Face Recognition via Adaptive Sparse Representations of Random Patches

Domingo Mery
Department of Computer Science
Pontificia Universidad Católica de Chile
http://dmery.ing.puc.cl

Kevin Bowyer
Department of Computer Science and Engineering
University of Notre Dame
http://www.nd.edu/∼kwb/

***Abstract*—Unconstrained face recognition is still an open problem, as state-of-the-art algorithms have not yet reached high recognition performance in real-world environments (*e.g.*, crowd scenes at the Boston Marathon). This paper addresses this problem by proposing a new approach called Adaptive Sparse Representation of Random Patches (ASR+). In the learning stage, for each enrolled subject, a number of random patches are extracted from the subject's gallery images in order to construct representative dictionaries. In the testing stage, random test patches of the query image are extracted, and for each test patch a dictionary is built concatenating the 'best' representative dictionary of each subject. Using this adapted dictionary, each test patch is classified following the Sparse Representation Classification (SRC) methodology. Finally, the query image is classified by patch voting. Thus, our approach is able to deal with a larger degree of variability in ambient lighting, pose, expression, occlusion, face size and distance from the camera. Experiments were carried out on five widely-used face databases. Results show that ASR+ deals well with unconstrained conditions, outperforming various representative methods in the literature in many complex scenarios.**

I. Introduction

Face recognition has been a relevant area of research in computer vision, making many important contributions since the 1990s. In recent years the focus of face recognition algorithms has been shifted to deal with unconstrained conditions including variability in ambient lighting, pose, expression, face size and distance from the camera [13]. In the last few years, many approaches have been proposed to deal with the aforementioned problems (see for example [18]).

Algorithms based on Sparse Representation Classification (SRC) have been widely used [26]. In the sparse representation approach, a dictionary is built from the gallery images, and matching is done by reconstructing the query image using a sparse linear combination of the dictionary. The identity of the query image is assigned to the class with the minimal reconstruction error. Several variations of this approach were recently proposed. In [22], registration and illumination are simultaneously considered in the sparse representation. In [5], an intra-class variant dictionary is constructed to represent the possible variation between gallery and query images. In [23], sparsity and correlation are jointly considered. In [8] and [24], structured sparsity is proposed for dealing with occlusion and illumination problem. In [6], the dictionary is assembled by the class centroids and sample-to-centroid differences. In [3], SRC is extended by incorporating the low-rank structure of data representation. In [9], a discriminative dictionary is learned using label information. In [14], a linear extension of graph embedding is used to optimize the learning of the dictionary. In [15], a discriminative and generative dictionary is learned based on the principle of information maximization. In [17], a sparse discriminative analysis is proposed using the $\ell_{1,2}$-norm. In [27], a sparse representation in two phases is proposed. In [4], sparse representations of patches distributed in a grid manner are used. These variations improve recognition performance significantly as they are able to model various corruptions in face images, such as misalignment and occlusion.

Other approaches with comparable performance are based on the similarity between features extracted from regions of the gallery images and from the query image [19]. Recently, one novel approach proposed a new representation of the face image that is a sequence of forehead, eyes, nose, mouth and chin in a natural order [25].

Reflecting on the problems confronting unconstrained face recognition, and on the solutions proposed in recent years, we believe that there are some key ideas that should be present in new proposed solutions. First, if the face image is somehow occluded, it is clear that the occluded parts are not providing any information of the subject. For this reason, such parts should be automatically detected and should not be considered by the recognition algorithm. Second, in recognizing any face, there are parts of the face that are more relevant than other parts (for example birthmarks, moles or large eyebrows, to name but a few). For this reason, relevant parts should be subject-dependent, and could be found using unsupervised learning. Third, in the real-world environment, and given that face images are not perfectly aligned and the distance between camera and subject can vary from capture to capture, analysis of fixed sub-windows can lead to misclassification. For this reason, feature extraction should not be in fixed positions, and can be in several random positions, and use a selection criterion that enables selection of the best regions. Fourth, the expression that is present in a query face image can be subdivided into 'sub-expressions', for different parts of the face (*e.g.*, eyebrows, nose, mouth). For this reason, when searching for similar gallery subjects it would be helpful to search for image parts in all images of the gallery instead of similar gallery images.

Inspired by these key ideas, this paper proposes a new method for face recognition that is able to deal with less constrained conditions. Two main contributions of our approach are: 1) A new representation for the gallery face images of a subject: this is based on representative dictionaries learned for each subject of the gallery, which correspond to a rich collection of representations of selected relevant parts that are particular to the subject's face. 2) A new representation for the query face image: this is based on *i*) a discriminative criterion that selects the best test patches extracted randomly from the query image and *ii*) and an 'adaptive' sparse representation of the selected patches computed from the 'best' representative dictionary of each subject. Using these new representations, the proposed method (ASR+) can achieve high recognition performance under many complex conditions, as shown in our extensive experiments.

