Visible and Infrared Face Identification via Sparse Representation

Pierre Buyssens, Marinette Revenu

Hindawi Publishing Corporation
ISRN Machine Vision
Volume 2013, Article ID 579126, 10 pages
http://dx.doi.org/10.1155/2013/579126

Research Article

Visible and Infrared Face Identification via Sparse Representation

Pierre Buyssens[1] and Marinette Revenu[2]

[1] *LITIS EA 4108-QuantIF Team, University of Rouen, 22 Boulevard Gambetta, 76183 Rouen Cedex, France*
[2] *GREYC UMR CNRS 6072 ENSICAEN-Image Team, University of Caen Basse-Normandie, 6 Boulevard Maréchal Juin, 14050 Caen, France*

Correspondence should be addressed to Pierre Buyssens; pierre.buyssens@gmail.com

Received 4 April 2013; Accepted 27 April 2013

Academic Editors: O. Ghita, D. Hernandez, Z. Hou, M. La Cascia, and J. M. Tavares

We present a facial recognition technique based on facial sparse representation. A dictionary is learned from data, and patches extracted from a face are decomposed in a sparse manner onto this dictionary. We particularly focus on the design of dictionaries that play a crucial role in the final identification rates. Applied to various databases and modalities, we show that this approach gives interesting performances. We propose also a score fusion framework that allows quantifying the saliency classifiers outputs and merging them according to these saliencies.

1. Introduction

Face recognition is a topic which has been of increasing interest during the last two decades due to a vast number of possible applications: biometrics, video surveillance, advanced HMI, or image/video indexation. Although considerable progress has been made in this domain, especially with the development of powerful methods (such as the Eigenfaces or the Elastic Bunch Graph Matching methods), automatic face recognition is not enough accurate in uncontrolled environments for a large use. Many factors can degrade the performances of facial biometric system: illumination variation creates artificial shadows, changing locally the appearance of the face; head poses modify the distance between localized features; facial expression introduces global changes; artefacts wearing, such as glasses or scarf, may hide parts of the face.

For the particular case of illumination, a lot of work has been done on the preprocessing step of the images to reduce the effect of the illumination on the face. Another approach is to use other imagery such as infrared, which has been showed to be a promising alternative. An infrared capture of a face is nearly invariant to illumination changes and allows a system to process in all the illumination conditions, including total darkness like night.

While visual cameras measure the electromagnetic energy in the visible spectrum (0.4–0.7 μm), sensors in the IR respond to thermal radiation in the infrared spectrum (0.7–14.0 μm). The infrared spectrum can mainly be divided into reflected IR (Figure 1(b)) and emissive IR (Figure 1(c)). Reflected IR contains near infrared (NIR) (0.7–0.9 μm) and short-wave infrared (SWIR) (0.9–2.4 μm). The thermal IR band is associated with thermal radiation emitted by the objects. It contains the midwave infrared (MWIR) (3.0–5.0 μm) and long-wave infrared (LWIR) (8.0–14.0 μm). Although the reflected IR is by far the most studied, we use thermal long-wave IR in this study.

Despite the advantages of infrared modality, infrared imagery has other limitations. Since a face captured under this modality renders its thermal patterns, a temperature screen placed in front of the face will totally occlude it. This phenomenon appears when a subject simply wears glasses. In this case, the captured face has two black holes, corresponding to the glasses, which is far more inconvenient than in the visible

(a) Visible (b) Reflected IR (c) Emissive IR

Figure 1: A face captured under (a) visible spectrum, (b) reflected IR spectrum, and (c) emissive IR spectrum respectively.

modality. Moreover, thermal patterns can change due to external conditions such as weather. However, since these two modalities do not present the same advantages/limitations, using information of both can decrease the disadvantages of each and globally enhance the identification rates [1].

Two main schemes are considered in a biometric system [2].

(i) The verification (of authentication) aims to compare the unknown face with the one of a claimed identity. It is a one-to-one comparison scenario, which often involves a threshold step to accept/reject the probe.

(ii) The identification aims to find an unknown identity (probe) among a set of known identities (gallery).

Most of the approaches that have been proposed in the literature for the face recognition problem are built with the same three-step scheme:

(i) preprocessing of the images,

(ii) extraction of features from faces,

(iii) classification of these features.

Preprocessing Step. The first step intends to locate a face, resize it if necessary, and apply some algorithms to enhance the quality of the images. Illumination can also be corrected to simplify the features extraction.

Features Extraction Step. This second step consists in extracting salient features from faces. This strategy can globally be divided into two main approaches:

(i) the local approaches, which act locally on the face by extracting salient interest points (like eyes or mouths), and combine them into a global model;

(ii) the global approaches which often rely on a projection of the whole image onto a new low-dimensional space (these methods are then named Subspace methods).

Numerous local approaches based on geometrical features have been proposed in the literature [3–7].

The most popular local approach, named Elastic Bunch Graph Matching (EBGM) [8], consists in modeling the salient features (like nose, mouth, etc.) by a graph. To each node is associated a so-called jet which encodes the local appearance around the feature obtained via a Gabor filter. The classification of a probe graph involves then a specific algorithm that takes into account a geometric similarity measure and the appearance encoded by the jets.

The main advantages of these local approaches are their ability to deal with pose, illumination, or facial expression variations. Nevertheless, these approaches require a good localization of the discriminant features, which can be a difficult task in case of degradations of the image.

