A Selection module for large-scale face recognition systems

Giuliano Grossi, Raffaella Lanzarotti, and Jianyi Lin

Dipartimento di Informatica, Università degli Studi di Milano
Via Comelico 39/41, Milano, Italy
{grossi,lanzarotti,lin}@di.unimi.it

Abstract. Face recognition systems aimed at working on large scale datasets are required to solve specific hurdles. In particular, due to the huge amount of data, it becomes mandatory to furnish a very fast and effective approach. Moreover the solution should be scalable, that is it should deal efficiently the growing of the gallery with new subjects. In literature, most of the works tackling this problem are composed of two stages, namely the selection and the classification. The former is aimed at significantly pruning the face image gallery, while the latter, often expensive but precise, determines the probe identity on this reduced domain. In this article a new selection method is presented, combining a multi-feature representation and the least squares method. Data are split into sub-galleries so as to make the system more efficient and scalable. Experiments on the union of four challenging datasets and comparisons with the state-of-the-art prove the effectiveness of our method.

1 Introduction

Face Recognition (FR) problem has been largely studied during the last decades [24], and the more it was investigated the more challenging developments were wished for. FR systems (FRSs) have been interested by progressively less controlled acquisition conditions concerning illumination, image quality and indoor/outdoor acquisition [17]. Also the constraints on the face appearance have been more and more relaxed: at present, a FRS is desired to work well on faces with different expressions and poses [6,1] and even having partial occlusions [13,22]. A last challenge recently risen concerns with large-scale datasets: high interest is devoted to systems able to deal with thousands of subjects [7,8], as it could happen either in security contexts [18,25] or in retrieval applications [21,23,16]. As deepened in section 2, several works have already been proposed to tackle this hurdle. Many of them are based on a two-stage scheme, consisting in a first *selection* of a subset of subjects, followed by the proper *classification* of the test image. Clearly, the behaviour of the selection step, strongly influences the system performances both in terms of recognition rate and computational costs, thus deserving an in-depth analysis.

In this paper we propose a new selection method, namely *fg-LS*, based on the least squares (LS) technique, and characterized by two key attributes: it works on

multi features (f), and organizes data in sub-galleries (g). These characteristics together guarantee effectiveness, robustness and efficiency. Experiments show that *fg-LS* outperforms both baseline and state-of-the-art methods.

The remainder of the paper is organized as follows. Section 2 summarizes related works, section 3 presents the selection method we propose. The experimental analysis is detailed in section 4, while section 5 draws some conclusions.

2   Related works

Two-stage FRSs apply subsequently the *selection* and the *classification* steps. The former aims at strongly shrinking the gallery in order to preserve the target subject with high confidence. Notice that, dealing with huge data, it is desirable that the selector be computationally efficient. The classification step determines the identity of the probe image referring to the reduced gallery and adopting a technique such as nearest neighbour (NN), Support Vector Machines (SVM), or Sparse Representation (SR). In the following FRSs are categorized according to the adopted selection technique.

A first category worth to be mentioned concerns the systems adopting the clustering method for the selection stage, aiming at implementing the *divide et impera* paradigm. In such approaches the image representation is crucial: raw data are often projected in some convenient feature space (PCA, LDA, ...) thus reducing noise, attenuating misalignment problems or better discriminating the subject classes. For example, in [11] the k-means clustering applied on PCA projected data is adopted to partition the gallery subjects according to a maximum-similarity criterion. Given a test image, the method searches for the most similar cluster, and then within it, the identity of the probe image. In [9] the authors propose an iterative approach that subsequently reduces the search space. The method repeats a LDA-projection of both the training and test data, a k-means clustering to partition the search space into K clusters, and a nearest-neighbour classifier to select K' clusters closest to the probe image. The process ends when $K' = 1$. Lu et al. in [12] partition the subjects into K maximal-separability clusters by a LDA-like technique, and adopt a two-level nearest-neighbour classifier to select firstly K subjects, one from each cluster, and attaining on them the final classification decision. Although the use of both the clustering and the dimensional reduction are very pertinent in case of large-scale databases, they both require to reorganize the gallery for each incoming of new subjects, making the training process too cumbersome [23].

