Face Recognition using Adaptive Sparse Representations

Anonymous ECCV submission

Paper ID 965

Abstract. Unconstrained face recognition is still an open question, as state-of-the-art algorithms have not yet reached high recognition performance in real-world environments. This paper attempts to make a contribution to this field by proposing a new approach called ASR – Adaptive Sparse Representation. ASR consists of two stages: learning and testing. In the learning stage, for each enrolled subject, several random small patches are extracted from its 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 less constrained conditions including some variability in ambient lighting, pose, expression, face size and distance from the camera. Experiments were carried out on six different databases. ASR could deal with unconstrained conditions well, achieving a good recognition performance in many complex scenarios. It outperformed other representative methods in the literature using the same descriptors based on intensity features. ...

Keywords: Face recognition, sparse coding, sparse representation classification, real word environments, pacth-based recognition, patch selection.

1 Introduction

Face recognition has been a relevant area of research in computer vision, making many important contributions since the 1990s, see for example holistic methods [1, 2], in which the face image is analyzed as one entity without considering different parts of the face. In the following two decades, high recognition performance was achieved in still frontal and aligned images. Methods based on local descriptors, such as local binary patterns [3] in combination with one 'image-to-class' approach [4] have achieved some levels of robustness against occlusion and moderate misalignment. Interesting variations of LBP can be found in [5, 6] to mention a few.

Nowadays, it is clear that holistic methods are not suitable when dealing with certain real problems such as occlusion or misalignment, because holistic features

can be distorted. Moreover, methods that extract features from sub-windows distributed in a regular grid of the face image, or from face landmarks (determined automatically from the face image) may suffer from unwanted descriptions (in wrong locations of the face) which can lead to misclassification.

For these reasons, 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 [7–10].

In the last few years, many approaches have been proposed to deal with occlusions. Algorithms based on sparse representations have been widely used [11–14]. In this approach, a dictionary built from the gallery images is used. The idea is to reconstruct 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. Variations of this approach were recently proposed: in [15], an intra-class variant dictionary is constructed to represent the possible variation between gallery and query images. In [16], the dictionary is assembled by the class centroids and sample-to-centroid differences. In [17], a class of structured sparsity-inducing norms is included. These variations improve recognition performance significantly as they are able to model various corruptions in face images, such as misalignment and occlusion. Other approaches are based on the similarity between features extracted from regions of the gallery images and from the query image [18, 19]. The idea is to use a training process to filter out those regions that can correspond to occluded parts. 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 [20], in which recognition is performed taking the order information into account using Dynamic Image-to-Class Warping, that computes the distance between a query face and an enrolled person.

When studying face recognition problems and the suggested solutions proposed in recent years, we believe that there are some key ideas that should be present in new solutions that attempt to deal with real-world environments. 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, by recognizing the face of a subject, there are parts of this that are more relevant than other parts (for example birthmarks, moles or large eyebrows, to name but a few). For this reason, the 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, thus using a selection criterion enables the best regions to be chosen. Fourth, the face expression that is present in a query image can be subdivided into 'sub-expressions', *i.e.*, of different parts of the face image. 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. The contributions of our approach, referred to as Adaptive Sparse Representation (ASR), are the following two:

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 subjects face.
2. A new representation of 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.

Combining these two key ideas, a classification approach –such as Sparse Representation Classifier (SRC) [11]– can achieve high recognition performance under many complex conditions, as we have shown in our experiments.

The rest of the paper is organized as follows: in Section 2, the proposed ASR method is explained in further detail. In Section 3, the experiments and results are presented. Finally, in Section 4, concluding remarks are given.

