Sparse Multi-Stage Regularized Feature Learning for robust Face Recognition

Mohamed Anouar Borgi[1], Demetrio Labate[2], Maher El Arbi[1], Chokri Ben Amar[1]

[1]Research Groups on Intelligent Machines,
University of Sfax, BP 1173,
Sfax 3038, Tunisia

[2] Department of Mathematics,
University of Houston, Houston,
TX 77204, USA
{anoir.borgi@ieee.org; dlabate@math.uh.edu; maher.elarbi@ieee.org; chokri.benamar@ieee.org}

Abstract—The major limitation in current facial recognition systems is that they do not perform very well in uncontrolled environments, that is, when faces present variations in pose, illumination, facial expressions and environment. This is a serious obstacle in applications such as law enforcement and surveillance systems. To address this limitation, in this paper we introduce an improved approach to ensure robust face recognition, that relies on two innovative ideas. First, we apply a new multiscale directional framework, called Shearlet Network (SN), to extract facial features. The advantage of this approach comes from the highly sparse representation properties of the shearlet framework that is especially designed to robustly extract the fundamental geometric content of an image. Second, we apply a refinement of the Multi-Task Sparse Learning (MTSL) framework to exploit the relationships among multiple shared tasks generated by changing the regularization parameter during the recognition stage. We provide extensive numerical tests to show that our Sparse Multi-Regularized Shearlet Network (SMRSN) algorithm performs very competitively when compared against different state-of-the-art methods on different experimental protocols, including face recognition in uncontrolled conditions and single-sample-per-person.

Keywords—*Neural Networks*, *Shearlets, Sparsity, Shearlet Networks, Wavelet Networks, Face Recognition.*

I. INTRODUCTION

Face recognition (FR) is one of the main areas of investigation in biometrics and computer vision. It has a wide range of applications, including access control, information security, law enforcement and surveillance systems. Despite the intense research activity in this area, however, many fundamental challenges remain and "at their current level of development, current facial recognition systems show promise but are not yet advanced enough to be considered mature technologies", according to a statement recently found on the face recognition page of the RAND Corporation (http://www.rand.org/natsec_area/products/facialrecog.html). In fact, current FR algorithms perform rather well under controlled conditions (i.e., with each subject looking directly into the camera, with good illumination, and filling the area of the photo completely) but the recognition rates suffer dramatically in the presence of variations in pose, illumination and facial expression. This is in contrast with the remarkable human ability for recognizing faces: an infant innately responds to face shapes at birth and can discriminate his/her mother's face from a stranger's within about 2 days. Thus, there is a need to develop more advanced face recognition methodologies that are capable of providing accurate recognition under realistic conditions. This is crucially important for all applications mentioned above.

The main objective of the ongoing research in FR is to improve the robustness of the algorithms with respect to the various factors indicated above and many of the methods proposed during the last decade have been trying to take advantage of the latest advances in statistical learning. In particular, an important class of recent FR methods focuses on the notion of *sparsity* to extract the salient facial features in such a way to obtain robust classification with respect to a wide range of changes in pose, illumination, etc. In nutshell, the philosophy underlying sparsity is that the essential information embedded in most multidimensional phenomena is intrinsically low dimensional, that is, there exists a sparse representation for such phenomena (Donoho et al. 1998). The human ability to recognize faces can be interpreted as a manifestation of our ability to sparsely represent perceptual data or, in other words, to efficiently reduce the dimensionality of the data.

One of the main novelties of the method proposed in this paper is the use of the shearlet representation to extract the essential geometric content of facial features. This approach, pioneered by one of the authors, is a powerful extension of the classical wavelet framework that combines multiscale analysis with directional sensitivity. Together with the curvelet approach (that shares some of its properties) the shearlet representation is the only method known to provide optimally sparse representations for images with edges (Guo and Labate, 2007), so that it can capture most efficiently edges and other landmark in images. These properties go far beyond the basic capabilities of PCA and their variants as well as conventional multiscale methods.

The other main novelty of our approach is the way we use the features extracted to learn from the data. The method that we use is a refinement of multi-task learning MTL, originally proposed by Caruana (Caruana, 1997). MTL attempts to learn classifiers for multiple tasks jointly and works under the assumption that all tasks should share some common features. In our algorithm, we use this idea to exploit the relationships among multiple shared stages of our recognition algorithm in such a way to obtain more robust recognition.

Thus, by combining the power of the sparse shearlet representation together with our refined version of multi-task learning, we introduce an improved framework for robust face recognition that we call Sparse Multi-Regularized Shearlet Network (SMRSN). This method includes a multi-regularization step inspired from multi-stage convex relaxation (Zhang, 2010) to upgrade from a non-convex optimization to a convex relaxation. As part of this work, we have extensively tested the performance of our FR algorithm under challenging uncontrolled conditions (i.e., changes of pose and illumination, and single-subject-per-person) and compared it against other standard and state-of-the-art algorithms. As we will show below, the performance of our method is outstanding.