The global approaches often take the face image as a whole and perform a statistical projection of the images onto a face space. The most popular technique called Eigenfaces (first used by Turk and Pentland [9]) is based on a principal components analysis (PCA) of the faces. It has also been applied to infrared faces by Chen et al. [10]. Another popular technique is the *Fisherfaces* method based on a Linear Discriminant Analysis (LDA), which divides the face images into classes according to the Fisher criterion. It has been early applied by Kriegman et al. [11].

Note that the nonlinear versions Kernel-PCA and Kernel-LDA have been, respectively, applied in [12, 13].

The main drawback of the global approaches is their sensitivity to the illumination changes for the visible light modality and the thermal distribution of the face over time for the infrared modality. When the illumination (or the thermal distribution) of a face changes, its appearance undergoes a nonlinear transformation, and due to the linear projection often performed by these global approaches, the classification can fail. In the case of nonlinear projections, the choice of the kernel is critical and is a nontrivial problem. Moreover, as pointed out in [14], nonlinear dimensionality reduction methods can perform poorly on natural datasets.

Classification Step. The last step consists in classifying the extracted features. There are plenty of methods, simple ones based on distances between features via classification algorithms such as the Nearest Neighbor [15], others based on learning methods such as support vector machine [16] or neural networks [17]. However, these last methods have a significant drawback: they learn to recognize a fix number of identities, that is, classes. As the number of classes may vary by adding new identities to the system, for example, the design of the learning machine has to be updated and the learning recomputed. More recently, a seminal paper [18] has introduced a novel classification method relying on parsimony. The algorithm, named SRC for Sparse Representation-based Classification, decomposes in a sparse manner a probe feature vector $y \in \mathbb{R}^m$ onto a dictionary $A \in \mathbb{R}^{m \times n}$ composed of the n feature vectors of the gallery. As it mainly relies on a sparse decomposition problem, this algorithm requires $m < n$ in order to have an underdetermined system and a unique sparsest solution. More recent algorithms that use sparse decompositions have been proposed in the literature, such as robust sparse [19], group sparse [20], or structure sparse [21]. To ensure $m < n$, these algorithms first proceed to a dimension reduction via PCA (Eigenfaces) or other dimensionality reduction techniques. In our work, the extracted features of a face image are sparse and have a higher dimension than the images. Since the number of vectors of the gallery is less than the dimension of the vectors, such sparse-based classification algorithms cannot be used. Moreover, these classification algorithms make the assumption that a probe face lies onto a subspace specific to each individual. This assumption involves many faces of the same individual in the gallery, which is not the case of the databases used in our experiments. Finally, these algorithms are unusable in case of a one-to-one face comparison since the number of columns of A is 1. For all these reasons, this paper focuses on the feature extraction and makes use of the simple *Nearest Neighbor* algorithm as classifier.

This paper only considers the identification scheme. Assuming that the searched identity is always in the gallery, we focus on the rank-1 identification rates.

Contributions. This paper is a direct extension of our previous work [22]. A parameter exploration on the main parameters that pilot the dictionary design is presented. These learned dictionaries play a crucial role in the efficiency of the extracted features and then in the final identification rates. We propose also a framework for the fusion of different matchers at the score level. Based on a saliency function, it weights the outputs of a classifier without any assumptions.

The rest of the paper is organized as follows: Section 2 is dedicated to the proposed sparse features extraction method. Section 3 is devoted to the proposed score-based fusion method. Experimental results on various face datasets are presented in Section 4. Finally we present our conclusions and further work in Section 5.

2. Features Extraction

In this section, we present the proposed methodology for the features extraction and the fusion steps. After a brief recall of notations and definitions of the sparse decomposition theory, we detail the proposed scheme for the face features extraction and the fusion framework.

2.1. Notations and Definitions. An *atom* is an elementary basis element of a signal or an image. A collection of atoms (Φ_i) is called a dictionary $\boldsymbol{\Phi}$.

In this paper, the considered dictionaries are $N \times M$ matrices where the M columns represent the atoms (of size N) of the dictionary. When $r = M/N > 1$, the dictionary is overcomplete (with redundancy term r). In such a case, given a signal $\mathbf{x} \in \mathbb{R}^N$, the equation $\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\lambda}$ leads to an underdetermined system with an infinite set of solutions for $\boldsymbol{\lambda}$.

2.1.1. Sparse Decomposition. Given a signal $x \in \mathbb{R}^N$ (or an image of size $\sqrt{N} \times \sqrt{N}$), we are looking for its decomposition according to a dictionary $\boldsymbol{\Phi}$ composed of M vectors $\boldsymbol{\phi}_m$ recovering $\mathbb{R}^N$. Let us define first the $\mathscr{L}_p$ norm of a vector $\mathbf{x}$

$$\|\mathbf{x}\|_p = \left(\sum_i |x_i|^p \right)^{1/p} \tag{1}$$

with the particular case of the "$\mathscr{L}_0$ norm" (defined as the number of nonzero elements of $\mathbf{x}$):

$$\|\mathbf{x}\|_0 = \sum_{0 \le i < N} a_i, \quad \text{where } a_i = \begin{cases} 1, & \text{if } x_i \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

When the dictionary is over-complete ($M > N$), there are an infinite set of coefficients α_i that may be used to decompose the signal onto the dictionary:

$$\mathbf{x} = \sum_{m=1}^{M} \alpha_m \boldsymbol{\phi}_m. \tag{3}$$

In the sparse decomposition framework, the optimal solution is the one with the minimum of non-zeros elements (or the maximum of zeros elements). In this case, the problem is written:

$$\min_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_0 \quad \text{such that } \mathbf{x} = \sum_{m=1}^{M} \lambda_m \boldsymbol{\phi}_m. \tag{4}$$

Unfortunately, this problem is NP-hard.