The rest of the paper is organized as follows: in Section II, the proposed method is explained in further detail. In Section III, the experiments and results are presented. Finally, in Section IV, concluding remarks are given.

II. Proposed Method

According to the motivation of our work, we believe that the robustness of the face recognition can be improved by using a patch-based approach. Thus, following a sparse representation methodology, in a learning stage several random patches can be extracted from each training image, and a dictionary can be built for each subject by concatenating its patches (stacking in columns). In the testing stage, several patches can be extracted and each of them can be classified using its sparse representation. The final decision can be made by majority vote. This baseline approach, however, shows three important disadvantages: *i*) The location information of the patch is not considered in the representation, *i.e.*, a patch of one part of the face could be erroneously represented by a patch of a different part of the face. This first problem can be solved by considering the (x, y) location of the patch in its description. *ii)* The method requires a huge dictionary for reliable performance, *i.e.*, each sparse representation process would be very time consuming. This second problem can be remedied by using only a part of the dictionary *adapted* to each patch. Thus, the whole dictionary of a subject can be subdivided into sub-dictionaries, and only the 'best' ones can be used to compute the sparse representation of a patch. *iii*) Not all query patches are relevant, *i.e.*, some patches of the face do not provide any discriminative information of the subject (*e.g.*, sunglasses). This third problem can be addressed by selecting the query patches according to a score value.

In this section we describe our approach taking into account the three mentioned improvements. As illustrated in Fig. 1, in the learning stage, for each subject of the gallery, several random small patches are extracted and described from their images (using both intensity and location features) in order to build representative dictionaries. In the testing stage, random test patches of the query image are extracted and described, and for each test patch a dictionary is built concatenating the 'best' representative dictionary of each subject. Using this adapted dictionary, each test patch is classified in accordance with the Sparse Representation Classification (SRC) methodology [26]. Afterwards, the patches are selected according to a discriminative criterion. Finally, the query image is classified by voting for the selected patches. Both stages will be explained in this section in further detail.

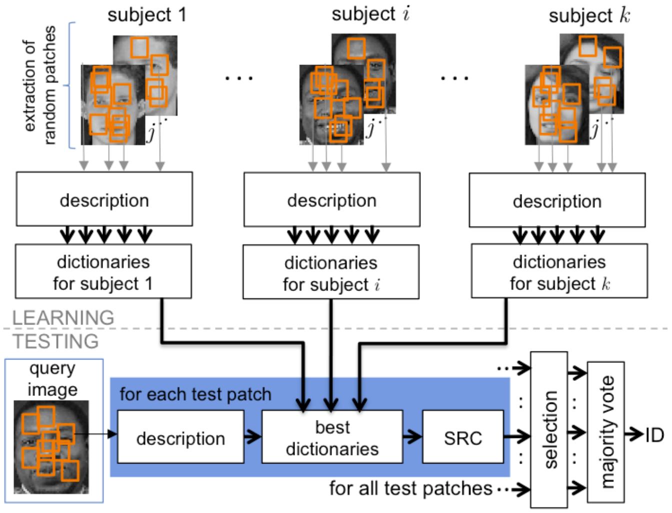

Fig. 1. Overview of proposed method ASR+.

A. Learning

In the training stage, a set of n face images of k subjects is available, where $\mathbf{I}^i_j$ denotes image j of subject i (for $i = 1 \ldots k$ and $j = 1 \ldots n$). In each image $\mathbf{I}^i_j$, m patches $\mathcal{P}^i_{jp}$ of size $w \times w$ pixels (for $p = 1 \ldots m$) are randomly extracted. They are centered in (x^i_{jp}, y^i_{jp}). In this work, the description of a patch $\mathcal{P}$ is defined as vector:

$$\mathbf{y} = f(\mathcal{P}) = [\ \mathbf{z}\ ;\ \alpha x\ ;\ \alpha y\ ] \in \mathcal{R}^{d+2} \tag{1}$$