2 Proposed Method

The proposed ASR method consists of two stages: learning and testing (see Fig. 1). In the learning stage, for each subject of the gallery, several random

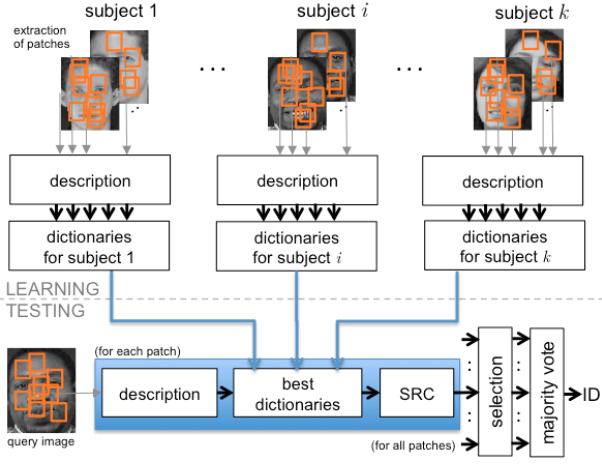

Fig. 1. Overview of proposed method ASR.

small patches are extracted and described from their images in order to built 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 [11]. 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.

2.1 Model learning

In the training stage, a set of n face images of k subjects is available, where $\mathbf{I}_j^i$ denotes image j of subject i (for $i = 1 \dots k$ and $j = 1 \dots n$) as illustrated in Fig. 2. In each image $\mathbf{I}_j^i$, m patches $\mathcal{P}_{jp}^i$ of size $w \times w$ pixels (for $p = 1 \dots m$) are randomly extracted. They are centred in (x_{jp}^i, y_{jp}^i). In this work, a patch $\mathcal{P}$ is defined as vector:

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

where $\mathbf{z} = g(\mathcal{P}) \in \mathcal{R}^d$ is a descriptor of patch $\mathcal{P}$ (*e.g.*, a local descriptor or the $d = w \times w$ gray values of the patch given by stacking its columns); (x, y)

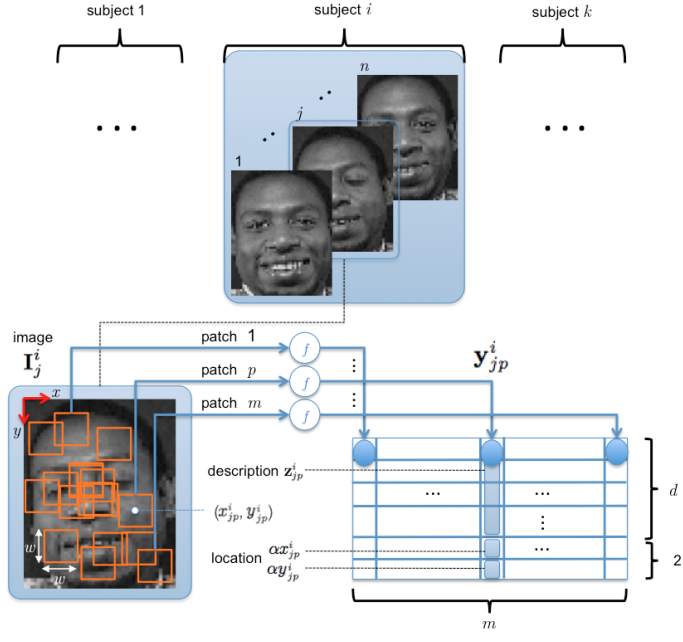

Fig. 2. Extraction and description of m patches of gallery image j of subject i.

are the image coordinates of the center of patch $\mathcal{P}$; and α is a weighting factor between description and location. Patch $\mathcal{P}$ is described using a vector that has been normalized to unit length:

$$\mathbf{y} = f(\mathcal{P}) = \frac{\mathbf{p}}{||\mathbf{p}||} \in \mathcal{R}^{d+2} \tag{2}$$

Using (2) 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 $j = 1 \dots n$ and $p = 1 \dots m$) as shown in Fig. 3a.

The description $\mathbf{Y}^i$ of subject i is clustered using k-means algorithm in Q clusters that will be referred to as *parent* clusters (Fig. 3b):