A second category of selector approaches consists in applying the K-Nearest Neighbor (K-NN) method to select the training samples closer to a test image. In [10,14] K-NN is adopted to reduce the search domain for the subsequent l1-solvers. In [25] the K-NN selector is followed by the nonnegative SR classifier. These approaches proved good performances and remarkable speed-up.

Very recently Ortiz and Becker [16] proposed the adoption of the speed least-squares approach to reduce the search gallery to feed to a SR classifier. The

authors claimed that linear regression (LR) approximates l1-minimization better than K-NN.

As an example of one-stage face identification on large datasets, it is worth to be mentioned the one proposed in [18]. It is based on a large set of feature descriptors using partial least squares (PLS) to perform multichannel feature weighting.

3   The *fg-LS* selection method

Given a huge gallery of face images, the selection method presented here aims at "pruning" a large part of subjects in order to make the classification stage more effective on a rather small gallery. The method relies on highly discriminative features projected on LDA subspaces. Residual errors, obtained by the least squares technique, are used for deciding which subjects to make boiling up to the classifier. As an automatic approach, all face images are cropped and normalized by means of the landmark localization method presented in [4], and then characterized extracting a pool of features $\mathcal{F}$. This characterization could be attained considering for instance simple raw data or a single feature only. However it has been proven that multi-features enrich the data description allowing one to capture more information [19,18,16]. We extend this consideration also to the illumination correction methods, each one being able to extract different discriminative characteristics from images acquired in various illumination conditions. Thus, we resort to a combination of illumination corrections (IC) and feature extractors (FE), so that $\mathcal{F} = \text{IC} \times \text{FE}$ is a collection of transformations $f^1, f^2, ..., f^J$, $J = |\mathcal{F}| = |\text{IC}| \cdot |\text{FE}|$, where each f^j maps an image I to the corresponding illumination-corrected feature vector $f^j(I) \in \mathbb{R}^{n_j}$.

Consider a large set of reference images $\mathcal{G} = \{I_1, \ldots, I_m\} \subset \mathbb{R}^N$ corresponding to the subjects (or classes) $\mathcal{C} = \{1, ..., c\}$. Based on a divide et impera approach, the main idea is to work on non-overlapping sub-galleries $\mathcal{G}_1, \ldots, \mathcal{G}_S \subset \mathcal{G}$. The first step of preprocessing consists in the construction of the matrix for the s-th sub-gallery in the j-th feature space by concatenating the feature column vectors: $G_s^j = [f^j(I)]_{I \in \mathcal{G}_s}$, $s = 1, ..., S$, $j = 1, ..., J$.

To perform dimensionality reduction while preserving as much as possible the class discriminative information, we apply the LDA projection. For each matrix G_s^j (i.e. each sub-gallery and feature) we construct the LDA projection matrix W_s^j, so maintaining lower computational costs (dominated by generalized eigenvalue problems) w.r.t. considering the gallery $\mathcal{G}$ all at once. Hence, the reference sub-galleries in LDA space are $\Phi_s^j = W_s^j G_s^j$, whose columns are feature-vectors in LDA space $\phi_s^j = \phi_s^j(I) = W_s^j f^j(I)$, for $I \in \mathcal{G}_s$.

Now, given a test image T, the same preprocessing is applied, locating and normalizing the face, extracting and projecting the features $f^j(T)$ onto the LDA spaces corresponding to every sub-gallery, hence obtaining $\tau_s^j = W_s^j f^j(T)$.

