biological researchers, psychologists, and computer scientists. This interest is motivated by the still big disparity between the performances achieved by existing automatic face recognition systems (FRSs) [1,2] and human ability in solving this task. FRSs can be classified in local-based or holistic. The first extract local features either on the whole face [3] or in correspondence to peculiar fiducial points [4]. By construction such methods are more robust to variations caused by either illumination or pose changes. Their main disadvantages are the computational cost and the fact that they require a certain image resolution and quality, which cannot be guaranteed in real world applications. The holistic approaches are more suitable in case of low quality images considering they do not require to design and extract explicit features. The most popular are Eigenface [5], Fisherface [6] and Laplacianface [7]. More recently a new approach [8] based on the sparse representation theory [9,10] has been proposed, proving its effectiveness. This method aims to recognize a test image as a sparse representation of the training set, assuming that each object covers a certain subspace. The main disadvantage of this method, and of all the holistic approaches in general, is that

it requires a very precise (quasi-perfect) alignment of all the images both in the training and in the test sets: even small errors affect heavily the performances [11]. Besides, they require to have numerous images per subject for training and it is computationally expensive. All these characteristics are not conceivable for real world applications.

The misalignment problem has been tackled in [12,13], showing a robustness increment of these systems.

In this paper we propose a completely automatic and fast FRS based on the sparse representation (SR) method. Both the training and the test sets are pre-processed with the off-the-shelf face detector presented in [14] plus the eyes and mouth locator presented in [15]. The obtained face sub-images are projected in the Fisher space and then sparsity is accomplished applying the recently proposed algorithm k-LiMapS [16]. Such method is based on suitable Lipschitzian type mappings providing an easy and fast iterative schema which leads to capture sparsity in the face subspace spanned by the training set.

We tested out method on the FRGC version 2.0 database [17], and compared it with the SRC method. These experiments prove that, despite the system is completely automatic, it is robust with respect to mis-alignments and variations in expression or illumination.

2 Sparse Recovery by k-LiMapS

In this section, we first briefly recall the general sparse recovery (or sparse representation, SR) framework, then we detail our proposed k-LiMapS algorithm.

2.1 Sparse Recovery

The mathematical problem statement of SR consists in finding the sparsest representation of a vector $x \in \mathbb{R}^n$ given an overcomplete dictionary $\Phi = [\phi_1, \ldots, \phi_m]$ assumed to be a collection of $m > n$ atoms or vectors in $\mathbb{R}^n$. A sparse representation for x can be expressed as a linear combination of atoms, i.e., $x = \sum_i \alpha_i \phi_i$, or equivalently in matricial form

$$\Phi\alpha = x, \tag{1}$$

and is measured in terms of ℓ_0-norm $\|\alpha\|_0$, simply representing the number of non-zero elements in α. More generally, it is not sensible to assume that the available data x obeys precise equality (1) with a sparse representation $\|\alpha\|_0 = k \ll n$. A more plausible scenario assumes sparse approximate representation in which there is an ideal noiseless signal x (admitting a sparse representation) corrupted by noise, leading to the model $x = \Phi\alpha + \varepsilon$, in which error or noise $\varepsilon \in \mathbb{R}^n$ gives rise, for instances, to measurements or estimates. Adopting this noisy setting, the general goal of finding the sparsest decomposition of the signal x can be rephrased as the constrained minimization problem in ℓ_2-norm

$$\min_{\alpha \in \mathbb{R}^m} \|x - \Phi\alpha\|^2 \quad \text{subject to} \quad \|\alpha\|_0 \leq k. \tag{P_0}$$

The optimization problem (P_0) is generally NP-hard. Therefore the objective becomes to find computationally efficient algorithms that can approximately solve (P_0), keeping the purpose of recovering as sparse as possible coefficient vectors α.

2.2 k-LiMapS Algorithm

To promote sparsest solutions to the underdetermined inhomogeneous system (1) we proposed the k-LiMapS algorithm (k-Coefficients Lipschitzian Mappings for Sparsity) [16]. For a desired sparsity level $k > 0$ fixed a priori, the method iterates a parametric family of nonlinear mappings along the affine space associated to the system favoring sparse near-feasible solutions. To recover in turn admissible solutions, an alternating stage envisages the use of an orthogonal projector onto the feasible space. The process yields a Cauchy sequence in the Hilbert space ℓ_2^m for which limit point exists regardless of the initial guess. At the end of the process, depending on whether the signal under exam x admits or not a k-sparse representation, an hard thresholding operation is applied to solution $\alpha \in \mathbb{R}^m$ so that $\| \alpha \|_0 = k$.