$$\mathbf{c}^i_q = \text{kmeans}(\mathbf{Y}^i, Q) \tag{3}$$

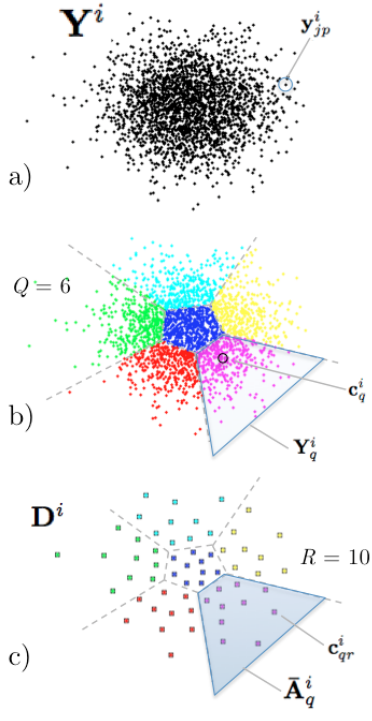

Fig. 3. Representation of subject i: a) each representation of a patch $\mathbf{y}^i_{jp}$ corresponds to a point in $\mathcal{R}^{(d+2)}$ (see black points of $\mathbf{Y}^i$). b) Set of points $\mathbf{Y}^i$ is clustered in Q parent clusters in which points are arranged in array $\mathbf{Y}^i_q$ with centroid $\mathbf{c}^i_q$. c) Each parent cluster $\mathbf{Y}^i_q$ is clustered in R child clusters in which centroids $\{\mathbf{c}^i_{qr}\}^R_{r=1}$ are arranged in array $\bar{\mathbf{A}}^i_q$. In this example $\mathbf{Y}^i$ has 2.400 points, $Q = 6$ (parent clusters) and $R = 10$ (child clusters).

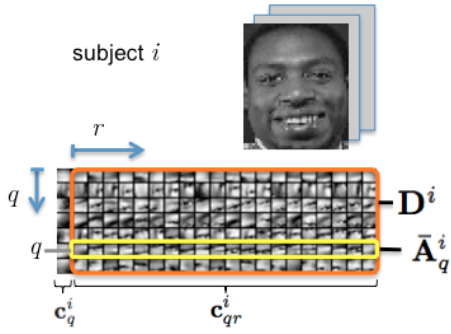

Fig. 4. Representative dictionaries of subject i for $Q = 32$ (only for $q = 1 \dots 7$ is shown) and $R = 20$. Left column shows the centroids $\mathbf{c}^i_q$ of parent clusters. Right columns (orange rectangle called $\mathbf{D}^i$) shows the centroids $\mathbf{c}^i_{qr}$ of child clusters. $\bar{\mathbf{A}}^i_q$ is row q of $\mathbf{D}^i$, *i.e.*, the centroids of child clusters of parent cluster q.

for $q = 1 \dots Q$, 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 $\mathbf{c}^i_q$. In order to select a reduced number of samples, each parent cluster is clustered again in R *child* clusters (Fig. 3c):

$$\mathbf{c}^i_{qr} = \text{kmeans}(\mathbf{Y}^i_q, R) \tag{4}$$

for $r = 1 \dots R$, where $\mathbf{c}^i_{qr} \in \mathcal{R}^{(d+2)}$ 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:

$$\bar{\mathbf{A}}^i_q = [\mathbf{c}^i_{q1} \dots \ \mathbf{c}^i_{qr} \dots \ \mathbf{c}^i_{qR}]^\top \in \mathcal{R}^{(d+2)\times R} \tag{5}$$

Thus, this arrange contains R representative samples of parent cluster q of subject i as illustrated in Fig. 4. The set of all centroids of child clusters of subject i ($\mathbf{D}^i$), represents Q representative dictionaries with R descriptions $\{\mathbf{c}^i_{qr}\}$ for $q = 1 \dots Q, r = 1 \dots R$.

2.2 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 $\mathcal{P}^t_p$ of size $w \times w$ pixels are extracted and described using (2) as $\mathbf{y}^t_p = f(\mathcal{P}^t_p)$ (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 $\mathbf{y} = \mathbf{y}^t_p$, a distance to each parent cluster q of each subject i of the gallery is measured:

$$h^i(\mathbf{y}, q) = \text{distance}(\mathbf{y}, \bar{\mathbf{A}}^i_q). \tag{6}$$