The selection is attained by evaluating the least squares residuals. Given a subject $i \in C$ in the s-th sub-gallery ($s = s(i)$ uniquely determined by i), let us denote the set of columns in sub-gallery Φ_s^j belonging to i by $\Phi^j(i)$. We then

compute the ℓ_2-distance between the test image vector τ_s^j and the column-space of $\Phi^j(i)$ associated to every subject $i \in \mathcal{C}$:

$$R_i^j = \min_x \|\tau_s^j - \Phi^j(i)x\|_2 = \|\tau_s^j - \Phi^j(i)[\Phi^j(i)]^\dagger \tau_s^j\|_2 \qquad \text{for } j = 1, ..., J$$

where $A^\dagger$ is the Moore-Penrose pseudo-inverse of a matrix A. Since these distances are calculated in several inhomogeneous spaces, we perform a z-score standardization z_i^j of R_i^j (over all R_h^j, h appearing in sub-gallery s) for each feature j; we then combine them into one residual $r_i = \sum_{j=1}^J z_i^j$ for every subject $i \in \mathcal{C}$.

Final selection is obtained gathering the K subjects in $\mathcal{C}$ with smallest residuals. We summarize this process in Alg. 1.

Algorithm 1: fg-LS

Input: Image gallery $\mathcal{G} = \{I_1, ..., I_m\} \subset \mathbb{R}^N$; test image T; integer K
Output: Subjects $i_1, ..., i_K \in \mathcal{C}$ with minimum residual for T
- Gallery construction (off-line):
Preprocessing of $\mathcal{G}$: cropping and normalization
Sub-galleries construction: $\mathcal{G}_1, ..., \mathcal{G}_S$
for $s = 1, ..., S$ **do**
    for $j = 1, ..., J$ **do**
        Computation of feature vectors: $f^j(I)$, for all I in $\mathcal{G}_s$
        Matrix construction: $G_s^j = [f^j(I)]_{I \in \mathcal{G}_s}$
        Compute LDA transformations: W_s^j
        Galleries in LDA spaces: $\Phi_s^j = W_s^j G_s^j$
    end
end
- Testing phase:
Preprocessing of T: cropping and normalization
for $j = 1, ..., J$ **do**
    Computation of feature vectors: $f^j(T)$
    LDA projections: $\tau_s^j = W_s^j f^j(T)$, $s = 1, ..., S$
    for $i \in \mathcal{C}$ **do**
        Computation of submatrices $\Phi^j(i)$ and pseudo-inverses $[\Phi^j(i)]^\dagger$
        Residuals: $R_i^j = \|\tau_s^j - \Phi^j(i)[\Phi^j(i)]^\dagger \tau_s^j\|_2$
    end
    for $i \in \mathcal{C}$ compute $z_i^j = $ z-score(R_i^j) over subjects h in gallery $s(i)$
end
Cumulative residuals: $r_i = \sum_{j=1}^J z_i^j, i \in \mathcal{C}$
Sort $\{r_1, ..., r_c\}$
Return indices $i_1, i_2, ..., i_K$ of the smallest residuals $r_{i_1}, r_{i_2}, ..., r_{i_K}$

---

4 Experimental analysis

The experiments reported in this section show that the idea of filtering out a substantial number of subjects from a huge gallery, while preserving the target

one with high confidence, plays a relevant role in FR task. This is mainly due to the fact that the most effective classification techniques are complex in nature while being highly discriminative when dealing with small galleries.

To show the robustness of the selector, we carry out experiments in challenging conditions, referring to a large pool of subjects, poorly represented in the gallery, and acquired in uncontrolled conditions, including variations in the environmental conditions (lighting, clutter background), variations of the face expression, and variation in the quality of the acquisition (focus/blurred).

4.1 Datasets

We collect the images of four public available databases acquired in uncontrolled conditions, thus obtaining a pool of $c = 725$ subjects [1], and thousands of images. For each experiment we randomly select 3 images per subject for the gallery construction, while the remaining images are used for testing. Specifically, we refer to the Extended Yale B (frontal), the BANCA Avderse, the FRGC v.2 Uncontrolled, and the Multi-PIE (frontal). In Table 1 we synthesize their peculiar characteristics, and in Fig. 1 some examples of face images are shown.