More specifically, let $\mathcal{F} = \{F_\lambda : \mathbb{R}^m \to \mathbb{R}^m \mid \lambda \in \mathbb{R}^+\}$ the one-parameter family of nonlinear mappings promoting sparsity via near-feasible points defined by

$$F_\lambda(x) = x \odot \left(1 - e^{-\lambda |x|}\right), \tag{2}$$

where $\odot$ denotes the Hadamard product. The orthogonal projector aimed to map every point falling in the range of (2) into the nearest point in the affine space $\mathcal{A}_{\Phi,x} = \{\alpha \in \mathbb{R}^m : \Phi\alpha = x\}$, supposed not empty, is the usual projector $P = I - \Phi^\dagger \Phi$, where I is the identity operator and $\Phi^\dagger = (\Phi^T \Phi)^{-1}\Phi^T$ the Moore-Penrose pseudo-inverse of the (assumed) full-rank matrix Φ. Moreover, the point $\nu = \Phi^\dagger x$ represents the closed-form least-squares solution of system (1), frequently used as starting point of the trajectory iteratively generated.

As a consequence, by composing mapping (2) and projector P, we are given with the new mapping $G_\lambda : \mathbb{R}^m \to \mathcal{A}_{\Phi,x}$, defined as

$$\begin{aligned} G_\lambda(\alpha) &= PF_\lambda(\alpha) + \nu \\ &= \alpha - P\alpha \odot e^{-\lambda|\alpha|}, \end{aligned} \tag{3}$$

whose iterations will help in finding sparse solutions of the system. In fact, the search is accomplished by the sequence of successive approximations inductively defined by

$$\begin{cases} \alpha_0 = \alpha \in \mathbb{R}^m \\ \alpha_n = \alpha_{n-1} - P\alpha_{n-1} \odot e^{-\lambda_n|\alpha_{n-1}|}, \quad n \geq 1 \end{cases} \tag{4}$$

where the sequence of positive parameters $\{\lambda_n\}_{n\geq 1}$ is suitably defined as follows. In order to meet the severe constraint of choosing k coefficients not null and discarding the remaining $m - k$, by denoting with $\hat{\alpha}_n$ the absolute values of α_n rearranged in descending order and $\hat{\alpha}_N(k+1)$ its $(k+1)$-th element, the sequence

of parameters $\{\lambda_n\}$ is adaptively assumed to be $\lambda_n = 1/\hat{\alpha}_n(k+1)$. In this way the algorithm preserves the k most significant coefficients annihilating little by little the remaining $m - k$.

The overall algorithmic schema explained above is summarized by the pseudo-code in Algorithm 1.

Algorithm 1. k-LiMapS

Require: Projector $P = I - \Phi^\dagger \Phi$, sparsity level k, initial guess $\nu = \Phi^\dagger x$

1: $\alpha \leftarrow \nu$
2: **while** $[\, ||P\alpha \odot e^{-\lambda|\alpha|}|| > \epsilon \,]$ **do**
3:     $\sigma \leftarrow \text{sort} (|\alpha|)$ $<descending\ order\ coefficients>$
4:     $\lambda \leftarrow 1/\sigma_k$ $<sparsity\ ratio\ update>$
5:     $\alpha \leftarrow \alpha - P\alpha \odot e^{-\lambda|\alpha|}$ $<orthogonal\ projection>$
6: **end while**
7: $\alpha_j \leftarrow 0 \quad \forall\, j$ s.t. $|\alpha_j| \leq \sigma_k$ $<thresholding>$

Ensure: an approx. solution of $s = \Phi\alpha$ s.t. $\|\alpha\|_0 \leq k$

Regarding the convergence analysis, it can be stated that the sequence of iterates $\{\alpha_n\}_{n \geq 0}$ in k-LiMapS converges to a fixed point of (3), by locally minimizing problem (P_0).

3   Face Recognition via k-LiMapS

Here we show that k-LiMapS can be used in a FRS, enforcing the belief that sparsest representation is naturally discriminative. Our algorithm achieves small classification error by selecting the most representative subset of atoms belonging to the same target subject, while rejecting all other possible but less compact representations. To achieve good performances both in time and in representation quality, we choose to work into the feature space yielded by the LDA transform, briefly sketched in the next sub-section.

3.1   LDA Subspace Analysis and Image Embeddings

In a typical setting for face recognition, we face the problem of correctly determining to which class belongs a given test image among c distinct classes or subjects. Arranging data in a matrix structure, the training samples from the i-th class are represented as column vectors of the matrix $A_i = [x_1, \ldots, x_{n_i}] \in \mathbb{R}^{n \times n_i}$. The training set collecting all subjects is then obtained by stacking all matrices A_i into matrix $A = [A_1, \ldots, A_c]$.

