Face Recognition in Uncontrolled Conditions Using Sparse Representation and Local Features

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
[2] Dipartimento di Informatica, Università degli Studi di Milano
Via Comelico 39, 20135 Milano, Italy
{adamo,grossi,lanzarotti}@di.unimi.it

Abstract. Face recognition in presence of either occlusions, illumination changes or large expression variations is still an open problem. This paper addresses this issue presenting a new local-based face recognition system that combines weak classifiers yielding a strong one. The method relies on sparse approximation using dictionaries built on a pool of local features extracted from automatically cropped images. Experiments on the AR database show the effectiveness of our method, which outperforms current state-of-the art techniques.

Keywords: Sparse representation, face recognition, face partial occlusions, expression variations, illumination variations, local features.

1 Introduction

Nowadays face recognition (FR) techniques perform very well in controlled conditions [17] but suffer when applied in real-world contexts [13,10]. This is the frontier we want to investigate in this work. In particular we focus the attention on three aspects: illumination variations, continuous occlusions, and large expression variations [12,9]. From the literature it is well known that local-based methods behave better than holistic ones [8,16] in presence of occlusions. This consideration can be extended to both light and expression variations which affect heavily only certain face regions, leaving others less altered [12].

Following these results, we introduce a new Local-based Face Recognition (LFR) system, namely k-LiMapS_LFR relying on a pool of local features and on the sparse representation paradigm [16] for classification. In particular, we adopt the k-LiMapS sparse approximation algorithm [1,2]: an easy and fast iterative schema based on suitable Lipschitzian type mappings which allows to capture sparsity in feature subspaces spanned by suitable dictionaries.

The system we propose is tested on face images either manually or automatically cropped, in order to verify the robustness to possible misalignments. In particular, the automatic cropping is attained adopting the face detector proposed in [14], and augmenting the precision applying the eyes and mouth locator (EML) presented in [4].

A. Petrosino (Ed.): ICIAP 2013, Part II, LNCS 8157, pp. 31–40, 2013.

The proposed technique can be setup adopting any local feature. In this work we use features which have demonstrated a good discriminative capability, such as, raw patches randomly extracted (namely Random Tessellation features), LBP [3], MSLBP [5], Gabor filters [15] and HLAC [6].

The rest of the article is organized as follows: we first recall the general sparse recovery (or sparse representation, SR) framework and the elements of novelty of our method; then we briefly recall the rationale of the k-LiMapS algorithm, and finally we present extensive results obtained on the AR database [7] comparing our method versus the well known SRC [16] and PFI [11] algorithms, both on manual and automatic face cropping.

2 Sparse Representation-Based Classification

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 pseudo-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 to the 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

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

where $\| \cdot \|$ denotes the ℓ_2-norm.

In [16] it has been proposed the SRC algorithm, a pioneering work adopting SR as FR system. Here we present a local FRS, namely the k-LiMapS_LFR, which constructs a set of dictionaries based on local features, and combines these weak classifiers via the majority vote rule. The idea is to characterize faces in the training set with a large pool of independent local features, in order to have a sufficiently rich face description even in presence of occlusions or deformations caused by strong expression variations. The k-LiMapS_LFR system requires as input face images at least roughly localized. To this end, in the case of automatic cropping, the images are preprocessed by the face detector proposed in [14], and the eyes and mouth detector (EML) presented in [4]. The EML, besides the primary functionality of improving the face alignment, reveals possible occlusions as mentioned below. In the following we describe our method.