Two approaches can be used to tackle this problem.

(i) The first one consists in a modification of the penalty term ($\|\mathbf{x}\|_0$) such that the problem becomes convex. Also known as Basis Pursuit (BP) [23] when turning the "$\mathscr{L}_0$" into an $\mathscr{L}_1$ norm, this approach gives equal results to the original problem under certain conditions (see [24] for more details). The problem becomes then

$$\min_{\boldsymbol{\lambda}} \left(\left\| \mathbf{x} - \sum_{m=1}^{M} \lambda_m \boldsymbol{\phi}_m \right\|_2^2 + \mu \|\boldsymbol{\lambda}\|_1 \right). \tag{5}$$

Numerous algorithms have been developed for this problem resolution (also known under the name Lasso for Least Absolute Shrinkage and Selection Operator) based on the interior point method [25] or on iterative thresholding [26].

(ii) The second method usually used in the community is based on greedy algorithms that build iteratively a sparse representation of a signal [27]. The class of Matching Pursuit (MP) algorithms selects at each iteration the atom that minimizes the residual between the signal and the reconstruction obtained at the last iteration. More details on the well-known variant Orthogonal Matching Pursuit can be found in [28].

2.1.2. Dictionary Learning. An overcomplete dictionary $\mathbf{\Phi}$ that leads to sparse representations can be chosen as a predefined set of functions adapted to the signal. For certain class of signals, this choice is appealing because it leads to simple and fast algorithms for the evaluation of the sparse decomposition. This is the case for overcomplete wavelets, curvelets, ridgelets, bandelets, Fourier transforms and more. Due to the morphological diversity contained in a natural image, it is often preferable to concatenate such basis to obtain the dictionary. Another way of constructing the dictionary is to learn it directly from data.

Many methods have been developed to perform this task such as those based on maximum likelihood [29–31], the one named Modeling of Optimal Directions (MOD) [32, 33], or those based on the a posteriori maximum [34, 35].

In this paper, we use the K-SVD algorithm proposed in [36] based on a singular value decomposition, which can be viewed as a generalization of the K-means, hence its name. Starting from a random initialization of the atoms, learning the dictionary proceeds in an iterative way, alternating the two steps:

(i) minimize (5) with respect to $\mathbf{x}$ keeping the dictionary elements ϕ_m constant;

(ii) update the atoms ϕ_m of the dictionary with $\mathbf{x}$ found at previous step.

2.2. Features Extraction Methodology. In this paper, we use sparse decompositions as features for the face identification. An appealing way would be to directly decompose faces onto a dictionary learned with a set of faces. This scheme is however impractical in practice for the following reasons.

(i) As a good sparse decomposition involves an overcomplete dictionary, one has to dispose of a dictionary whose size is at least equal to the signal dimension. As the signal (an image) is high-dimensional, the dictionary would be huge, and the decomposition would be very slow.

(ii) Because of the morphological diversity contained in the images of faces, a sparse decomposition would be more efficient when processed on a learned dictionary, which involves a number of training samples at least equal to the size of the dictionary.

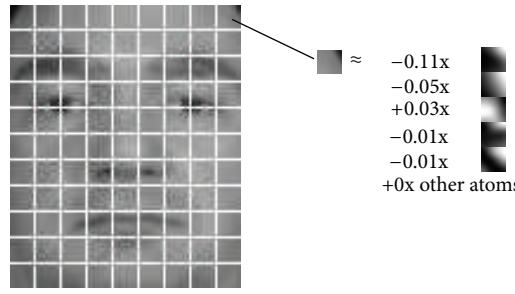


Figure 2: Schematic view of the feature extraction process.

For example, with images sizes of 40×50 (which is small for the face recognition task), the minimum number of atoms as well as the minimum number of training samples would be 2000. Moreover, within the K-SVD algorithm, one has to apply a singular value decomposition on matrices whose height is equal to the atoms dimension, which can be impractical in case of high-dimensional data. For these reasons, the sparse decomposition is processed on parts of the images.

Once the preprocessing is applied, the sparse features extraction of a face image acts in 3 steps:

(i) the image is splitted into n nonoverlapping square patches of size $\Gamma \times \Gamma$;

(ii) each patch is independently decomposed into a sparse vector $\mathbf{x}_k$ ($k \in \{1 \cdots K\}$) onto a dictionary $\mathbf{\Phi}$ by minimizing (5);

(iii) the sparse vectors $\mathbf{x}_k$ are concatenated to form the final sparse feature vector $\mathbf{x}$ of the face.

A schematic view of the feature extraction process is shown in Figure 2.

In a first time, the dictionary used for the decomposition of the patches is learned from data with the algorithms OMP for the sparse code computation and K-SVD to update of the atoms.

In a second time, the features are computed with the algorithm FISTA proposed in [37] based on a two-step iterative soft-thresholding, which is a fast algorithm that solves (5).

Size of the Features. The size of the features depends on several parameters:

(i) the size Γ of the patches,

(ii) the redundancy r of the dictionary,

(iii) the size $w \times h$ of the image.