Table 1. Databases and their characteristics: **N. sbj**, **N. Images**, **Background**, **Illumination** (*varies*: oriented light, *good*: homogenous light, *poor*: image underexposure), **Expression** (*varies*: different face expressions, *reading*: reading subjects, *neutral*), **Timing** (*no*: single acquisition section, *yes*: several sessions spanning over several months), **Img quality** (*good*: high resolution, focused images, *bad*).

Database	N. sbj	N. Images	Background	Ill.	Expr.	Timing	Quality
Ext. YaleB (frontal)	38	2432	homogeneous	varies	neutral	no	good
BANCA Adv.	52	2.080	clutter	varies	reading	yes	bad
FRGC v.2 Uncontr.	289	5.248	clutter	varies	varies	yes	bad
Multi-PIE (frontal)	337	≈ 50.000	homogeneous	varies	varies	yes	good

4.2 Feature extraction

The method requires the image preprocessing consisting in face detection and normalization of both the gallery and test images. This is accomplished by applying the method presented in [4] that allows us to crop the portion of the image corresponding to the face precisely, and to rescale it to 80x70 pixels (Fig. 1).

The pool of features corresponding to IC and FE has been chosen as a trade-off between performances and computational costs. In particular we adopt a

---

[1] we consider only those subjects with at least four images correctly, or at least approximately localized

Fig. 1. Examples of original images automatically cropped. Images show some of the possible hurdles present in the database: variations in expression, pose, lighting, and focus.

pool of three illumination corrections, that is the linear stretching, the MSQ (Multi-Scale Quotient) and ASSR (Adaptive Single-Scale Retinex) techniques [20], aiming at solving various illumination problems (low contrast, strong shadows, over/under-exposure). Also the FE pool consists of three methods, that is the HoG (Histogram of Oriented Gradients) [5], Shearlet [3] and MSLBP (Multi-Scale Local Binary Pattern) features [15], each one able to catch a specific image property, that is the discontinuities, the granularity, and the texture respectively. This choice leads to nine combined-feature spaces.

The parameters used in illumination correction and feature extraction have been set as follows. In MSQ we adopt Gaussian filters with four standard deviations in the range $\sigma \in (1, 1.6)$. In ASSR the number of iterative convolutions is set to 10 and the weights needed for the filter are $\delta = 10e^{-\mathbb{E}_{x,y}[|\nabla I(x,y)|]/10}$ (that is based on the expected value of the image gradient), and $h = 0.1e^{-10\tau}$, with τ being the normalized average of local intensity differences. Concerning the HoG features, we refer to 15×15 patches, concatenating the obtained 8-bin histograms. In MSLBP we maintain the same window and histogram sizes as in HoG, setting the circle radius equal to $1, 3, 5$, respectively. The shearlet feature has been implemented using the Meyer-type filter, and adding together the detail coefficients, while excluding the first scale (low frequencies).

4.3   Tests on the fg-LS and other assessed methods

The fg-LS method essentially depends on two parameters, the first is the number $M = [c/S]$ of subjects per sub-gallery, and the second is the number $K \in \{1, \ldots, c\}$ of subjects yielded by the selector. We analyse their impact on the selector setting-up two experiments. The first aims at studying the behaviour of fg-LS method varying M, the latter addresses a comparison with some assessed methods varying K. In addition we carry out a last experiment aiming at highlighting the effects of our selector on a complete FR system.

Experiment 1. We run fg-LS varying M between 50 and 725, the last being the number c of subjects available in the referred dataset. In Fig. 2 we report for each tested M the performances achieved by the system in terms of the percentage ρ of presence of the target subjects in the selected pools. Formally, given the test images $T_1, ..., T_n$ with corresponding labels $\ell_1, ..., \ell_n \in \mathcal{C}$, and denoting by $\mathcal{C}_T = \text{fg-LS}(\mathcal{G}, T, K)$ the output of the selector, such percentage is computed as