Suppose we are dealing with c different classes or subjects, labelled $1, \ldots, c$, and there are exactly k training samples for each subject $s \in \{1, \ldots, c\}$. Each training image is characterized applying a pool of features, $f_1, \ldots, f_d$, to local subimages or patches cropped around to a fixed set of randomly selected pixels $p_1, \ldots, p_h$, thus obtaining $d \times h$ dictionaries. According to the i-th feature applied in the j-th pixel, the dictionary $\Phi_{i,j} \in \mathbb{R}^{n \times k\,c}$ (with $n < k\,c$) represents the entire training set (consisting of k images for each of the c subjects) obtained by stacking the feature-vectors as columns:

$$\Phi_{i,j} = \left[f_i^{(1,1)}(p_j), \ldots, f_i^{(c,k)}(p_j) \right], \qquad i = 1, \ldots, d, \quad j = 1, \ldots, h.$$

For a given test image I, the FR system extracts the corresponding local features

$$z_{i,j} = f_i(p_j), \qquad i = 1, \ldots, d, \quad j = 1, \ldots, h,$$

and performs the k-LiMapS algorithm (see next section) to find the sparse vector α such that $\Phi_{i,j}\,\alpha \approx z_{i,j}$. In the purpose of solving the membership $\chi_{i,j}$ of the local feature $z_{i,j}$, the algorithm looks for the linear span of the training samples in $\Phi_{i,j}$ associated with the subject $s \in \{1, \ldots, c\}$ that better approximates the feature vector $z_{i,j}$. In other words, by denoting with $\hat{\alpha}_s$ the coefficient vector whose only nonzero entries are the ones in α associated to class s, it classifies $z_{i,j}$ minimizing its residual with the linear combination $\Phi_{i,j}\hat{\alpha}_s$, i.e. applying the following rule:

$$\chi_{i,j} = \underset{s \in \{1,..,c\}}{\operatorname{argmin}} \|z_{i,j} - \Phi_{i,j}\hat{\alpha}_s\|, \qquad i = 1, \ldots, d, \quad j = 1, \ldots, h.$$

The final classification of I is obtained applying the majority vote rule over all local classifier results $\chi_{1,1}, \ldots, \chi_{d,h}$. In Algorithm 1 we sketch the whole algorithm.

3 *k*-LiMapS Rationale

To solve the underdetermined inhomogeneous system (1) our FRS applies the k-LiMapS algorithm (k-Coefficients Lipschitzian Mappings for Sparsity) proposed in [1], which has demonstrated both its efficacy and its low computational costs. Briefly, for a desired sparsity level $k > 0$ fixed a priori, the method iterates a parametric family of nonlinear shrinking mappings along the affine space $\mathcal{A}_{\Phi,x} = \{\alpha \in \mathbb{R}^m : \Phi\alpha = x\}$, associated to the system favoring sparse near-feasible solutions. To recover in turn admissible sparse solutions, an alternating stage envisages the use of an orthogonal 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, onto the feasible space. The process yields a Cauchy sequence $\{\alpha_t\}_{t \geq 1}$ 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 $\|\,\alpha\,\|_0 = k$.

Algorithm 1. k-LiMapS_LFR

Dictionary construction:

Require: c subjects, k training images per subject, d local features

1: Randomly generate h key points
2: **for all** Training set images I_t **do**
3: EML(Viola-Jones(I_t))
4: **for all** Key points p_j and Features f_i **do**
5: Compute $f_i^t(p_j)$
6: **end for**
7: **end for**
8: Dictionaries: $\Phi_{i,j} = \left[f_i^{(t)}(p_j) \right], \quad t = 1, \ldots, c \cdot k$

Testing phase on image I:

1: EML(Viola-Jones(I))
2: Feature extraction: $z_{i,j} = f_i(p_j), \qquad i = 1, \ldots, d, \quad j = 1, \ldots, h$
3: Sparse solution α via k-LiMapS algorithm (see next session)
4: Local classifications: $\chi_{i,j} = \text{argmin}_{s \in \{1,..,c\}} \|z_{i,j} - \Phi_{i,j} \hat{\alpha}_s\|$.
5: Majority vote on $(\chi_{1,1}, \ldots, \chi_{d,h})$

To better understand the computation of the final solution, the pseudo-code of k-LiMapS is reported in Algorithm 2.

Two aspects affecting k-LiMapS performance deserve to be discussed: how to tune the sparsity level k in problem (P_0) and how to define the stop condition of the while loop of the algorithm. Guided by empirical evidence on the performance assessment, we faced the first question by fixing k equal to the number of images per subject in the training set. In this way very few coefficients are preserved while the most part is discarded as shown in Figure 1 *(left)*, where the absolute values of the coefficients are plotted in descent order (blue saved and red discarded).