Since each extracted patch is decomposed onto the dictionary (composed of m atoms), and the feature vector is the concatenation of the p extracted patches, the size of a feature vector is computed as

$$\text{size} = p \times m, \tag{6}$$

where $m = r \times \Gamma^2$ and $p = \lceil (w/\Gamma) \times (h/\Gamma) \rceil$. If w (or h) is not divisible by Γ, the image is padded with zero. This padding

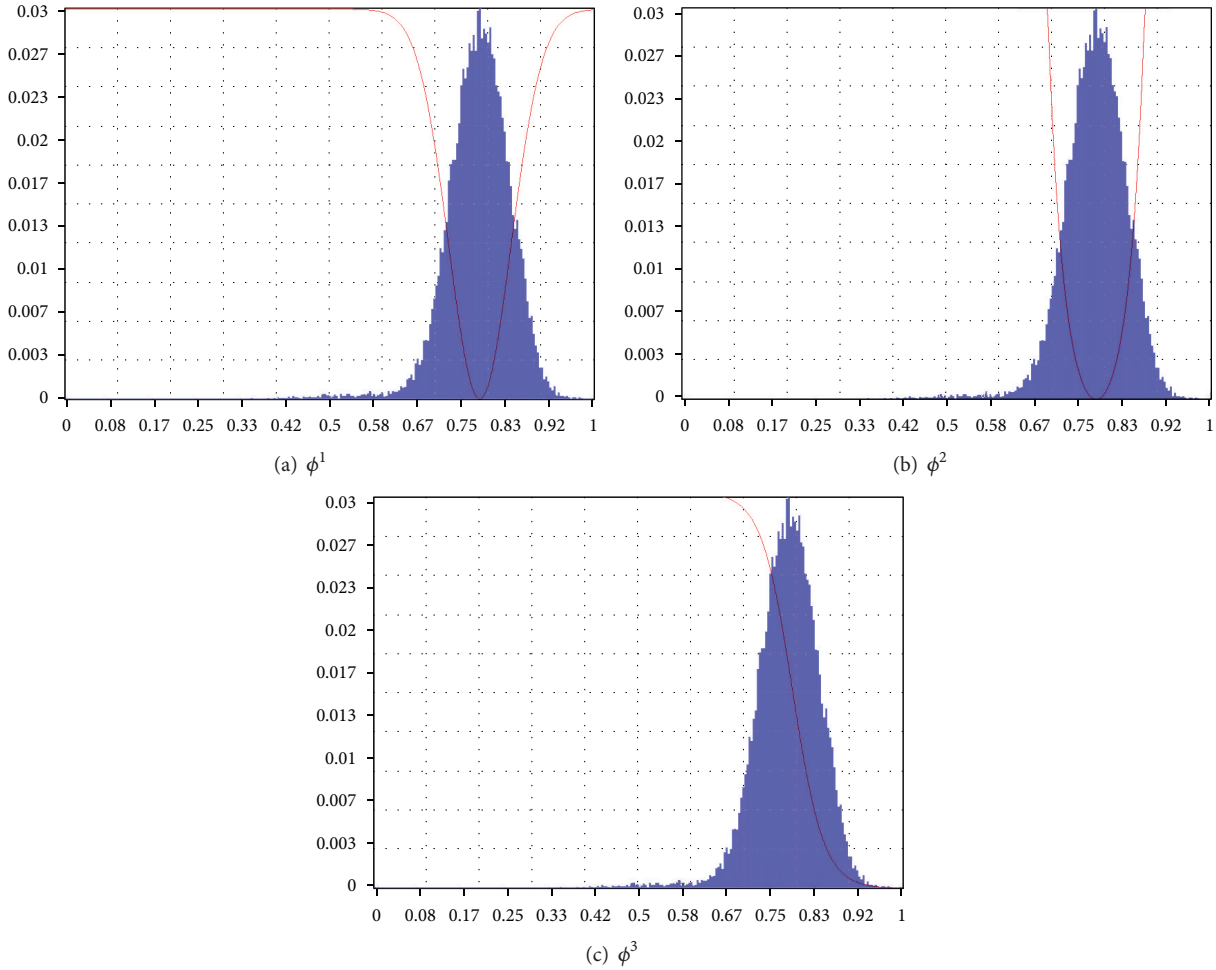


Figure 3: Distribution of the outputs of a classifier in blue. Associated saliency function in red.

has no effect on the recognition behavior since all the images are padded in the same way.

The dimension of the resulting feature vectors may be quite high (higher than the image dimension) but are very sparse, that is, containing few nonzero entries.

3. Score Fusion

Given classifiers that yield score rankings as results, we consider a fusion framework that weights the outputs of these classifiers without any assumptions.

Assuming that classifiers do not have the same accuracy, we propose a merging methodology that uses measures of saliency computed dynamically for each classifier. This fusion scheme can be divided into three steps:

(i) the scores produced by different classifiers may be heterogeneous, so a normalization step is required. Several normalization methods exist such as linear, logarithmic, or exponential normalizations;

(ii) a function of saliency is computed onto a score distribution according to some statistical measure, and a unique saliency value is attributed to each score;

(iii) final scores are computed as a weighted sum of the scores according to the saliencies.

Given a probe sample I, the distances to the labeled samples of the gallery $\mathbf{G}$ are computed giving a distribution of distances $\mathbf{D}$:

$$\begin{aligned} \mathbf{D} &= \{d_k\}, \\ d_k &= \|I - \mathbf{G}_k\|, \end{aligned} \tag{7}$$

where $\mathbf{G}_k$ is a feature vector of a gallery sample. After a normalization of $\mathbf{D}$, its mean μ and standard deviation σ are computed.