where $\mathbf{z} = g(\mathcal{P}) \in \mathcal{R}^d$ is a descriptor of patch $\mathcal{P}$, (x, y) are the image coordinates of the center of patch $\mathcal{P}$, and α is a weighting factor between description and location[1]. Using (1) all extracted patches are described as $\mathbf{y}^i_{jp} = f(\mathcal{P}^i_{jp})$. Thus, for subject i an array with the description of all patches is defined as $\mathbf{Y}^i = \{\mathbf{y}^i_{jp}\} \in \mathcal{R}^{(d+2)\times nm}$ (for $j = 1 \ldots n$ and $p = 1 \ldots m$). The description $\mathbf{Y}^i$ of subject i is clustered using a k-means algorithm in Q clusters that will be referred to as *parent* clusters:

$$\mathbf{c}^i_q = \text{kmeans}(\mathbf{Y}^i, Q) \tag{2}$$

for $q = 1 \ldots Q$, where $\mathbf{c}^i_q \in \mathcal{R}^{(d+2)}$ is the centroid of parent cluster q of subject i. We define $\mathbf{Y}^i_q$ as the array with all samples $\mathbf{y}^i_{jp}$ that belong to the parent cluster with centroid $\mathbf{c}^i_q$.

[1] In our experiments, $\mathbf{z}$ corresponds to the intensity values of the patch subsampled by 2 in both directions, *i.e.*, $d = (w \times w)/4$ given by stacking its columns normalized to unit length in order to deal with different illumination conditions; (x, y) are normalized coordinates (values between 0 and 1); and $0.25 \leq \alpha \leq 4$.

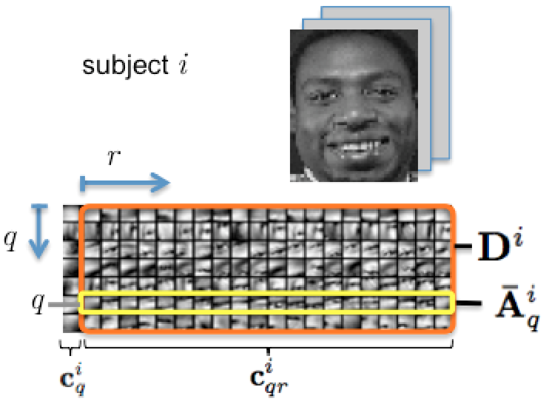

Fig. 2. Dictionaries of subject i for $Q = 32$ (only for $q = 1 \dots 7$ is shown) and $R = 20$. Left column shows the centroids $\mathbf{c}_q^i$ of parent clusters. Right columns (orange rectangle called $\mathbf{D}^i$) shows the centroids $\mathbf{c}_{qr}^i$ of child clusters. $\bar{\mathbf{A}}_q^i$ is row q of $\mathbf{D}^i$, *i.e.*, the centroids of child clusters of parent cluster q.

In order to select a reduced number of samples, each parent cluster is clustered again in R *child* clusters[2]

$$\mathbf{c}_{qr}^i = \text{kmeans}(\mathbf{Y}_q^i, R) \tag{3}$$

for $r = 1 \dots R$, where $\mathbf{c}_{qr}^i \in \mathcal{R}^{(d+2)}$ is the centroid of child cluster r of parent cluster q of subject i. All centroids of child clusters of subject i are arranged in an array $\mathbf{D}^i$, and specifically for parent cluster q are arranged in a matrix:

$$\bar{\mathbf{A}}_q^i = [\mathbf{c}_{q1}^i \dots \ \mathbf{c}_{qr}^i \dots \ \mathbf{c}_{qR}^i]^\mathsf{T} \in \mathcal{R}^{(d+2)\times R} \tag{4}$$

Thus, this arrangement contains R representative samples of parent cluster q of subject i as illustrated in Fig. 2. The set of all centroids of child clusters of subject i ($\mathbf{D}^i$), represents Q representative dictionaries with R descriptions $\{\mathbf{c}_{qr}^i\}$ for $q = 1 \dots Q, r = 1 \dots R$.

B. Testing

In the testing stage, the task is to determine the identity of the query image $\mathbf{I}^t$ given the model learned in the previous section. From the test image, s selected test patches $\mathcal{P}_p^t$ of size $w \times w$ pixels are extracted and described using (1) as $\mathbf{y}_p^t = f(\mathcal{P}_p^t)$ (for $p = 1 \dots s$). The selection criterion of a test patch will be explained later in this section. For each selected test patch with description $\mathbf{y} = \mathbf{y}_p^t$, a distance to each parent cluster q of each subject i of the gallery is measured:

$$h^i(\mathbf{y}, q) = \text{distance}(\mathbf{y}, \bar{\mathbf{A}}_q^i). \tag{5}$$