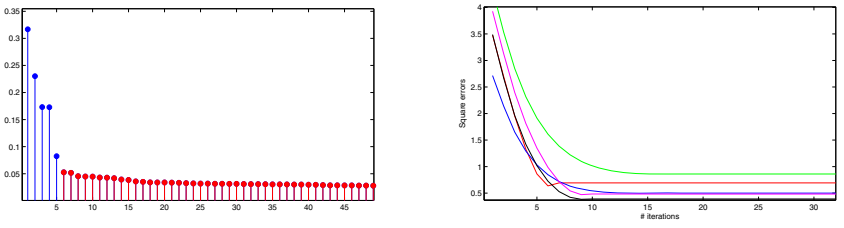

Fig. 1. *(Left)* Absolute values of the coefficients in descent order with sparsity constraint $k = 5$. The first five coefficients (blue stems) are preserved while the remaining (red stems) are discarded. *(Right)* Approximation error in ℓ_2-norm referred to some sparse solutions during the iterations of k-LiMapS.

Algorithm 2. k-LiMapS

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

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

Ensure: An approx. solution $s \approx \Phi\alpha$ s.t. $\| \alpha \|_0 \leq k$

Secondly, a suitable stop condition for the SR process may be grasped observing the minimization of the least squares objective function defined in (P_0), under the sparsity constraint forced by the ℓ_0-pseudo norm. In Figure 1 *(right)* the error variations together with the number of iterations are drawn for some test images. The graphic captures quite well the typical behavior of k-LiMapS as minimizer for the problem, highlighting that the local minimum is reached within very few iterations of the while loop at the heart of Algorithm 2. This suggests a stop condition for such a loop: end when no cost function reduction is attained.

Notice that the stop condition also influences the time complexity of the algorithm. Although a fair rule to stop the iterative process does not exist, using the above stop condition in the most part of the cases the number of iterations drops to few instances giving a good approximation.

4 Experimental Results

In this section we present the experiments on the k-LiMapS_LFR[1]. The system could work referring to any pool of local features. In this work we choose five features aiming at capturing a wide variety of image information useful in the FR task: the LBP and its generalization, the MSLBP, are good texture descriptors, while the Gabor filters and the HLAC features capture salient edges and characteristic shapes. In addiction to these well justified filters, we also use simple raw data consisting in squared patches (RT) centered in each chosen point and normalized, aiming at reducing possible illumination problems.

The experiments refer to the AR database [7] which consists of face images of 126 subjects (70 men and 56 women) acquired in two sessions, each one varying 13 different conditions covering both illumination and expression variations as well as occlusions caused by either scarves or sunglasses. The experiments have been conducted referring to a pool of 100 subjects (50 men and 50 women), averaging the results over 50 trials so guaranteeing a high confidence level.

[1] MATLAB code of k-LiMapS_LFR and all tests done are available on the website `http://dalab.di.unimi.it/klimaps`.

For comparison purpose, we run both SRC [16] and PFI [11] algorithms. The first represents the state-of-the-art in the sparsity framework, while the second uses a large feature set extracted locally on each image. The SRC algorithm has been developed to manage possible occlusions, so we apply it directly. On the contrary, the PFI algorithm has been proposed for non occluded faces, so, in order to apply it in the occluded case, we discard the corrupted half face (the lower half for scarf and the upper half for sunglasses).

We setup four kinds of experiments described below. In all cases, the training sets consist of non occluded images, with $k = 5$ (the number of images for subject in training). This is a tradeoff between the necessity of representing subjects in several conditions (potentially constructing a complete base) and the requirement to keep k small in order to emulate realistic scenarios.

Sunglasses. Training set: all images of subjects acquired in the first session with different illumination conditions, with neutral or smiling expression and no occlusion (labelled as either 1, 2, 5-7). Test set: subjects wearing sunglasses and acquired in different illumination conditions (AR images labelled as 8-10, 21-23).