A saliency s_k is then given to each d_k according to a function depending on μ and σ:

$$s_k = \phi_{\mu,\sigma}(d_k). \tag{8}$$

In this work, we propose three saliency functions ϕ^1, ϕ^2, and ϕ^3 (Figure 3) that are of the form

$$\phi^1_{\mu,\sigma}(d_k) = \frac{1}{\sigma\sqrt{2\pi}}\left(1 - e^{-(1/2)((d_k-\mu)/\sigma)^2}\right),$$
$$\phi^2_{\mu,\sigma}(d_k) = \sigma\sqrt{2\pi}\frac{1}{e^{-(1/2)((d_k-\mu)/\sigma)^2}}, \quad (9)$$
$$\phi^3_{\mu,\sigma}(d_k) = \left(1 + \frac{1}{2}\tanh\left(\frac{1}{\sigma}(d_k - \mu)\right)\right)^{-1}.$$

This fusion scheme works with any 2-class classifiers that give a distance (or a similarity) measure as output. The saliency functions allow weighting the output of a classifier according to its response on a set of inputs (the gallery), without any ad hoc assumptions.

Note that the proposed functions deal with distances measures, but other functions can easily be used with similarity measures.

Saliency functions ϕ^1 and ϕ^2 tend to highly weight an uncommon measure, even if it is high (i.e., a probe sample far from any gallery samples). ϕ^3 specializes this idea by more penalizing higher distances than common distances and then favors small distances.

For a given classifier, this procedure then gives a distribution of distances which are weighted by their respective saliencies.

Given several classifiers C_i, the final fusion distances are computed as a weighted sum of the outputs:

$$d_k = \frac{\sum_i d_{k_i} \times s_{k_i}}{\sum_i s_{k_i}} \quad \forall k. \quad (10)$$

As for the single-classifier experiments, the classification is performed via the *Nearest Neighbor* classifier.

4. Experiments and Results

In this section, we detail the experiments of both feature extraction performance and score fusion on different public databases. In all the experiments, the images are cropped to ensure that the eyes are roughly at the same position and scaled to the size 110×90.

4.1. Extended Yale B Database. The extended Yale B database is composed of 2414 frontal-face images of 38 individuals [38]. Faces were captured under various laboratory-controlled lighting conditions. This experiment is mainly dedicated to show the effectiveness of the approach in term of recognition rates. The main parameters that pilot the dictionary learning are fixed to:

$$r = 2,$$
$$\Gamma = 10, \quad (11)$$
$$n_{\text{OMP}} = 5.$$

With these parameters, the size of a face feature vector is $11 \times 9 \times 2 \times 10^2 = 19800$.

The dictionary is learned with a small number of images of the database. Figure 4 shows the atoms of the learned dictionary. One can see that some atoms encode low-frequency patterns, while others are more oriented edge selective. The database is then divided into disjoint training and testing parts (as in [18]), and face images are decomposed onto the dictionary following the methodology explained in Section 2.2. For the experiments, we randomly select 8, 16, and 32 images per individuals for training. Randomly dividing the database ensures that the results do not depend on a favorable choice of the training set. The mean rank-1 identification rates over 10 different executions are shown in Table 1. Identification rates are competitive with those given in [18], although our method performs the classification with a simple *Nearest Neighbor* classifier.

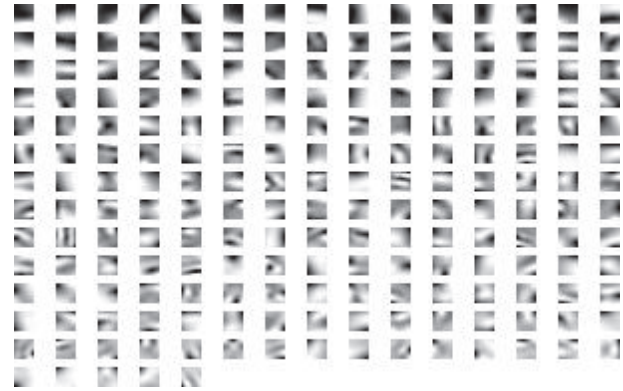


Figure 4: Learned atoms for $\Gamma = 10$ (patches size: 10×10) and $n_{\text{OMP}} = 5$ sorted by variance.

4.2. FERET Database. The FERET database [39] is a well-known database composed of thousands of individuals. We focus on two subsets named *fa* and *fb*:

(i) *fa* contains 994 images of 994 individuals (one image per individual) and is used as gallery;

(ii) *fb* contains 992 images of 992 individuals (one image per individual) and is used as probe.

This experiment is mainly dedicated to evaluate the proposed score fusion methodology.

To this end, we extract simple random features from faces. Linear random projections are generated by Gaussian random matrices, hence the name of this technique RandomFaces [40]. A random projection matrix is extremely efficient to generate. Its entries are independently sampled from a zero-mean normal distribution, and each row is normalized to unit length.

In this experiment, we generate three different random projection matrices to map the faces to three random subspaces of dimension 50, 100, and 150.

Various normalization techniques and fusion methods have been implemented for comparison purposes. Normalization techniques used are the MinMax (MM), the Decimal Scaling (DeSc), the Z-Score (ZS), the Median Absolute Deviation (MAD), and the Hyperbolic Tangent (Tanh) techniques (see [41] for more details on these normalization techniques). Fusion methods used are classical ones of the literature [42]:

Table 1: Main results. (a) Identification rates on the extended Yale B database. (b) Identification rates for each random space on the FERET database and a comparison between different normalization and fusion techniques (best score per column in bold). (c) Identification rates on the Notre-Dame database obtained by score fusion of visible and infrared modalities (best score per column in bold).