$$\rho = \frac{\#\{i : \ell_i \in \mathcal{C}_{T_i}, i = 1, ..., n\}}{n}.$$

From these graphs, we can observe that the system behaves better when the sub-galleries have intermediate sizes. Specifically to these experiments, sub-galleries with M between one and two hundreds of subjects (about 14-28% of all subjects) behave better than adopting bigger or smaller values for M. Recalling that the operations carried out at sub-gallery domain concern the LDA projection and the z-score standardization, these results highlight the opportunity to work in a sufficiently expressive domain. Indeed, adopting big values of M restricts the LDA capability of separating the different classes, while too small M's would make the standardized residuals too uniform across sub-galleries (for instance think about the limit $M = 1$; in this case all the cumulative residuals r_i^j would be very concentrated), making their comparison meaningless.
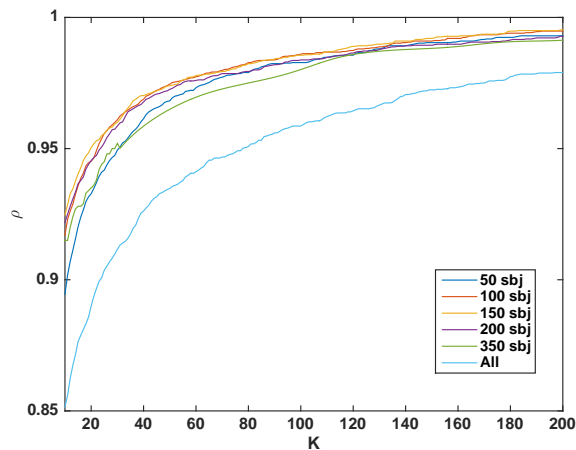
Fig. 2. Experimental analysis of the fg-LS method for different value of M (number of subjects per sub-gallery, 'sbj' in the legend) varying the selector parameter K. Performances are expressed by the percentage ρ of presence of the target subjects in the selected pools.

Concerning the computational cost of the online testing phase, the dependence of the time consumption on the gallery size (Fig. 3) is mainly due to the cost of multiplication with the LDA transform matrices W_s^j, making it convenient referring to small sub-galleries. As a trade-off between performances and computational costs, for the subsequent experiments, we set $M = 150$.
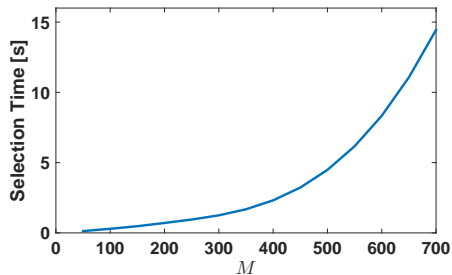

Fig. 3. Computational time (in seconds) of the fg-LS selection method, varying M.

Experiment 2. For comparisons, we set up an experimental analysis aiming at evaluating the performances of different selectors varying the number of selected subjects K. In particular we assess two baseline methods, namely the K-NN and LS approaches, both projecting data into the LDA space, and two state-of-the-art methods, namely the f-LS [16] and the f-PLS [18]. Both these systems represent data in a multi-feature space, varying the selection criterion: in f-LS the linear regression is applied on the PCA projection of the concatenation of LBP, HoG, and Gabor features. In f-PLS the features are extracted locally, and then the PLS (Partial Least Squares) is applied to derive a weighted model for each subject.

In Fig. 4 we plot for each assessed method the percentage ρ of presence of the target subjects in the selected pools, varying K over a challenging range. The results prove the effectiveness of the LS approach, above all if associated to multi-feature representation. In particular our selector, fg-LS, exceeds the more sophisticated f-PLS method that requires a training for each subject, simply combining the multi-feature representation with the LDA projection and the partitioning into sub-galleries.