Scarf. Training set: as in the previous experiment, that is all images labelled as either 1, 2, 5-7. Test set: subjects wearing scarves and acquired in different illumination conditions (images labelled as 11-13, 24-26).

Illuminations. Training set: as in the previous experiment, that is all images labelled as either 1, 2, 5-7. Test set: images with no occlusion corrupted by strong illumination variations (images labelled as 14, 15, 18-20).

Expressions. Training set: all images of subjects acquired in the first session, having different facial expressions (labelled as either 1-4 or 7). Test set: subjects showing strong expression variations and acquired in the second session (images labelled as 15-17).

These kinds of experiments involve many issues. In particular, our investigation has been guided by the following questions: "Which is the most critical condition compromising the FRS performance?". "Does the feature aggregation increases the recognition rate?" and, if yes, "Which is the best combination?". Secondly: "How much is the amount of inspected local patches influential?" and finally "How hard is to deal with the misalignment?".

To answer to these questions, we tested the system behavior increasing both the number of features from 1 to 5 (considering all possible feature combinations), and the number of characterized points (specifically with $N = 50, 100, 150,$ 200 randomly selected in non occluded regions), and we run all the experiments on both manual and automatic cropped images.

In Fig. 2 and Fig. 3 we report the performance obtained on manually and automatically cropped images respectively. In order to highlight the dependency on the feature pool cardinality and at the same time keeping the results concise, we depict for each feature cardinality the best performance we obtained, together with the performance of both the SRC and the PFI algorithms (which are obviously both independent of the value N).

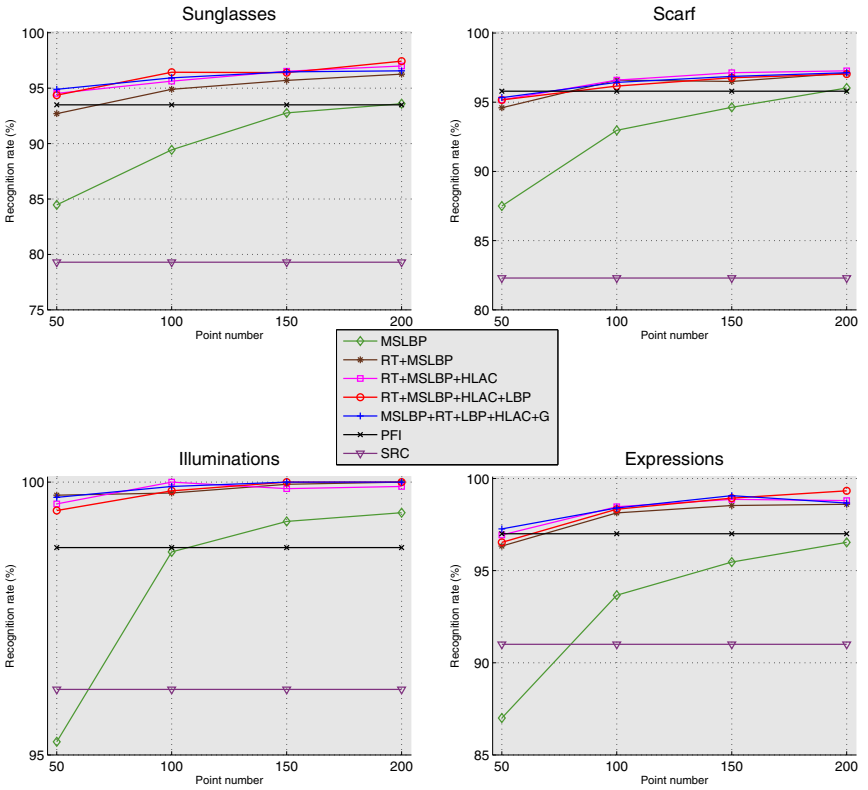

Fig. 2. Recognition rates achieved by the k-LiMapS_LFR varying the pool of features, and by the SRC and the PFI, all considering **manually** cropped images

Such results allow us to draw some conclusions. Firstly, we notice that the FR task is particularly hard when the eyes are occluded, revealing the high relevance of such face portion. Secondly, it is evident that the feature aggregation helps the recognition system. In particular, the highest performance enhancement is obtained passing from one to two features. Limiting ourselves to the five considered features, the most characterizing one has turn out to be the MSLBP, while the best pair corresponds to (MSLBP, RT).