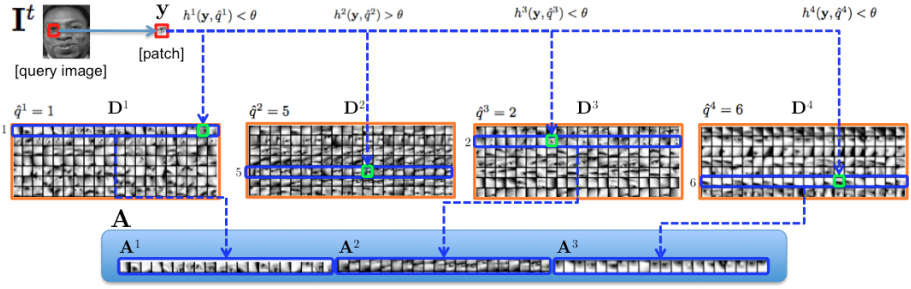

Fig. 5. Adaptive dictionary $\mathbf{A}$ of patch $\mathbf{y}$. In this example there are $k = 4$ subjects in the gallery. For this patch only $k' = 3$ subjects are selected. Dictionary $\mathbf{A}$ is built from those subjects by selecting all child clusters (of a parent cluster -see blue rectangles-) which has a child cluster with the smallest distance to the patch (see green squares). In this example, subject 2 does not have child clusters that are similar enough, *i.e.*, $h^2(\mathbf{y}, \hat{q}^2) > \theta$.

Several distance metrics were tested in our experiments. For instance, the Euclidean distance to centroid of parent cluster $h^i(\mathbf{y}, q) = ||\mathbf{y} - \mathbf{c}_q^i||$ achieves good results and can be implemented very fast using kd-trees structures, however, the best performance 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. 5. As $\mathbf{y}$ and $\mathbf{c}_{qr}^i$ are pre-normalized to have unit ℓ_2 norm, it can be rewritten as:

$$h^i(\mathbf{y}, q) = 1 - \max < \mathbf{y}, \mathbf{c}_{qr}^i > \quad \text{for } r = 1 \dots R \tag{7}$$

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:

$$\hat{q}^i = \underset{q}{\text{argmin}} \; h^i(\mathbf{y}, q), \tag{8}$$

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. 5. 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. 5):

$$\mathbf{A}(\mathbf{y}) = [\ \mathbf{A}^1 \dots \mathbf{A}^{i'} \dots \mathbf{A}^{k'}\ ] \in \mathcal{R}^{(d+2)\times Rk'} \tag{9}$$

With this adaptive dictionary $\mathbf{A}$, built for patch $\mathbf{y}$, we can use *Sparse Representation Classification* (SRC) methodology [11]. That is, we look for a sparse representation of $\mathbf{y}$ using the ℓ_1-minimization approach:

$$\hat{\mathbf{x}} = \text{argmin}||\mathbf{x}||_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y} \tag{10}$$

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

$$r_{i'}(\mathbf{y}) = ||\mathbf{y} - \mathbf{A}\delta_{i'}(\hat{\mathbf{x}})|| \tag{11}$$

where $\delta_{i'}(\hat{\mathbf{x}})$ is a vector of the same size of $\hat{\mathbf{x}}$ 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

$$\hat{i}(\mathbf{y}) = v(\hat{i'}) \tag{12}$$

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 $p = 1 \dots s$:

$$\text{identity}(\mathbf{I}^t) = \text{mode}(\hat{i}(\mathbf{y}_1^t), \dots \hat{i}(\mathbf{y}_p^t), \dots \hat{i}(\mathbf{y}_s^t)) \tag{13}$$

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 (2): $\mathbf{y}_j^t$, 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 the mentioned adaptive sparse representation according to (10).

iii) The *sparsity concentration index* (SCI) of each patch is computed in order to evaluate how spread are its sparse coefficients [11]. SCI is defined by

$$S_j := \text{SCI}(\mathbf{y}_j^t) = \frac{k \cdot \max(||\delta_{i'}(\hat{\mathbf{x}}_j^t)||_1)/||\hat{\mathbf{x}}_j^t||_1 - 1}{k - 1} \tag{14}$$

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 in a descended way.

v) 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 (13) will be taken with the first s' patches.