Experiment 3. Aiming at highlighting the selection effects on a complete FR system, we adopt the FR classifier we presented in [2], namely k-LiMapS, that is a sparsity-based classifier. Note that in principle any other could be exploited (e.g. NN, LS, SRC, ...). In Fig. 5 we report the results obtained varying K. The blue line corresponds to the selector, plotting the percentage ρ of success achieved setting $M = 150$, as explained above. The curve corresponding to the classifier (red line) expresses the success ratio of the k-LiMapS when applied onto subsets correctly produced by fg-LS of K subjects, that is when they include the target one. So doing we break down the possible errors between
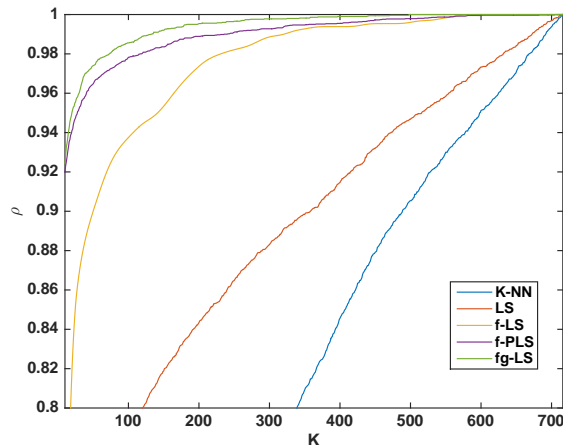
Fig. 4. Comparison of different selectors, varying the parameter K. Specifically, K-NN, and LS are well-known baseline methods, f-LS corresponds to the method presented in [16], f-PLS to the one introduced in [18], and fg-LS is ours. Performances are expressed by the percentage ρ of presence of the target subjects in the selected pools.

the selector and the classifier. As already argued, the selector increases its percentage of success when increasing K. Contrarily, the classifier takes advantage of the pruning, showing a rapid drop of performance with the increase of the selected pool size. The overall recognition ratio (yellow line) is obtaining as a product between these two partial performances, thus being influenced by these two opposite trends simultaneously. Obviously, adopting a too strict selector (e.g. $K = 30$) it penalizes the total ratio because of the low ρ, while a high K makes more difficult (and also computationally expensive) the classifier's task. This suggests that the choice of the selection parameter K should above all take into account the overall system behaviour.

5 Conclusions

In this article we presented a new selection method, namely fg-LS, conceived as first step for any two-stage large scale FRS. The method exploits the richness of the description given by multi-feature representation, the discriminative power given by the LDA projection, and the efficiency given by the divide et impera paradigm. The characteristic of the last method is crucial also for easily dealing with new incoming subjects, that can be added to any sub-gallery, requiring the re-computation of the only corresponding LDA and feature normalization, while leaving unaltered most of the gallery information. The method is based on the least squares technique, applied on small sub-galleries of face image characterized with multi-features. In order to test the method on large and challenging
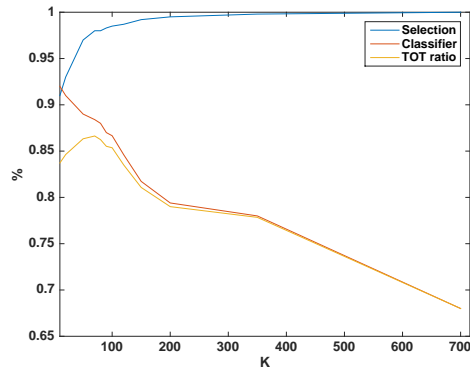
Fig. 5. Analysis of a Selection+Classifier FRS, namely fg-LS + k-LiMapS. *Blue line*: partial performances of the fg-LS selector. *Red line*: partial performances of the classifier k-LiMapS. *Yellow line*: overall system performances.

datasets, we collected the face images of four public databases, considering the images acquired in uncontrolled conditions.

The experimental analysis proves that multi-features actually give a powerful description of the data, and that the partitioning is advantageous specially if well tuned: neither too small nor too big sub-galleries are adequate in facing the selection problem. Concerning the parameter K, we have shown how it influences inversely the selection and the classifier, thus its setting should above all take into account the overall system behaviour. Comparisons proved that fg-LS is effective and efficient, performing better than well-known methods in the field.