In order to explore the feature structure of A_i, one of the most popular discriminative tool is the so called Linear Discriminant Analysis (LDA) [6] or Fisherfaces. Defining the between-class scatter matrix S_B and the within-class scatter

matrix S_W for the training set A as in [6], LDA searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. Such a projection is chosen to maximize the Fisher Discriminant Criterion, i.e.,

$$W_{\text{OPT}} = \arg\max_W \frac{|W^T S_B W|}{|W^T S_W W|}.$$

The main drawback of LDA is related to the potential singularities of the matrix S_W because the dimension of the sample space is typically larger than the number of samples in the training set. To overcame this problem, the Fisherfaces method foresees a Principal Components Analysis (PCA) [5] stage in order to project the original data on a lower dimensional space where the within-class scatter matrix may be nonsingular. Formally, the complete transform is given by

$$W_{\text{LDA}}^T = W_{\text{OPT}}^T W_{\text{PCA}}^T, \tag{5}$$

where the space is reduced to the dimension $c - 1$, which is in general much smaller than n.

3.2   FRS Based on k-LiMapS

Our FRS requires to construct a dictionary Φ on the basis of the training images (k images per subject). At first all the training samples corresponding to the c subjects to recognize are collected in a matrix A (as described in subsection 3.1). Then, applying the LDA projection in (5), the dictionary $\Phi = W_{\text{LDA}} A$ is created, being each atom a $(c-1)$-dimensional vector. Successively, in the classification stage, given a test image x, we calculate the projected sample $y = W_{\text{LDA}} x$ and then we perform k-LiMapS to find sparse vector α such that $\Phi\alpha \approx y$.

In the purpose of solving the membership i of the test image x, we look for the linear span (i.e., LDA subspace) of the training samples associated with the subject i that better approximates the feature vector y. In other words, by denoting with $\hat{\alpha}_i$ the coefficient vector whose only nonzero entries are the ones in α associated to class i, we classify y minimizing its residual with the linear combination $\Phi\hat{\alpha}_i$, i.e., applying the following rule:

$$j = \min_{i \in [1,..,c]} \|y - \Phi\hat{\alpha}_i\|.$$

The class assigned to x will be the j so found.

4   Experimental Results

We tested the proposed technique on the FRGC version 2.0 database [17]. The dataset reports images of 466 people acquired in several sessions (from 1 to 22, varying from person to person), over two periods (Fall 2003 and Spring 2004). A session consists of six images: four *controlled* and two *uncontrolled*, both acquired with either neutral or smiling face expression. Controlled images are

Table 1. The face recognition rate (%) on the FRGC 2.0 controlled, varying the cardinality. In brackets we report the number of features which brought to such percentage.

# Subj	50	100	150	200	239
$k = 3$	97.6 (100)	96.4 (180)	95.6 (200)	94.9 (340)	93.9 (360)
$k = 4$	98.4 (100)	98.3 (200)	97.0 (250)	96.9 (390)	95.4 (490)
$k = 5$	98.8 (160)	98.2 (230)	98.2 (280)	97.2 (340)	97.2 (390)

acquired in frontal pose, with homogeneous illumination, while the uncontrolled ones represent smaller faces, often blurred and acquired in several illumination conditions. For our experiments we considered only the subjects with at least three sessions per period. This brought us to 239 subjects.

All the experiments have been carried out on images automatically localized with the face detector proposed in [14] followed by the eyes and mouth locator presented in [15]. *No human intervention* is required. The misalignment we deal with is exemplified in Fig. 1.
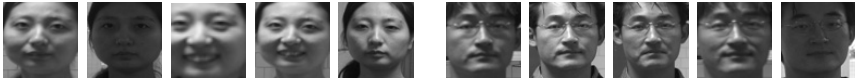
Fig. 1. Examples of automatic cropping on uncontrolled images of two subjects

Furthermore the number of images in the training set has been deliberately kept low (k varying between 3 and 5) even if the database would allow a richer subject representation. This has been done in order to emulate real world settings. The results we report have been obtained mediating over 20 experiments; at each iteration, k images are randomly selecting for training and the remaining are used to construct the test set. Comparisons have been carried out with the state-of-the-art SRC [8], with a feature space dimension equal to 100, which is a good compromise between the performances and the computational costs.

We set up several experiments[1]: first, we explored the *system scalability*: considering only the controlled images of people with neutral expressions, we tested the system performances incrementing the subjects cardinality. As shown in Table 1, the decrease of performances is more important for small values of k.

Second, we investigated how the *expression variation* influences the performances. In the first two columns of Table 2 we report the results obtained by both our algorithm and the SRC, varying k and the pool of images: either neutral or neutral and smiling. In all these experiments we considered 239 subjects. As we can see, the expression variation causes a loss of less than one percentage point for both our method and the SRC, showing a desirable invariance to the expressions.