The rest of this paper is organized as follows. In Sec. 2, we describe the related work on regularization theory and MTL. In Sec. 3, we provide the necessary background on shearlets. In Sec. 4, we present the proposed Sparse Multi-Regularized Shearlet Network (SMRSN) algorithm. In Sec. 5, we present extensive numerical experiments to demonstrate the efficacy of the proposed algorithm and compare it against standard and state-of-the-art methods. In Sec. 6, we have concluding remarks.

II. RELATED WORK

Face recognition has been an active field of research for more than two decades and many approaches have been proposed. Among the most classical and representative methods, we recall Eigenface (Turk and Pentland, 1991), Fisherface (Belhumeur et al., 1997) and SVM (Heisele et al., 2001). As we mentioned above, the focus of current research is the development of improved FR algorithms that are highly efficient also in *uncontrolled*

environments. One of the most successful contributions appeared in recent years is the algorithm proposed by (Wright et al., 2009), that applies a sparse coding technique to FR called Sparse Representation-based Classification (SRC). By coding a test (or query) image as a sparse linear combination of all the training samples, SRC classifies the image by evaluating which class could result in the minimal reconstruction error. One of the remarkable properties of this approach is its robustness to face occlusion and corruption which is due in part to the sparsity constraint applied on the coding coefficients. Following this idea, several variants of SRC have been proposed, such as the Regularized Robust Coding (RRC) by (Yang et al., 2011 and Yang et al., 2013), that aims to reduce the computational cost of SRC and improve robustness. Its main idea consists in using a linear regression approach with regularized regression coefficients where, by assuming that the coding residual and the coding coefficient are respectively independent and identically distributed, RRC seeks for a maximum a posterior solution of the coding problem. The approach that we adopt in this paper follows the general idea of RRC. Also in our method, that we will describe in the Section IV, after encoding the image information using the shearlet framework, we will represent faces as sparse linear combinations of the training faces to achieve a classification. In addition to taking advantage of shearlets to efficiently encode the geometric information of images, the main difference of our approach with respect to RRC, is that we use a more sophisticated regularization approach for classification, which is inspired by ideas from Multi-Task-Learning.

Recall that, among the classification methods recently remerged in the machine learning literature, Multi-Task Learning (MTL) has been especially successful due to its remarkable performance. MTL aims to learn classifiers for multiple tasks jointly and works under the assumption that different tasks should share some common features. Many variants of MTL were proposed, including the multi-stage multi-task feature learning (MSMTFL) recently introduced by (Gong et al., 2013), who defines a non-convex formulation for multi-task feature learning based on a novel non-convex regularization, called *capped-ℓ_1, ℓ_1 regularized model* for multi-task feature learning. This approach aims to simultaneously learn the features specific to each task as well as the common features shared among tasks. A variant of this approach was proposed by (Zhang, 2010 and Zhang, 2012) whose main novelty is a multi-stage convex relaxation scheme for solving problems with non-convex objective functions. Despite its success, very few publications have attempted to study the FR problem using MTL. One notable contribution in this direction is the work of (Wang et al., 2009), that presents a modified multi-task learning (MTL) framework, called boosted MTL, for face verification with limited training data. This algorithm learns classifiers for multiple persons by sharing a few boosting classifiers in order to avoid overfitting. In our algorithm, we will adopt a refined version of MTL that uses a multi-task regularization to update the residual at every stage and a more effective approach to share classifiers. This improvement will be discussed in detail in Sec. IV.

III.   SHEARLET

The shearlet transform, introduced by one of the authors and his collaborators (Labate et al., 2005), is a multiscale framework for image analysis that is especially designed to represent information not only across several scales and locations, as the conventional wavelet transform, but also across several orientations, in such a way that it can more efficiently encode geometric features such edges and other landmarks in images. The

analyzing shearlet functions are scaled using a combination of shear matrices B_s, $s \in \mathbb{R}$ and anisotropic dilation matrices A_a, $a > 0$, defined by:

$$B_s = \begin{pmatrix} 1 & \text{-}s \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix} \tag{1}$$

It follows that the shearlet filters are highly anisotropic and can be oriented along any directions as illustrated in Fig. 1. Thanks to these properties, the shearlet transform has the ability to detect very efficiently the geometry of edges and other elongated features that usually constitute the dominating landmarks in many types of images, most notably face images. Related to this, shearlets provide optimally sparse approximation properties for the class of cartoon-like images, a simplified model of natural images, outperforming conventional wavelets and other traditional methods (Guo and Labate, 2007).
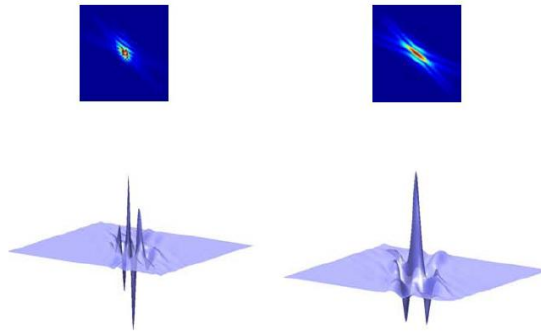

Fig. 1. Directional filters of shearlet.