3 Experimental Results

In this Section, results obtained from different databases under varying conditions are reported. In the databases there were K subjects, each of them had at least N images. 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 experiment: from these K subjects, we randomly selected $k \leq K$ subjects. From each selected subject, N images were randomly chosen, from them $n = N - 1$ images were used for training and the other one for testing, *i.e.*, in our experiment there were kn images for training (k subjects with n images each) and k images for testing (one of each subject). Four different face recognition algorithms were tested. For a fair comparison, we used the same descriptors based on intensity features. The algorithms that we tested are: *i)* NBNN [4] using intensity features normalized to the unit length in 6×6 partitions, *ii)* SRC [11] where the images were subsampled to 22×18 pixels building features of dimension $d = 396$, *iii*) TPTSR based on a two-phase test sample sparse representation approach [14], and *iv)* ASR (our proposed method) using the parameters described below. The parameters of all methods were set so as to obtain the best performance. In order to obtain a better confidence level in the estimation of face recognition accuracy, the test was repeated 100 times randomly selecting new k subjects and N images each time. **The reported accuracy η in all of our experiments was the average calculated from the accuracy of 100 tests**.

3.1 Databases

In our experiments, we tested our method on six set of face images:

• **ORL**: The database called 'The ORL Database of Faces' [21] 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 reference only.

• **Yale**: The database contains the original and extended 'Yale Database B' [22]. It consists of 38 subjects with 64 different images taken with many variations of lighting conditions. In this case we use the Tan-Triggs normalization [23] that obtain better results.

• **Cas-Peal**: The database, called 'CAS-PEAL-R1' [24], consists of several face images taken under different conditions. We used the 66 subjects contained in subset 'Aging'. Thus, we collected the images of these subjects in all other subsets ('Gallery', 'Accessory', 'Background', 'Distance', 'Expression' and 'Lighting'). The number of images per subject (in the selected 66 subjects) is between 26 and 37. In this case we use the Tan-Triggs normalization [23] that obtain better results

• **RR**: This is a new database. Database 'RR' was obtained in real world from an access control system. The images were collected by an enterprise[2] during

the last year (on different days) using an iPad mini at the entrance of an office building. Each subject was asked to look at the iPad screen, on which a life video of the frontal camera was presented. When a face was detected (using iPad's face detector[1]) an image was automatically captured using the frontal camera, and the face image was cropped according to the parameters proposed by the aforementioned face detector. The database consists of 114 subjects. The number of images per subject is between 10 and 38. There are several problems in these images: they are not well aligned and centered; the size (in pixels) of a face varied from day to day; many images are captured when the subject was talking with other subjects or talking by phone, smiling, smoking, etc.; and many of the images present some degree of blur. For security reasons, we are not permitted to publish these images.

• **AR**: The images of database 'AR' [25] 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. In order to test the algorithms on face images that contain real disguise, the experiments performed on this database took randomly the training images from not occluded faces only (face images with different facial expressions and illumination conditions), and the query images from occluded faces only (face images with sun glasses and scarf).

• **LFW**: The images database 'Labeled Faces in the Wild–LFW' contains real-life face images taken under unconstrained conditions and collected from the web [26]. We used the deep funneled version [27]. We selected those subjects that have at least 10 images per subject. Thus, the database consists of 157 subjects. The number of images per subject is between 10 and 323 (the average is 17.3 images per subject). The face images of LFW database have a large amount of intra-class variability, due to factors such as pose, background, expression and lighting.

3.2 Implementation

We used the implementation of k-means from [28] and SPAMS library for sparse representation[2]. The remaining algorithms were implemented in MATLAB. Our best parameters (obtained by trial and error) were: $\alpha = 0.1$ (weighting factor for location coordinates), $Q = 32$ (number of parent clusters), $R = 20$ (number of child clusters), $m = 600$ (number of patches per image), $w = 16$ (size of patch), $\theta = 0.1$ (threshold for minimal distance between the test patch and child cluster), $\tau = 0.2$ (threshold for SCI) and $s = 300$ (number of selected patches). In our experiments it was noted that the use of these thresholds incremented the accuracy to about 3%. Additionally, the number of atoms selected from the dictionary 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