Acknowledgments

References

1. Abate, A.F., Nappi, M., Riccio, D., Sabatino, G.: 2d and 3d face recognition: A survey. Pattern Recognition Letters 28(14), 1885–1906 (2007)
2. Adamo, A., Grossi, G., Lanzarotti, R.: Sparse representation based classification for face recognition by k-limaps algorithm. In: Image and Signal Processing - 5th Int'l Conf., ICISP 2012. LNCS, vol. 7340, pp. 245–252. Springer (2012)
3. Borgi, M., Labate, D., ElArbi, M., Ben Amar, C.: Shearlet network-based sparse coding augmented by facial texture features for face recognition. In: Petrosino, A. (ed.) Image Analysis and Processing ICIAP 2013, LNCS, vol. 8157, pp. 611–620. Springer (2013)

4. Cuculo, V., Lanzarotti, R., Boccignone, G.: Using sparse coding for landmark localization in facial expressions. In: Proc. of Int'l Conf. EUVIP. IEEE (2014)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition. CVPR 2005. vol. 1, pp. 886–893 (2005)
6. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-PIE. Image Vision Comput. 28(5), 807–813 (2010)
7. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007)
8. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and Simile Classifiers for Face Verification. In: IEEE Int'l Conf. on Computer Vision (ICCV) (2009)
9. Kyperountas, M., Tefas, A., Pitas, I.: Face recognition via adaptive discriminant clustering. In: Proc. of the Int'l Conf. on Image Processing, ICIP 2008. pp. 2744–7 (2008)
10. Li, C., Guo, J., Zhang, H.: Local sparse representation based classification. In: 20th International Conference on Pattern Recognition, ICPR 2010. pp. 649–652 (2010)
11. Liu, W., Wang, Y., Li, S., Tan, T.: Null space-based kernel fisher discriminant analysis for face recognition. In: Int'l Conf. on Automatic Face and Gesture Recognition. pp. 369 – 374. IEEE (2004)
12. Lu, J., Plataniotis, K.: Boosting face recognition on a large-scale database. In: Int'l Conf. on Image Processing. pp. 109–112. IEEE (2002)
13. Martinez, A., Benavente, R.: The AR face database. CVC Tech. Rep. 24 (1998)
14. Nan, Z., Jian, Y.: K nearest neighbor based local sparse representation classifier. IEEE 2010 Chin. Conf. on Pattern Recognition, CCPR pp. 1–5 (2010)
15. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Recognition and Machine Intelligence 24(7), 971–987 (2002)
16. Ortiz, E.G., Becker, B.C.: Face recognition for web-scale datasets. Computer Vision and Image Understanding 118, 153–170 (2014)
17. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K.: Overview of the face recognition grand challenge. Proc. IEEE Conf. CVPR 1, 947–954 (2005)
18. Schwartz, W., Guo, H., Choi, J., Davis, L.: Face identification using large feature sets. IEEE transactions on image processing 21(4), 2245–2255 (2012)
19. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. Trans. on Image Processing 19, 1635–1650 (2010)
20. Štruc, V., Pavešić, N.: Photometric normalization techniques for illumination invariance. In: Zhang, Y. (ed.) Advances in Face Image Analysis: Techniques and Technologies, pp. 279–300. IGI-Global (2011)
21. Wolf, L., Hassner, T., Taigman, Y.: Descriptor based methods in the wild. In: European Conference on Computer Vision (ECCV) (2008)
22. 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(2), 210–27 (2008)
23. Yan, J., Lei, Z., Yi, D., Li, S.: Towards incremental and large scale face recognition. In: Proc. of Int'l Joint Conf. on Biometrics (IJCB 2011). pp. 1–6 (2011)
24. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys 35(4), 399–458 (2003)
25. Zheng, W., Hu, B., Kong, X.: Two-stage nonnegative sparse representation for large-scale face recognition. IEEE Trans. on Neural Networks and Learning Systems 24(1), 35–46 (2013)