The properties of shearlets have been already successfully employed in a number of image processing application, including denoising, edge detection and feature extraction (Easley and Labate, 2012 and Easley et al., 2014 and Yi et al., 2009). It is therefore natural to take advantage of the properties of shearlets as part of our face recognition approach.

IV.   SPARSE MULTI-REGULARIZED SHEARLET NETWORK (SMRSN)

In this section, we describe our novel method for face recognition that we call Sparse Multi-Regularized Shearlet Network (SMRSN).

In a nutshell, our SMRSN scheme for FR is defined as a cascade of a feature extraction module, taking advantage of the shearlet transform, followed by a recognition module. The architecture of our algorithm is shown in Fig. 2. The figure shows that an image is fed into a Shearlet Network (SN) and this generates a collection of features associated with a range of locations, scales and orientation; next, a regularization routine is applied to handle to task of recognition. Let us describe each module in detail.
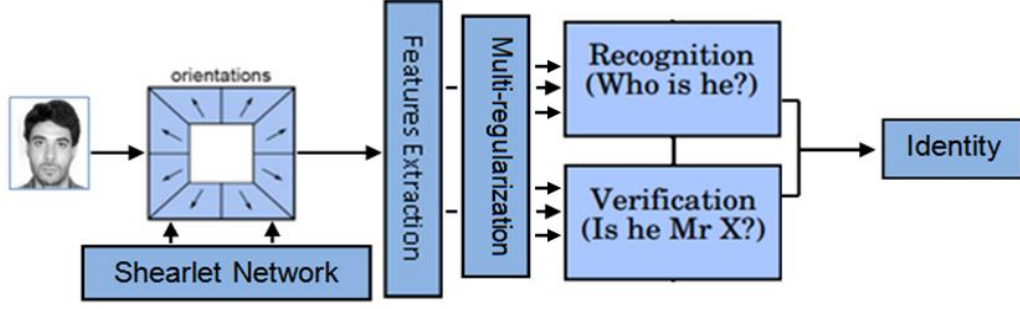
Fig. 2. SMRSN face recognition scheme.

A. Multi-Stage Regularization-based Convex Relaxation approach

We will model the FR problem using the statistical framework called Subset Selection (or L_0 regularization) (Zhang, 2012). This method consists in solving the following constrained optimization problem:

$$\hat{w}_{L_0} = \arg\min_{w \in R^d} \|Xw - y\|_2^2 \text{ subject to } \|w\|_0 \leq \delta \tag{2}$$

where y is an $n \times 1$ matrix representing a normalized test face and X is an $n \times d$ matrix representing a gallery of faces and δ is a tuning parameter. In other words, FR is handled as a multivariable regression problem from sparse data. This is an ill-posed problem and a classical way to solve it is by using a regularization approach where, rather than looking for an exact solution, we look for an approximate one (Bertero, 1986 and Bertero et al., 1988). A very popular and effective regularization method is the so-called Lasso approach (Tibshirani et al., 1996), defined as the minimization:

$$\hat{w}_{L_1} = \arg\min_{w \in R^d} \left[\frac{1}{n} \|Xw - y\|_2^2 + \lambda \|w\|_1 \right] \tag{3}$$

where $\lambda > 0$ is an appropriately chosen regularization parameter.

A refinement of the Lasso approach is obtained by replacing the L1 regularization term with a more general regularization function as follows (Zhang, 2012):

$$\hat{w} = \arg\min_{w} \left[\frac{1}{n} \|Xw - y\|_2^2 + \lambda \sum_{j=1}^{d} g\left(|w_j|\right) \right] \tag{4}$$

where $g\left(|w_j|\right)$ is a general regularization function. In particular, in this paper we choose $g(u) = \min(u, \theta)$ corresponding to the *capped-L_1 regularization,* where θ is a parameter (associated with a thresholding operation). While this approach can be very powerful, it is a non-convex formulation and, thus, it does not guarantee the existence of a global optimal solution (Zhang, 2010). Fortunately, the non-convexity can be addressed by using convex relaxation, so that we can ensure the uniqueness of the solution. In this paper we adopt a method of multi-stage convex relaxation (Zhang, 2010) that is defined as:

5

$$\hat{w}^{(\ell)} = \arg\min_{w \in R^d} \left[\frac{1}{n} \left\| Xw - y \right\|_2^2 + \sum_{j=1}^{d} \lambda_j^{(\ell-1)} \left| w_j \right| \right] \tag{5}$$

where we have a set of parameters $\lambda_j^{(0)} = \lambda$, $\ell = 1, 2, \ldots$ and $j = 1, \ldots, d$

The next section describes the feature extraction step using the shearlet network.