We tested with several distance metrics. The best performance, however, was obtained by $h^i(\mathbf{y}, q) = \min_r ||\mathbf{y} - \mathbf{c}_{qr}^i||$, which is the smallest distance to centroids of child clusters of parent cluster q as illustrated in Fig. 3. Normalizing $\mathbf{y}$ and $\mathbf{c}_{qr}^i$ to have unit ℓ_2 norm, (5) can be rewritten as:

$$h^i(\mathbf{y}, q) = 1 - \max < \mathbf{y}, \mathbf{c}_{qr}^i > \quad \text{for } r = 1 \dots R \tag{6}$$

where the term $< \bullet >$ corresponds to scalar product that provides a similarity (cosine of angle) between vectors $\mathbf{y}$ and $\mathbf{c}_{qr}^i$. The parent cluster that has the minimal distance is searched:

$$\hat{q}^i = \underset{q}{\text{argmin}} \ \ h^i(\mathbf{y}, q), \tag{7}$$

which minimal distance is $h^i(\mathbf{y}, \hat{q}^i)$. For patch $\mathbf{y}$, we select those gallery subjects that have a minimal distance less than a threshold θ in order to ensure a similarity between the test patch and representative subject patches. If k' subjects fulfill the condition $h^i(\mathbf{y}, \hat{q}^i) < \theta$ for $i = 1 \dots k$, with $k' \leq k$, we can build a new index $v_{i'}$ that indicates the index of the i'-th selected subject for $i' = 1 \dots k'$. For instance in a gallery with $k = 4$ subjects, if $k' = 3$ subjects are selected (*e.g.*, subjects 1, 3 and 4), then the indices are $v_1 = 1$, $v_2 = 3$ and $v_3 = 4$ as illustrated in Fig. 3. The selected subject i' for patch $\mathbf{y}$ has its dictionary $\mathbf{D}^{v_{i'}}$, and the corresponding parent cluster is $u_{i'} = \hat{q}^{v_{i'}}$, in which child clusters are stored in row $u_{i'}$ of $\mathbf{D}^{v_{i'}}$, *i.e.*, in $\mathbf{A}^{i'} := \bar{\mathbf{A}}_{u_{i'}}^{v_{i'}}$.

Therefore, a dictionary for patch $\mathbf{y}$ is built using the best representative patches as follows (see Fig. 3):

$$\mathbf{A}(\mathbf{y}) = [\ \mathbf{A}^1 \dots \mathbf{A}^{i'} \dots \mathbf{A}^{k'} \ ] \in \mathcal{R}^{(d+2)\times Rk'} \tag{8}$$

With this adaptive dictionary $\mathbf{A}$, built for patch $\mathbf{y}$, we can use SRC methodology [26]. That is, we look for a sparse representation of $\mathbf{y}$ using the ℓ_1-minimization approach:

$$\hat{\mathbf{x}} = \text{argmin}||\mathbf{x}||_1 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{y} \tag{9}$$

The residuals are calculated for the reconstruction for the selected subjects $i' = 1 \dots k'$:

$$r_{i'}(\mathbf{y}) = ||\mathbf{y} - \mathbf{A}\delta_{i'}(\hat{\mathbf{x}})|| \tag{10}$$

where $\delta_{i'}(\hat{\mathbf{x}})$ is a vector of the same size as $\hat{\mathbf{x}}$ whose only nonzero entries are the entries in $\hat{\mathbf{x}}$ corresponding to class $v(i') = v_{i'}$. Thus, the class of selected test patch $\mathbf{y}$ will be the class that has the minimal residual, that is it will be

$$\hat{i}(\mathbf{y}) = v(\hat{i}') \tag{11}$$

where $\hat{i}' = \text{argmin}_{i'} \, r_{i'}(\mathbf{y})$. Finally, the identity of the query subject will be the majority vote of the classes assigned to the s selected test patches $\mathbf{y}_p^t$, for $p = 1 \dots s$:

$$\text{identity}(\mathbf{I}^t) = \text{mode}(\hat{i}(\mathbf{y}_1^t), \dots \hat{i}(\mathbf{y}_p^t), \dots \hat{i}(\mathbf{y}_s^t)) \tag{12}$$

The selection of s patches of query image is as follows:
i) From query image $\mathbf{I}^t$, m patches are randomly extracted and described using (1): $\mathbf{y}_j^t$, for $j = 1 \dots m$, with $m \geq s$.
ii) Each patch $\mathbf{y}_j^t$ is represented by $\hat{\mathbf{x}}_j^t$ using (9).