(a)

Extended Yale B database			
Number of images	8	16	32
Identification rate %	88.39	95.74	98.04

(b)

FERET database					
Random subspace dimension			50	100	150
Identification rate %			67.43	73.28	74.69
	MM	DeSc	ZS	MAD	Tanh
PROD	68.95	77.92	77.92	77.92	78.12
SUM	78.02	73.18	78.12	78.12	78.12
MAX	77.21	67.64	0.10	0.10	75.30
MIN	68.95	**73.48**	74.19	74.29	74.19
ϕ^1	78.32	64.51	78.22	78.22	78.22
ϕ^2	76.00	13.91	74.79	74.39	74.79
ϕ^3	**78.42**	73.08	**78.42**	**78.42**	**78.42**

(c)

Notre-Dame database					
	MM	DeSc	ZS	MAD	Tanh
PROD	83.12	83.12	83.12	83.12	83.12
SUM	92.32	78.52	93.17	93.18	87.41
MAX	89.86	75.40	84.99	83.99	87.41
MIN	83.12	**87.41**	92.82	91.99	75.47
ϕ^1	85.86	11.19	91.80	91.15	57.77
ϕ^2	92.83	73.50	93.02	92.11	87.41
ϕ^3	**93.47**	51.71	**94.06**	**93.88**	**87.99**

the Product rule (PROD), the Sum rule (SUM), the Max rule (MAX), and the Min rule (MIN). Note that other score fusion methods exist such as the one based on a Gaussian Mixture Model [43], but they often rely on the need of several biometric samples from the same individual, which is not the case of our experiments.

Table 1 summarizes the identification rates of RandomFaces together with the different score fusion techniques. Despite the relative high number of individuals (about 1000), the difficult one-image-to-enroll scenario, and the simple extracted features, the identification rates are quite high (over 78%), and the proposed fusion scheme almost always outperforms the classical score fusion methods.

4.3. Notre-Dame Database. The database from the University of Notre-Dame (Collection $X1$) [42] is a public collection of 2D visible/thermal face images. This database has two advantages:

(i) a visible picture and its thermal counterpart are taken at the same time;

(ii) a well-defined test protocol is included with the database, which allows a fair comparison between previously published results on this database.

4.3.1. Details of the Database. The database is divided into two disjoints parts: the first one, named Train Set, is composed of 159 subjects. For each, one visible and one thermal images are available. The second part, named Test Set, is composed of 82 subjects. This set contains 2292 visible images and 2292 thermal images.

While the *Train Set* contains neither facial expressions nor illumination/thermal variations, the *Test Set* contains such variations.

Two experiments, named same session and time lapse, have been designed to test the facial identification algorithms across illumination variations and through time, respectively. In this work, we do not report identification rates on the same session experiment since it contains too few images and is too easy: most of the classical face recognition algorithms obtain identification rates close to 100%. We then report only the identification rates on the time lapse experiment which is a more challenging subdataset. The pictures have been taken within weeks/months which involves variations in faces appearance.

For this experiment, the test protocol consists in 16 subexperiments allowing picking galleries and probes of different facial expressions (neutral or smiling) and different lighting (FERET or Mugshot styles).

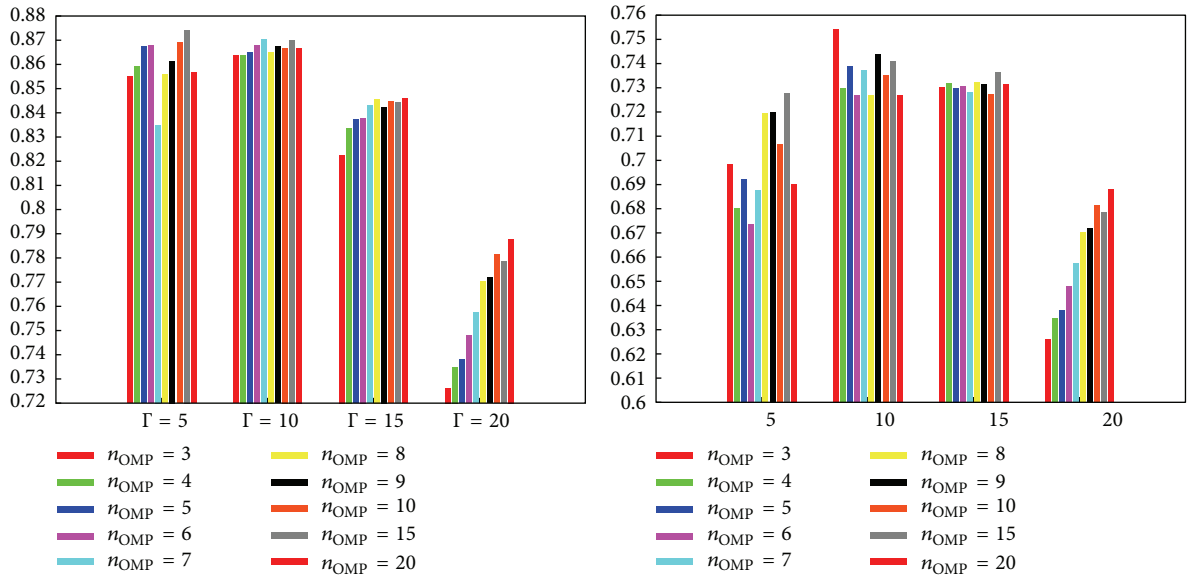
Note that each gallery contains only one image per subject (one-image-to-enroll scenario).