---

[1] Matlab code of k-LiMapS used in the tests is available on the website
http://dalab.dsi.unimi.it/software/klimaps-face-recognition.tgz.

As last case, we explored the system behavior on *uncontrolled* images reporting the results in the last column of Table 2. This is the more realistic and challenging scenario, where the subjects are non-collaborative and the acquisition conditions non-optimal. In this case the performances are poorer, reflecting the challenge of the task. The low quality of these images affects the recognition percentage in two ways: first the face locator is less precise, resulting in more mis-aligned faces (see Fig. 1). Second, the feature extractor itself has to deal with less discriminative information deleted by blurring, and even misleading information caused by shadows or glasses. What we highlight however is the large gap between the performance we achieve and the SRC ones. This confirms that our method is more robust in presence of misalignment and unfavorable conditions.

Table 2. The face recognition rate (%) on 239 subjects of the FRGC 2.0 controlled, neutral versus neutral and smiling and FRGC 2.0 uncontrolled

	Neutral		Neutral and Smiling		Uncontrolled	
	k-LiMapS	**SRC**	**k-LiMapS**	**SRC**	**k-LiMapS**	**SRC**
$k = 3$	93.9 (360)	92.8	93.2 (380)	91.8	77.1 (390)	68.4
$k = 4$	95.4 (490)	95.3	94.6 (500)	94.7	82.8 (360)	74.7
$k = 5$	97.2 (390)	96.6	96.3 (460)	96.2	87.2 (380)	79.1

Finally, we make two general considerations. First, we remark that the feature space dimension is not a critical parameter for our method: the reported performances are achieved for a wide range of feature dimensions (± 200 features from the optimal); here we report the best one to give an idea of the corresponding order of magnitude. Second, to hint at computational time, it turns out that in relation to the largest feature spaces considered in the experiments (equal to 1000), *k*-LiMapS processes a test image in about 0.05 seconds, making it a very fast heuristic for face recognition.

5   Conclusions

In this work, we outline a new approach for face recognition based on the paradigm of sparsity recovery. We have experimentally shown that it produces well separated classes in low-dimensional subspaces (e.g., Fisherspace) exhibiting good performances also in case of misalignments and large variations in lighting and facial expressions. In particular it should be stressed that such a technique results particularly suitable for realistic world applications where one has to deal with not only uncontrolled conditions but also very few examples available for training purpose. Future work suggests to face up to more challenging problems like recognition of noisy or partially occluded images as well as to show the independence of any database. This would prove the applicability of our method in a wide variety of real world applications of FRSs.

References

1. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: a literature survey. ACM Computing Surveys 35, 399–458 (2003)
2. Rabia, J., Hamid, R.: A survey of face recognition techniques. Journal of Information Processing Systems 5 (2009)
3. Perez, C., Cament, L., Castillo, L.E.: Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count. Pattern Recognition 44, 951–963 (2011)
4. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. IEEE Trans. on Pattern Analysis and Machine Intelligence 19, 775–779 (1997)
5. Turker, M., Pentland, A.: Face recognition using Eigenfaces. Journal of Cognitive Neuroscience 3, 71–86 (1991)
6. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Analysis and Machine Intelligence 19, 711–720 (1997)
7. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.: Face recognition using laplacianfaces. IEEE Trans. Pattern Analysis and Machine Intelligence 27, 328–340 (2005)
8. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Analysis and Machine Intelligence 31, 210–227 (2008)
9. Donoho, D.L.: For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. Comm. Pure Appl. Math. 59, 797–829 (2004)
10. Candes, E., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. 59, 1207–1223 (2005)
11. Delac, K., Grgic, M. (eds.): Face Recognition. I-Tech Education and Publishing (2007)
12. Yan, S., Wang, H., Liu, J., Tang, X., Huang, T.: Misalignment-robust face recognition. IEEE Transactions on Image Processing 19, 1087–1096 (2010)
13. Wagner, A., Wright, J.: Toward a practical face recognition system: Robust alignment and illumination by sparse representation. IEEE Trans. Pattern Analysis and Machine Intelligence 34, 372–386 (2012)
14. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 511–518 (2001)
15. Campadelli, P., Lanzarotti, R., Lipori, G.: Precise eye and mouth localization. International Journal of Pattern Recognition and Artificial Intelligence 23 (2009)
16. Adamo, A., Grossi, G.: A fixed-point iterative schema for error minimization in k-sparse decomposition. In: Proceedings of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2011), pp. 167–172 (2011)
17. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K.: Overview of the face recognition grand challenge. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2005)