[2] If n_q^i, the number of samples of $\mathbf{Y}_q^i$, is less than R, $\mathbf{c}_{qr}^i$ is built by taking the R first samples of a replicated version of the samples $[\mathbf{Y}_q^i \ \mathbf{Y}_q^i \dots]$. This dictionary with R words is equivalent to have a dictionary of n_q^i words only.

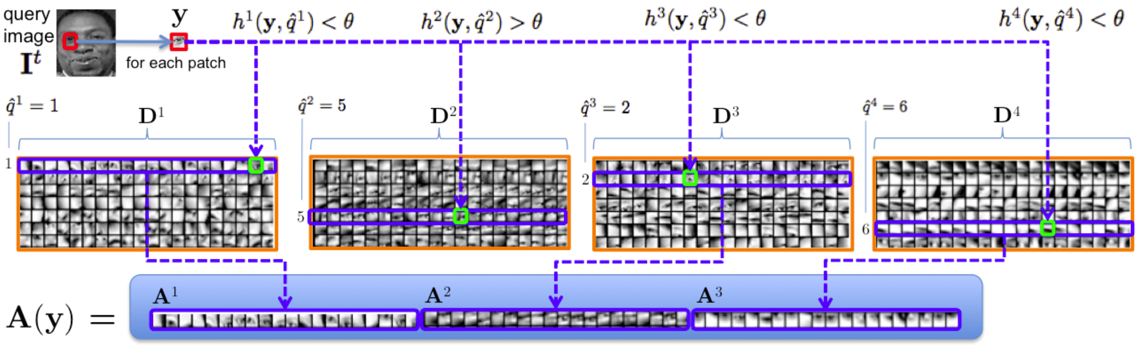

Fig. 3. Adaptive dictionary $\mathbf{A}$ of patch $\mathbf{y}$. In this example there are $k = 4$ subjects in the gallery. For this patch only $k' = 3$ subjects are selected. Dictionary $\mathbf{A}$ is built from those subjects by selecting all child clusters (of a parent cluster -see blue rectangles-) which have a child with the smallest distance to the patch (see green squares). In this example, subject 2 does not have child clusters that are similar enough to patch $\mathbf{y}$, *i.e.*, $h^2(\mathbf{y}, \hat{q}^2) > \theta$.

iii) The *sparsity concentration index* (SCI) of each patch is computed in order to evaluate how spread are its sparse coefficients [26]. SCI is defined by

$$S_j := \mathrm{SCI}(\mathbf{y}_j^t) = \frac{k\ \max(||\delta_{i'}(\hat{\mathbf{x}}_j^t)||_1)/||\hat{\mathbf{x}}_j^t||_1 - 1}{k-1} \qquad (13)$$

If a patch is discriminative enough it is expected that its SCI is large. Note that we use k instead of k' because the concentration of the coefficients related to k classes must be measured.

iv) Array $\{S\}_{j=1}^m$ is sorted into a descended order of SCI value. The first s patches in this sorted list in which SCI values are greater than a τ threshold are then selected. If only s' patches are selected, with $s' < s$, then the majority vote decision in (12) will be taken with the first s' patches.

III. Experimental Results

Experiments were carried out on five widely-used face databases under varying conditions as explained in Section III-A (see details of our implementation in Section III-B). We demonstrate the performance of our ASR+ approach with a combination of two types of experiments: 1) We compare performance of ASR+ to performance of our re-implemented versions of five well-known face recognition algorithms using the databases and experimental protocol described above (see Section III-C). 2) We compare performance of ASR+ against recent published performance results of a variety of algorithms using the database and experimental protocol used in the paper about each algorithm (see Section III-D).

A. Databases

In our experiments, the following databases were evaluated:

1) ORL: The database called 'The ORL Database of Faces' [16] consists of 40 subjects with 10 different images taken with some variation of lighting, face expressions and face details (glasses / no glasses). This is a very easy database, any face recognition algorithm should obtain more than 99% performance. It is used in our experiments as baseline only.

2) Yale: The database contains the original and extended 'Yale Database B' [10]. It consists of 38 subjects with 64 different images taken with many variations of lighting conditions. In this case we use the Tan-Triggs illumination normalization [20] that obtains better results than the raw images.

3) AR and **AR$_\times$**: The images of database 'AR' [11] were taken from 100 subjects (50 women and 50 men) with different facial expressions, illumination conditions, and occlusions with sun glasses and scarf (we used the cropped version). The number of images per subject is 26. We distinguish between AR and AR$_\times$: In AR, training and testing images are selected randomly from the 26 available images; whereas in AR$_\times$, training images are selected randomly from the images with no disguise, and testing from the images with disguise.