4.3.2. Details of the Experiment. Our experiment is mainly dedicated to a parameter exploration of the main parameters that pilot the dictionary design and to evaluate these parameters on the final identification rates.

The experiments have been conducted with different values of the considered hyperparameters: the size $\Gamma \times \Gamma$ of the square patches and the maximum number of atoms n_{OMP} allowed for the sparse decomposition within the algorithm OMP. These hyperparameters directly influence the learned dictionary, and then the extracted feature vectors. A grid search is performed onto these two parameters: Γ varies into $\{5, 10, 15, 20\}$ and n_{OMP} in $\{3, 4, 5, 6, 7, 8, 9, 10, 15, 20\}$. Note that each experiment is performed separately for the visible and infrared modality.

In order to learn the dictionary, for each couple (Γ, n_{OMP}), 10000 patches of size $(\Gamma \times \Gamma)$ with sufficient standard deviation (to avoid too uniform patches) are randomly extracted from the Train Set. The maximum number of atoms allowed for the OMP algorithm is then fixed to n_{OMP}, which means that each training pattern is decomposed into a sum of n_{OMP} atoms, the coefficients of the other atoms being 0.

For all the experiments, the redundancy of the dictionary is set to 2 which means $2 \times \Gamma^2$ atoms to learn. The learning process is applied until convergence.



(a) Visible (b) IR

FIGURE 5: Rank-1 mean identification rates for different values of Γ and n_{OMP}. (a) Visible, (b) IR.

For each couple (Γ, n_{OMP}), a dictionary is learned, then the face features are extracted following the proposed scheme (Section 2).

4.3.3. Results. Figures 5(a) and 5(b) show the identification results at rank-1 for different values of Γ and n_{OMP} for the two modalities. For the sake of clarity, identification rates have been averaged: each bin represents the mean identification rate of the 16 subexperiments of the time lapse experiment.

Although the best identification rate for the visible modality (87.41%) is obtained with $\Gamma = 5$ and $n_{OMP} = 15$, one can see that identification rates for $\Gamma = 10$ are the most stable according to n_{OMP}, each bin exceeding 86%. Identification rates with $\Gamma = 15$ are worse, and those obtained with $\Gamma = 20$ are the worst.

Similar results can be observed with the infrared modality. Although identification rates seem more stable with $\Gamma = 15$, results with $\Gamma = 10$ are globally the better (73.60% avg.).

4.3.4. Modality Fusion. Identification rates obtained above show that visible modality performs better than LWIR modality. This result has already been showed in [42] or [44].

Best couples (Γ, n_{OMP}) for each modality found above are retained, and the fusion scheme presented in Section 3 is performed. As for the FERET database experiment, our score fusion scheme is compared with various normalization and fusion techniques. A summary the results and a comparison of identification rates previously published in the literature are shown in Table 2. Our method outperforms other methods in visible modality but gives lower identification rates in infrared. The lack of texture in this modality could explain that our sparse features approach gives such identification rates. Note that we previously published in [22] better identification rates with these sparse features conjointly classified with the Sparse Representation-based Classification algorithm (SRC, [18]). Nevertheless, these results are not completely exact since the dimension of the features exceeds the number of elements of the gallery, which implies an overdetermined system within the SRC algorithm.

TABLE 2: Comparison of methods for the time-lapse experiment of the Notre-Dame database. Mean identification rates over the 16 subexperiments, standard deviation in parenthesis. Best score in bold.

	Time lapse		
	[42]	[44]	This paper
Visible	82.66	72.50	**87.41**
	(7.75)	(4.01)	(4.32)
IR	**77.81**	40.06	75.40
	(3.31)	(3.47)	(2.60)
Fusion	92.5	80.12	**94.06**
	(2.71)	(4.13)	(2.08)

5. Conclusion and Future Work

We presented a facial feature extraction method based on sparse decompositions of patches of face images. It decomposes a face image onto a dictionary that has been learned from data. Applied to various databases and modalities, it offers comparable identification results to the state-of-the-art on the Notre-Dame database according to its specific protocol.

Modalities fusion offers an alternative to unimodal biometric systems. From the hypothesis that different modalities can offer complementary informations (which is often the

case), fusion of these allows enhancing the reliability of a system.

We proposed a decision level fusion scheme based on a per-score measure of saliency. It does not depend on ad hoc assumptions and allows increasing rank-1 identification rates. Moreover, it is sufficiently general to be used with any number of features, biometrics, or classifiers.

Further work will involve the integration of our feature extraction scheme into a multiscale approach. A better selection could also enhance final decision scores. In this work, every patch is equally treated, even those containing hair, for example. This is obviously suboptimal and a selection or a weighting of discriminant patches will improve identification rates.

A limitation of our approach is also that faces have to be carefully aligned. The extracted features may not be robust to pose changes. However, recent works on the design of dictionaries that are robust to affine transformations could help to tackle this limitation. verification (of authentication) aims to compare the unknown face with the

References

[1] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent advances in visual and infrared face recognition—a review," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103–135, 2005.

[2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.

[3] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, 1993.

[4] Y. Gao and M. K. H. Leung, "Face recognition using line edge map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 764–779, 2002.

[5] B. Heisele, P. Ho, J. Wu, and T. Poggio, "Face recognition: component-based versus global approaches," *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 6–21, 2003.

[6] J. R. Price and T. F. Gee, "Face recognition using direct, weighted linear discriminant analysis and modular subspaces," *Pattern Recognition*, vol. 38, no. 2, pp. 209–219, 2005.