[1] http://developer.apple.com

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

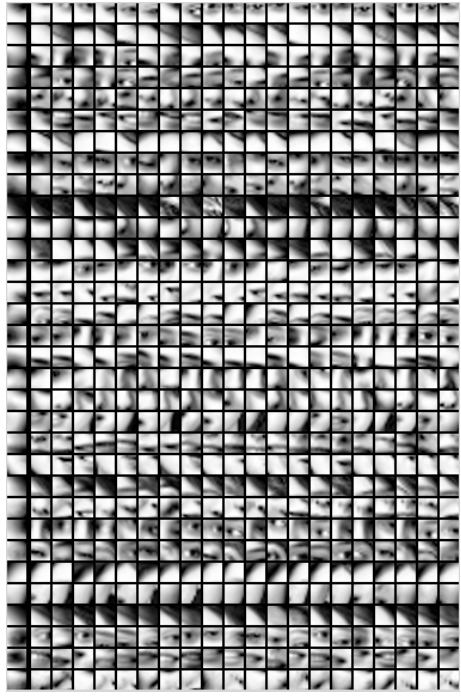

Fig. 6. The rows illustrate the representative dictionaries (of subject k from Fig. 1) computed by our algorithm ASR with $Q = 32$ and $R = 20$. The left column shows the parent clusters. The right columns show the child clusters.

present a reference, the testing results for $k = 20, N = 5$ and for $k = 40, N = 10$ were obtained after 1 and 2 seconds per subject respectively on a iMac OS X 10.8.5, processor 2.9 GHz Intel Core i7, memory of 8GB RAM 1600 MHz DDR3. The code of the MATLAB implementation is available on our webpage[3].

3.3 General experiments

Two general experiments were carried out using the mentioned algorithms and databases:

- Experiment 'A': $k = 20$ subjects with $N = 5$ images per subject and
- Experiment 'B': $k = 40$ subjects with $N = 10$ images per subject.

An example of the dictionaries computed for one subject of database ORL is shown in Fig. 6. 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 1. The ability of our algorithm ASR to discriminate the classes is illustrated in Fig. 7. It

[3] Not given because this is an anonymous submission.

Table 1. Comparison of our algorithm ASR.

	η_A for $k = 20$, $N = 5$				η_B for $k = 40$, $N = 10$			
Method → Database	NBNN [%]	SRC [%]	TPTSR [%]	ASR [%]	NBNN [%]	SRC [%]	TPTSR [%]	ASR [%]
ORL	91	96	94	98	95	97	98	100
Yale	98	92	99	96	100	96	100	98
Cas-Peal	68	72	71	75	76	77	85	88
RR	58	70	71	93	68	75	80	95
AR	70	43	55	87	72	47	67	95
LFW	25	28	25	65	33	31	28	75

(*) for η_B only $k = 38$ subjects were available.

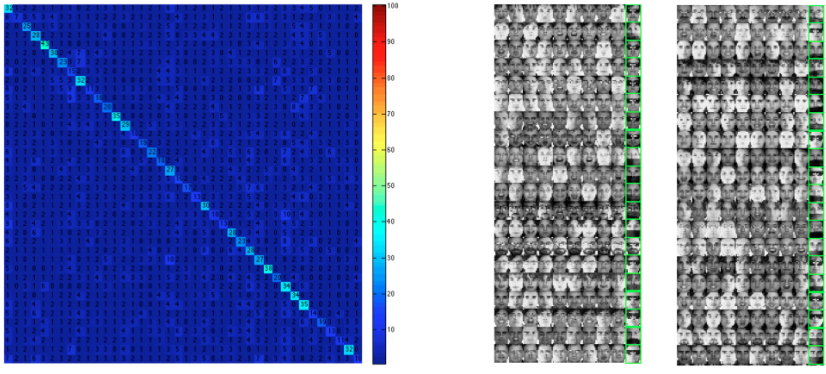

Fig. 7. Face recognition using AR database with k=40 subjects (9 images for training and 1 for testing). Left) Voting process. The element (i, j) of this matrix represents the percentage of votes that test subject i obtained from gallery subject j. The majority vote computed by (13) is concentrated in the diagonal, *i.e.*, subject i is classified correctly. Right) 40 subjects (20 in each column) are correctly recognized. The query image with real disguise (green square) was correctly identified using nine images without occlusions.

was observed that ASR –in comparison with the mentioned methods– achieved similar or better performance. The accuracy of our method is considerably better when faces are taken under unconstrained conditions, for example when faces are occluded or not well aligned.