4) MPIE: The 'multi-PIE' database [7] contains more than 750,000 images taken from 337 subjects in four different sessions showing different expressions under 15 viewpoints and 19 illumination conditions. In our experiments, we used the frontal viewpoint only with all illuminations, expressions and sessions. All face images were cropped using the same fixed coordinates, thus the horizontal and vertical alignment of the faces varies between images.

5) FWM: The database 'The Face We Make' [12] contains images from 224 subjects (140 women and 84 men) with 10 different expressions that covey feelings related to common 'emoticons', *e.g.* :) smile, :-O surprised, :(sad, etc.

B. Protocol and Implementation

In the databases, there were K subjects and more than n images per subject. All images were resized to 110 × 90 pixels and converted to a grayscale image if necessary. In each dataset, we collected all available images for each subject, *e.g.*, gallery images, different aging, illumination conditions, expressions, camera distances, etc. We defined the following protocol: from these K subjects, we randomly selected $k \leq K$ subjects. From each selected subject, n images were randomly chosen for training and one for testing. In order to obtain a better confidence level in the estimation of face recognition accuracy, the test was repeated 100 times by randomly selecting new k subjects and $n + 1$ images each time. The reported accuracy η in all of our experiments is the average calculated over the 100 tests.

TABLE I

COMPARISON OF OUR ALGORITHM ASR+ FOR DIFFERENT NUMBER OF SUBJECTS k AND TRAINING IMAGES n

Database	Method	η_{A} [%]	η_{B} [%]	η_{C} [%]
		$k{=}20, n{=}4$	$k{=}40^*, n{=}9$	$k{=}100, n{=}14^{**}$
ORL	NBNN	91	97	
	LBP	94	**100**	
	SRC	96	98	
	TPTSR	94	**100**	***
	LAD	94	99	
	ASR+	**98**	**100**	
	$\Delta \rightarrow$	+2	0	
Yale	NBNN	99	**100**	
	LBP	69	82	
	SRC	90	**100**	
	TPTSR	**99**	**100**	***
	LAD	50	74	
	ASR+	98	**100**	
	$\Delta \rightarrow$	-1	0	
AR	NBNN	75	91	90
	LBP	75	93	98
	SRC	78	93	92
	TPTSR	86	95	94
	LAD	75	91	97
	ASR+	**91**	**100**	**100**
	$\Delta \rightarrow$	+5	+5	+2
$\mathrm{AR}_\times$	NBNN	72	70	68
	LBP	82	94	94
	SRC	46	43	40
	TPTSR	61	55	60
	LAD	57	67	77
	ASR+	**96**	**100**	**100**
	$\Delta \rightarrow$	+14	+6	+6
MPIE	NBNN	60	79	89
	LBP	84	94	**98**
	SRC	67	85	94
	TPTSR	62	70	92
	LAD	84	83	88
	ASR+	**91**	**98**	**98**
	$\Delta \rightarrow$	+7	+4	0
FWM	NBNN	68	79	79
	LBP	94	98	95
	SRC	79	90	84
	TPTSR	63	76	74
	LAD	94	**99**	**97**
	ASR+	**95**	**99**	**97**
	$\Delta \rightarrow$	+1	0	0

* For Yale: $k = 38$. ** For FWM: $n = 9$. *** Not enough subjects for this experiment.

TABLE II

COMPARISON OF OUR ALGORITHM ASR+ WITH OTHER STATE-OF-ART METHODS FOR DIFFERENT NUMBER OF SUBJECTS k AND TRAINING IMAGES n

Method (X)		Database	k	n	η_X [%]	$\eta_{\mathrm{ASR+}}$[%]	Δ
ASRC	[23]	ORL	40	5	96	**99**	+ 3
		AR	100	7	95	**98**	+ 3
InfoMax	[15]	Yale	38	33	95	**100**	+ 5
L21FLDA	[17]	ORL	40	3	84	**94**	+ 10
			40	5	93	**99**	+ 6
			40	7	97	**99**	+ 2
		Yale	38	10	89	**99**	+ 10
			38	20	96	**100**	+ 4
			38	30	98	**100**	+ 2
		MPIE	68	10	86	**98**	+ 12
			68	20	92	**100**	+ 8
			68	30	95	**100**	+ 5
DLRR	[3]	Yale	38	16	96	**100**	+ 4
			38	32	99	**100**	+ 1
		AR	100	7	94	**98**	+ 2
			100	9	90	**97**	+ 7
		MPIE	68	12	94	**96**	+ 2
LC-KSVD	[9]	Yale	38	15	95	**100**	+ 5
			38	33	97	**100**	+ 3
		AR	100	5	94	**95**	+ 1
			100	20	98	**100**	+ 2
SSRC	[6]	AR	100	13	99	**100**	+ 1
		$\mathrm{AR}_\times$	100	9	90	**99**	+ 9
DICW	[25]	$\mathrm{AR}_\times$	100	8	**99**	**99**	0
		FWM	55	8	82	**97**	+ 15
LGE-KSVD	[14]	Yale	38	32	96	**100**	+ 4
ESRC	[5]	AR	80	13	93	**100**	+ 7
DKSVD	[28]	Yale	38	32	96	**100**	+ 4
		AR	100	20	95	**100**	+ 5