Concerning the choice of the best feature set among all proposed of the same cardinality, we also notice a desirable property: by adding a new feature to all l-length subsets (with $l \in \{1, \ldots, 4\}$), the best pool of size $l + 1$ includes the previous best pool of size l, so exhibiting good stability and an incremental way to extend the system. For instance, in the case of manual alignment, passing from the best pair (MSLBP, RT) to the best triplet, produces the pool (MSLBP, RT, HLAC), that maintains the best couple.

Naturally, another key parameter conditioning the performance is number of points or patches N referred by the system. Besides the obvious proportionality

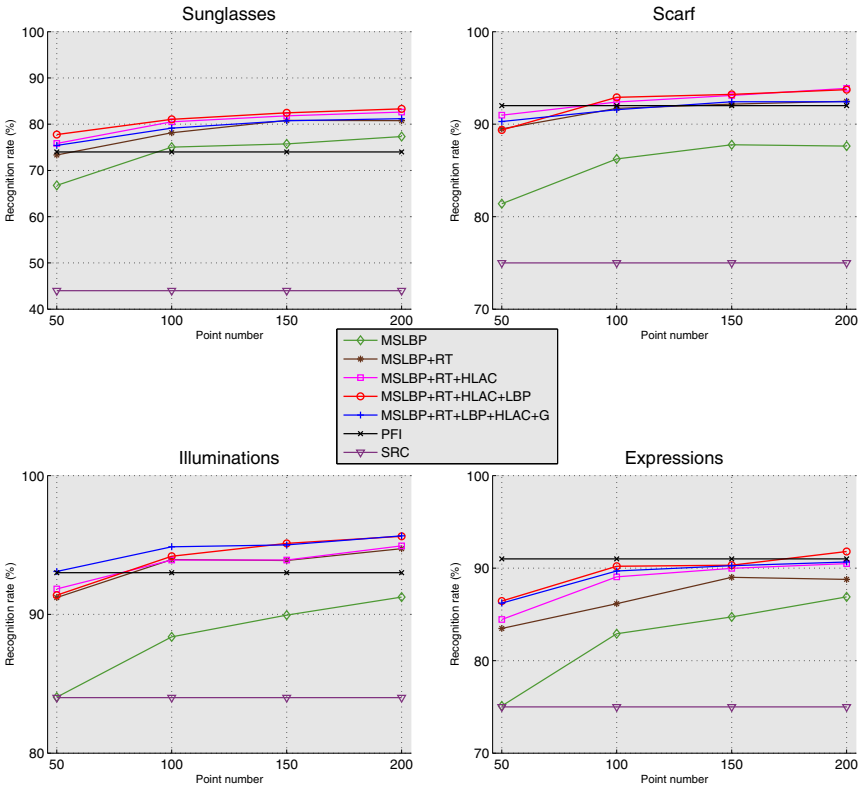

Fig. 3. Recognition rates achieved by the k-LiMapS_LFR varying the pool of features, and by the SRC and the PFI, all considering **automatically** cropped images

between N and the recognition rate (the denser the face description, the better the performance), we also notice an inverse relationship between N and the standard deviation of the recognition rate, so providing a good stability according to the increasing of N.

All these considerations cannot be regarded irrespectively of the computational costs: each increment of either the number of features or the number of points N increases the computational time. This implies a tradeoff between performance and speed. As far as the computational time of the k-LiMapS algorithm is concerned, in case of one feature and $N = 50$ a test image is processed in about 0.05 seconds on a Intel® Core™ i5 processor at 64 bit, with 3.6 GHz, and 8 Gb of memory. Naturally, the computational time scales proportionally increasing either N or the number of features. Furthermore we remark the independence of both the weak classifiers and the local features, which would made it possible to evaluate in parallel the weak classifiers merging the results subsequently.