[7] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*, pp. 138–142, December 1994.

[8] L. Wiskott, J. M. Fellous, N. Krüger, and C. D. Von Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[9] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, June 1991.

[10] X. Chen, P. J. Flynn, and K. W. Bowyer, "PCA-based face recognition in infrared imagery: baseline and comparative studies," in *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 127–134, IEEE Computer Society, 2003.

[11] D. J. Kriegman, J. P. Hespanha, and P. N. Belhumeur, "Eigenfaces vs. fisherfaces: recognition using class-specific linear projection," in *Proceedings of the European Conference on Computer Vision*, vol. 1, pp. 43–58, IEEE, 1996.

[12] B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[13] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Muller, "Fisher discriminant analysis with kernels," in *Proceedings of the 9th IEEE Workshop on Neural Networks for Signal Processing (NNSP '99)*, pp. 41–48, August 1999.

[14] L. J. P. van der Maaten, E. O. Postma, and H. J. van den Herik, "Dimensionality reduction: a comparative review," Tech. Rep., 2009.

[15] S. Yang and C. Zhang, "Regression nearest neighbor in face recognition," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, pp. 515–518, August 2006.

[16] H. J. Jia and A. M. Martinez, "Support vector machines in face recognition with occlusions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 136–141, 2009.

[17] L. Bai and Y. Liu, "Neural networks and wavelets for face recognition," in *Proceedings of the 4th International Conference on Enterprise Information Systems (ICEIS '02)*, pp. 334–340, Ciudad Real, Spain, 2002.

[18] 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*, vol. 31, no. 2, pp. 210–227, 2009.

[19] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 625–632, IEEE, 2011.

[20] Y.-W. Chao, Y.-R. Yeh, Y.-W. Chen, Y.-J. Lee, and Y.-C. F. Wang, "Locality-constrained group sparse representation for robust face recognition," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, B. Macq and P. Schelkens, Eds., pp. 761–764, IEEE, Brussels, Belgium, 2011.

[21] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 1873–1879, Providence, RI, USA, 2011.

[22] P. Buyssens and M. Revenu, "IR and visible identification via sparse representation," in *Biometrics: Theory, Applications and Systems*, IEEE, Was, USA, 2010.

[23] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," Tech. Rep., Department of Statistics, Stanford University, 1996.

[24] D. L. Donoho and M. Elad, "Maximal sparsity representation via l^1 minimization," *Proceedings of National Academy of Sciences*, vol. 100, pp. 2197–2202, 2003.

[25] C. Meszaros, "On the sparsity issues of interior point methods for quadratic programming," Tech. Rep., Laboratory of Operations Research and Decision Systems, Hungarian Academy of Sciences, 1998.

[26] M. J. Fadili and J.-L. Starck, "Sparse representation-based image deconvolution by iterative thresholding," *Astronomical Data Analysis*, vol. 6, 18, 2006.

[27] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[28] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," Tech. Rep., inst-courant-cs, 1992.

[29] M. S. Lewicki, H. Hughes, and B. A. Olshausen, "A probabilistic framework for the adaptation and comparison of image codes," *Journal of the Optical Society of America*, vol. 16, no. 7, pp. 1587–1601, 1998.

[30] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[31] K. Engan, S. O. Aase, and J. H. Husoy, "Multi-frame compression: theory and design," *Signal Processing*, vol. 80, no. 10, pp. 2121–2140, 2000.

[32] K. Engan, B. Rao, and K. Kreutz-Delgado, "Frame design using FOCUSS with method of optimized directions (MOD)," in *Proceedings of the Nordic Signal Processing Symposium*, pp. 65–69, Oslo, Norway, September 1999.

[33] K. Engan, S. O. Aase, and J. H. Husoy, "Frame based signal compression using method of Optimal Directions (MOD)," in *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems (ISCAS '99)*, pp. V-1–V-4, Orlando, Fla, USA, June 1999.

[34] K. Kreutz-Delgado and B. D. Rao, "Focuss-based dictionary learning algorithms," in *Wavelet Applications in Signal and Image Processing*, vol. 41, pp. 19–53, IEEE, 2000.

[35] J. F. Murray and K. Kreutz-Delgado, "An improved focuss-based learning algorithm for solving sparse linear inverse problem," in *Proceedings of the International Conference on Signals, Systems and Computers*, vol. 41, pp. 19–53, 2001.

[36] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: design of dictionaries for sparse representation," in *Proceedings of the Signal Processing with Adaptive Sparse Structured Representations (SPARS '5)*, vol. 5, pp. 9–12, 2005.

[37] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[38] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[39] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," Tech. Rep., 1999.

[40] S. Kaski, "Dimensionality reduction by random mapping: fast similarity computation for clustering," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '98)*, vol. 1, pp. 413–418, 1998.

[41] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[42] X. Chen, P. J. Flynn, and K. W. Bowyer, "IR and visible light face recognition," *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 332–358, 2005.

[43] J. Wang, Y. Li, X. Ao, C. Wang, and J. Zhou, "Multi-modal biometric authentication fusing iris and palmprint based on GMM," in *Proceedings of the IEEE/SP 15th Workshop on Statistical Signal Processing (SSP '09)*, pp. 349–352, September 2009.

[44] P. Buyssens, M. Revenu, and O. Lepetit, "Fusion of IR and visible light modalities for face recognition," in *Proceedings of the 3rd International Conference on Biometrics: Theory, Applications and Systems (BTAS '09)*, Wash, USA, September 2009.