In the implementation of ASR+, we used open source libraries like VLFeat [21] for k-means and SPAMS for sparse representation[3]. Our best parameters (obtained by trial and error) were as follows. Number of parent and child clusters: $Q = 80$ and $R = 50$ respectively. Number of patches: $m = 800$. Weighting factor for location coordinates: $\alpha = 4$ (for Yale, AR and $\mathrm{AR}_\times$), $\alpha = 2.5$ (for FWM) and $\alpha = 0.25$ (for ORL and MPIE)[4]. Size of patches: $w = 16$ pixels. Threshold for minimal distance between the test patch and child cluster: $\theta = 0.05$. Threshold for SCI $\tau = 0.1$. Number of selected patches $s = 300$. Additionally, the number of words ('atoms') selected from the dictionary in (9) is 20 k'/k, where k' is the number of selected subjects for the adaptive sparse representation, and k is the number of subjects in the gallery. The time computing depends on the number of subjects of the gallery, however, in order to present a reference, the testing results for $k = 40, n = 9$ were obtained after 0.8s per subject on a Mac Mini Server OS X 10.9.3, processor 2.6 GHz Intel Core i7 with 4 cores and memory of 16GB RAM 1600 MHz DDR3. The remaining algorithms were implemented in MATLAB. The code of the MATLAB implementation is available on our webpage[5].

C. General experiments

Our algorithm was compared with five well known face recognition methods: *i)* NBNN [2] using intensity features normalized to the unit length in 6 × 6 partitions, *ii)*, LBP [1] using 6 × 6 partitions, *iii)* SRC [26] where the images were sub-sampled to 22 × 18 pixels building features of dimension $d = 396$, *iv)* TPTSR based on a two-phase test sample sparse representation approach [27], and *v)* LAD [4] based on locally adaptive sparse representation of patches distributed in a grid. We coded these methods in Matlab according to the specifications given by the authors in their papers. The parameters were set so as to obtain the best performance.

Three general experiments were carried with different number of subjects (k) and training images (n): A) $k = 20, n = 4$, B) $k = 40, n = 9$, and C) $k = 100, n = 14$. An example of the dictionaries computed for one subject of database ORL is shown in Fig. 2. We observed that the dictionaries (rows of representations) corresponded to relevant parts of the subject viewed under different conditions (expressions, locations and size). The results are summarized in Table I, clearly demonstrating the ability of ASR+ to discriminate the classes. ASR+ achieves similar or better performance compared to a broad

[3] SPArse Modeling Software available on http://spams-devel.gforge.inria.fr

[4] Three values of α for a low, middle and high misalignment respectively.

[5] See http://dmery.ing.puc.cl/index.php/material.

range of state-of-the-art algorithms across a range of widely-used face databases. In row '$\Delta \rightarrow$', we observe the difference between the accuracy of ASR+ and the best accuracy of the other 5 methods. The accuracy advantage of our method is better when faces are taken under less constrained conditions, for example when faces are occluded or not well aligned.

D. Comparison with the state of the art

In order to compare the performance of the proposed method ASR+ with other state-of-art approaches, we collected the results in face recognition published in the last five years in prestigious journals and conferences. We followed the same protocols as in those papers, and the accuracy of ASR+ was measured for comparison. The comparison is shown in Table II. Once again, the results are consistent: ASR+ deals well with unconstrained conditions, outperforming various representative methods in the literature in many complex scenarios as we can see in the column Δ where the difference between the accuracy of ASR+ and the other methods is shown.

IV. Conclusions

In this paper, we have presented ASR+, a new algorithm that is able to recognize faces automatically in cases with less constrained conditions, including some variability in ambient lighting, pose, expression, size of the face and distance from the camera. The robustness of our algorithm is due to two reasons: *i)* the dictionaries learned for each subject of the gallery in the learning stage corresponded to a rich collection of representations of relevant parts which were selected and clustered; *ii)* the testing stage was based on 'adaptive' sparse representations of several patches using the dictionaries estimated in the previous stage which provided the best match with the patches.