Another comment concerns the comparison with the SRC and the PFI systems. As can be seen in the graphs, k-LiMapS_LFR outperforms SRC even with a small value of N and adopting only one feature. Regarding the comparison with PFI, k-LiMapS_LFR behaves better when setting $N = 100$ or more, and adopting at least two features. The only exception is the experiment referring to subjects varying their expression, where the PFI behaves very well. In this case the k-LiMapS_LFR obtains the same performance only with $N = 200$ and four features.

Finally, let us notice that all the systems worsen their performance passing from manual to automatic cropped images. In particular the average loss is of about the 17% for the SRC, the 9% for the PFI, and the 7% for the k-LiMapS_LFR, showing its greater robustness with respect to this critical aspect.

5 Conclusions

In this paper a new local-FRS has been illustrated. This approach allows to cope with local alteration of the face images, due to either partial occlusions or illumination or expression variations. The system refers to a pool of local features aiming at extracting most of the peculiar uncorrupted information and uses it to define a pool of weak classifiers. The final recognizer is obtained combining the weak classifier via the majority rule. The discriminative strategy of the weak classifier is committed to the sparse recovery paradigm which has recently turned to be successfully applied in face recognition.

Experimental results prove the system effectiveness and robustness, above all when compared with the state of the art in this field. The encouraging results motivate us to investigate this topic furthermore. In particular, we are interested in exploring some selection rules which would allow to maintain the same performance retaining only a fewest set of dictionaries.

References

1. Adamo, A., Grossi, G.: A fixed-point iterative schema for error minimization in k-sparse decomposition. In: Proceedings of the 2011 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2011, pp. 167–172. IEEE Computer Society (2011)
2. Adamo, A., Grossi, G., Lanzarotti, R.: Sparse representation based classification for face recognition by k-liMapS algorithm. In: Elmoataz, A., Mammass, D., Lezoray, O., Nouboud, F., Aboutajdine, D. (eds.) ICISP 2012. LNCS, vol. 7340, pp. 245–252. Springer, Heidelberg (2012)
3. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. Proc. Eur. Conf. Comput. Vis., 469–481 (2004)
4. Campadelli, P., Lanzarotti, R., Lipori, G.: Precise eye and mouth localization. International Journal of Pattern Recognition and Artificial Intelligence 23(3) (2009)
5. Chan, C.H., Kittler, J., Messer, K.: Multi-scale local binary pattern histograms for face recognition. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 809–818. Springer, Heidelberg (2007)

6. Liu, F., Wang, Z., Wang, L., Meng, X.: Facial expression recognition using hlac features and wpca. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784, pp. 88–94. Springer, Heidelberg (2005)
7. Martínez, A., Benavente, R.: The ar face database. Tech. Rep. 24, Computer Vision Center, Bellatera (June 1998), http://www.cat.uab.cat/Public/Publications/1998/MaB1998
8. Martinez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Trans. Pattern Analysis and Machine Intelligence 24(6), 748–763 (2002)
9. Naseem, I., Togneri, R., Bennamoun, M.: Linear regression for face recognition. IEEE Transactions on Pattern Analysis and Machiner Intelligence 32(11), 2106–2112 (2010)
10. Rabia, J., Hamid, R.A.: A survey of face recognition techniques. Journal of Information Processing Systems 5 (2009)
11. Schwartz, W.R., Guo, H., Choi, J., Davis, L.S.: Face identification using large feature sets. IEEE Transactions on Image Processing 21, 2245–2255 (2012)
12. Tan, X., Chen, S., Zhou, Z., Liu, J.: Face recognition under occlusions and variant expressions with partial similarity. IEEE Transactions on Information Forensics and Security 4(2), 217–230 (2009)
13. Tolba, A., El-Baz, A., El-Harby, A.: Face recognition: A literature review. Int. J. Signal Process. 2, 88–103
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. Wiskott, L., Fellous, J.M., Krüger, N., Malsburg, C.V.D.: Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 19, 775–779 (1997)
16. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 31(2), 210–227 (2009)
17. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys 35(4), 399–458 (2003)