3.4 Experiments with occlusion

In order to evaluate the robustness of the algorithms against occlusion, we corrupted the test images with a black square of size $a \times a$ pixels located randomly, for $a = 6, 12, 18 \ldots 54$. For $a = 54$, the occluded area corresponded to 29.5% of the image. The experiments were carried out for $k = 40$ random subjects with the first $N = 10$ images per subject on database Cas-Peal. The experiment

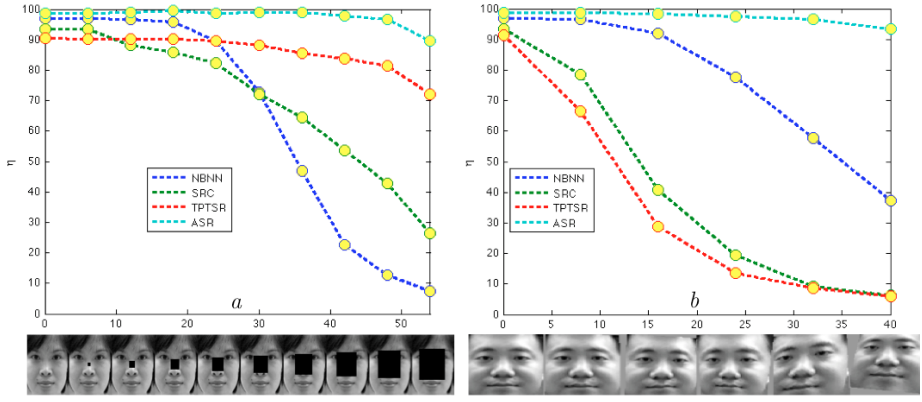

Fig. 8. Performance in Cas-Peal database with occlusion and changes in the size and alignment (see text). Left) occlusion of the faces by superimposition random black squares of size $a \times a$ pixels onto the query image. Right) change of the alignment and size in the query image by shifting the corners randomly $\pm b/2$ pixels. The size of the images is 90×110 pixels.

was repeated 100 times and the accuracy was averaged. The results are plotted in Fig. 8-left. The robustness of our algorithm is clearly better even by high occlusions.

3.5 Experiments with misalignment and size

In order to evaluate the robustness of the algorithms against misalignment and size of the face, the test images were transformed using a 2D general projective (homography) transformation [29], in which the four corners of the original test image of size 110×90 pixels: $\mathbf{m}_1 = (1,1)$, $\mathbf{m}_2 = (1,90)$, $\mathbf{m}_3 = (110,1)$ and $\mathbf{m}_4 = (110,90)$, were transformed into the new coordinate $\mathbf{m}'_i = \mathbf{m}_i + \mathbf{r}_i$, where $\mathbf{r}_i = (x_i, y_i)$ are random numbers between $-b/2$ and $b/2$ distributed uniformly. We tested for b =8, 16, 32 and 40 (see an example in Fig. 8 right-bottom). Again, the experiments were carried out for $k = 40$ random subjects with the first $N = 10$ images per subject on a Cas-Peal database. The experiment was repeated 100 times and the accuracy was averaged. The results are plotted in Fig. 8-right. The robustness of our algorithm is clearly better even with respect to high changes of location, rotation, scale and geometric distortion.

4 Conclusions

In this paper, we have presented 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. Combining these two key ideas, the algorithm can deal with the mentioned unconstrained conditions extremely well, achieving a good recognition performance in many complex conditions and outperforming the other tested algorithms using the same descriptors based on intensity features. We believe that this new adaptive sparse representation can be used to solve other kinds of computer vision problems in which there are similar unconstrained conditions.

The proposed model is very flexible and obviously it can be used with other descriptors. In terms of future work, we will accelerate computation by using faster proximity search algorithms. Additionally, we will extend this approach to face recognition using videos and other object-recognition problems.