It is worth mentioning that our extensive empirical evaluation has been performed in two directions: *i)* Other representative methods from the literature have been re-implemented and compared against using our methodology; and *ii)* our algorithm has been evaluated using the methodology of other papers to get a result that can be compared to their published result(s) on the selected datasets. In both scenarios, ASR+ can deal with the unconstrained conditions extremely well, achieving a high recognition performance in many complex conditions and outperforming the other tested algorithms.

We believe that ASR+ can be used to solve other kinds of recognition problems. The proposed model is very flexible and obviously it can be used with other descriptors. In terms of future work, we will extend this approach to face recognition using videos and other object-recognition problems.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[2] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, 2008.

[3] J. Chen and Z. Yi. Sparse representation for face recognition by discriminative low-rank matrix recovery. *Journal of Visual Communication and Image Representation*, 25:763–773, 2014.

[4] Y. Chen, T. T. Do, and T. D. Tran. Robust face recognition using locally adaptive sparse representation. In *IEEE International Conference on Image Processing (ICIP 2010)*, pages 1657–1660, 2010.

[5] W. Deng, J. Hu, and J. Guo. Extended SRC: Undersampled face recognition via intraclass variant dictionary. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1864–1870, 2012.

[6] W. Deng, J. Hu, and J. Guo. In defense of sparsity based face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013)*, pages 399–406, 2013.

[7] R. Gross, I. Matthews, J. Cohn, and T. Kanade. Multi-PIE. *Image and Vision Computing*, 28:807–813, 2010.

[8] K. Jia, T.-H. Chan, and Y. Ma. Robust and practical face recognition via structured sparsity. In *European Conference on Computer Vision (ECCV 2012)*, pages 331–344, 2012.

[9] Z. Jiang, Z. Lin, and L. S. Davis. Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2651–2664, 2013.

[10] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.

[11] A. Martinez and R. Benavente. The AR face database, June 1998. CVC Tech. Rep, No. 24.

[12] D. Miranda. The Face We Make, 2011. http://thefacewemake.org/.

[13] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'clocke, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer. An introduction to the good, the bad, & the ugly face recognition challenge problem. In *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*, pages 346–353, 2011.

[14] R. Ptucha and A. Savakis. LGE-KSVD: Flexible Dictionary Learning for Optimized Sparse Representation Classification. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2013)*, pages 854–861, 2013.

[15] Q. Qiu, V. M. Patel, and R. Chellappa. Information-theoretic Dictionary Learning for Image Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014. (to be published).

[16] F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pages 138–142, 1994.

[17] X. Shi, Y. Yang, Z. Guo, and Z. Lai. Face Recognition by Sparse Discriminant Analysis via Joint $L_{2,1}$-norm Minimization. *Pattern Recognition*, 2014.

[18] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)*, 2014.

[19] X. Tan, S. Chen, Z.-H. Zhou, and J. Liu. Face recognition under occlusions and variant expressions with partial similarity. *IEEE Transactions on Information Forensics and Security*, 4(2):217–230, 2009.

[20] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Analysis and Modeling of Faces and Gestures*, pages 168–182, 2007.

[21] A. Vedaldi and B. Fulkerson. VLfeat: an open and portable library of computer vision algorithms. In *MM '10: Proceedings of the international conference on Multimedia*, pages 1469–1472, New York, Oct. 2010.

[22] A. Wagner, Z. Zhou, J. Wright, H. Mobahi, A. Ganesh, and Y. Ma. Toward a Practical Face Recognition System: Robust Alignment and Illumination by Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):372–386, 2012.

[23] J. Wang, C. Lu, M. Wang, P. Li, S. Yan, and X. Hu. Robust Face Recognition via Adaptive Sparse Representation. *IEEE Transactions on Cybernetics*, 2014. (to be published).

[24] X. Wei, C.-T. Li, and Y. Hu. Robust face recognition under varying illumination and occlusion considering structured sparsity. In *International Conference on Digital Image Computing Techniques and Applications (DICTA 2012)*, 2012.

[25] X. Wei, C.-T. Li, and Y. Hu. Face recognition with occlusion using dynamic image-to-class warping DICW. In *IEEE International Conference on Automatic Face Gesture Recognition, (FG 2013)*, 2013.

[26] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[27] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang. A Two-Phase Test Sample Sparse Representation Method for Use With Face Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(9):1255–1262, 2011.

[28] Q. Zhang and B. Li. Discriminative K-SVD for dictionary learning in face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pages 2691–2698, 2010.