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3**(1) (1991) 71–86
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7) (1997) 711–720
3. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(12) (2006) 2037–2041
4. Boiman, O., Shechtman, E., Irani, M.: In defense of nearest-neighbor based image classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). (2008) 1–8
5. Maturana, D., Mery, D., Soto, A.: Face recognition with decision tree-based local binary patterns. In: Asian Conference on Computer Vision (ACCV 2010). (2010) 618–629
6. Xie, S., Shan, S., Chen, X., Meng, X., Gao, W.: Learned local gabor patterns for face representation and recognition. Signal Processing **89**(12) (2009) 2333–2344
7. Phillips, P.J., Beveridge, J.R., Draper, B.A., Givens, G., O'Toole, A.J., Bolme, D.S., Dunlop, J., Lui, Y.M., Sahibzada, H., Weimer, S.: An introduction to the good, the bad, & the ugly face recognition challenge problem. In: Proceedings of IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011). (2011) 346–353
8. Berg, T., Belhumeur, P.N.: Tom-vs-Pete Classifiers and Identity-Preserving Alignment for Face Verification. In: British Machine Vision Conference BMVC. (2012)
9. Cao, Q., Ying, Y., Li, P.: Similarity Metric Learning for Face Recognition. In: IEEE International Conference on Computer Vision (ICCV). (2013)
10. Barkan, O., Weill, J., Wolf, L., Aronowitz, H.: Fast High Dimensional Vector Multiplication Face Recognition. In: IEEE International Conference on Computer Vision (ICCV). (2013)

11. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(2) (2009) 210–227
12. Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Ma, Y.: Towards a practical face recognition system: Robust registration and illumination by sparse representation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009). (2009) 597–604
13. Wei, X., Li, C.T., Hu, Y.: Robust face recognition under varying illumination and occlusion considering structured sparsity. In: International Conference on Digital Image Computing Techniques and Applications (DICTA 2012). (2012) 1–7
14. Xu, Y., Zhang, D., Yang, J., Yang, J.Y.: A Two-Phase Test Sample Sparse Representation Method for Use With Face Recognition. IEEE Transactions on Circuits and Systems for Video Technology **21**(9) (2011) 1255–1262
15. Deng, W., Hu, J., Guo, J.: Extended SRC: Undersampled face recognition via intraclass variant dictionary. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(9) (2012) 1864–1870
16. Jia, K., Chan, T.H., Ma, Y.: Robust and practical face recognition via structured sparsity. In: European Conference on Computer Vision (ECCV 2012). Springer (2012) 331–344
17. Deng, W., Hu, J., Guo, J.: In defense of sparsity based face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013). (2013) 399–406
18. Tan, X., Chen, S., Zhou, Z.H., Liu, J.: Face recognition under occlusions and variant expressions with partial similarity. IEEE Transactions on Information Forensics and Security **4**(2) (2009) 217–230
19. Martínez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(6) (2002) 748–763
20. Wei, X., Li, C.T., Hu, Y.: Face recognition with occlusion using dynamic image-to-class warping DICW. In: IEEE International Conference on Automatic Face Gesture Recognition, (FG 2013). (2013)
21. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: Proceedings of the Second IEEE Workshop on Applications of Computer Vision. (1994) 138–142
22. Lee, K., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions of Pattern Analysis and Machine Intelligence **27**(5) (2005) 684–698
23. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Analysis and Modeling of Faces and Gestures. Springer (2007) 168–182
24. Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., Zhao, D.: The CAS-PEAL large-scale chinese face database and baseline evaluations. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans **38**(1) (2008) 149–161
25. Martinez, A., Benavente, R.: The AR face database (June 1998) CVC Tech. Rep, No. 24.
26. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst (October 2007)
27. Huang, G.B., Mattar, M.A., Lee, H., Learned-Miller, E.G.: Learning to Align from Scratch. NIPS (2012)

28. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008) http://www.vlfeat.org/.
29. Hartley, R.I., Zisserman, A.: Multiple view geometry in computer vision. Second edn. Cambridge University Press (2003)