B. SN for Features Extraction

Our SMRSN algorithm is initialized by training a shearlet network (SN) (Borgi et al., 2013) to model the faces. This scheme is shown in Fig. 3 and summarized as follows. The Gallery faces are mapped into a set of biometric signature that we compute using the shearlet transform. This mapping results in a significant dimensionality reduction, since the biometric signatures are very compact thanks to the sparse approximation properties of the shearlet representation. Similarly, the test (Probe) faces are projected on the shearlet network of the Gallery face and new weights specific to this face are produced. Note that the family of shearlets remains unchanged (this is the Gallery face).
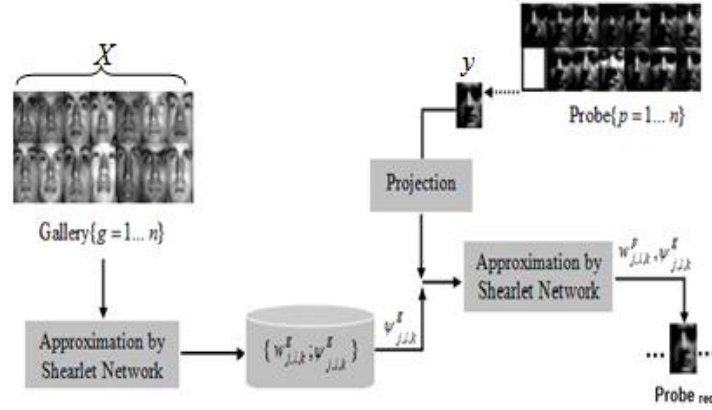

Fig. 3. Overview of SN for features extraction.

Recall that the shearlet basis is a tight frame, meaning that, for any image in the space of square integrable functions, we have a self-dual reproducing formula. In particular, in our setting, we can represent or synthetize any reconstructed face $\tilde{f}$ according to the formula

$$\tilde{f} = \sum_{j,l \in \mathbb{Z}, k \in \mathbb{Z}^2} < f, \psi_{j,l,k} > \psi_{j,l,k} = \sum_{j,l,k}^{i} w_i \psi_{j,l,k} \tag{6}$$

where f is the unknown face image, $\psi_{j,l,k}$ are the shearlet basis functions and w_i are appropriate weights.

Similar to the wavelet network (Said et al., 2009), the Shearlet Network is a combination of a RBF neural network and the shearlet decomposition. During the optimization stage, each shearlet coefficient from the library is processed through the hidden layer of the network. The calculation of the weights connection in every stage is obtained by projecting the signal to be analyzed into the shearlet basis. Since shearlets form a tight frame, we don't need to use the dual functions. Figure 4 shows the SN architecture.
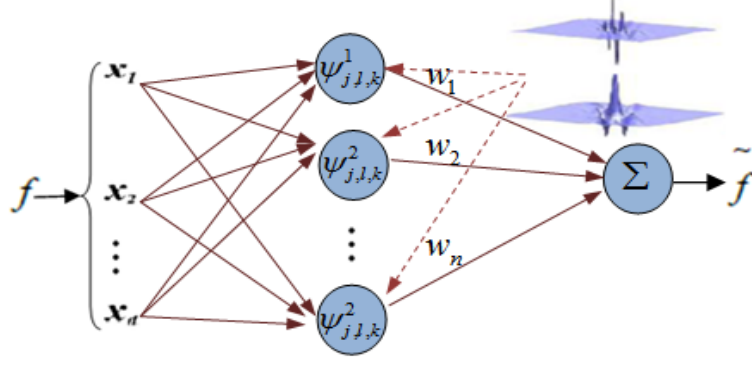
Fig. 4. Overview of SN architecture.

In our approach, the mother shearlet used to construct the tight frame of shearlets $\left\{\psi_{j,l,k}\right\}$ is the second derived of the Beta function (Ben Amar et al., 2005) that is defined as:

$$\beta(x; p, q, x_0, x_1) = \begin{cases} \left(\dfrac{x - x_0}{x_c - x_0}\right)^p \left(\dfrac{x_1 - x}{x_1 - x_c}\right)^q & if \quad x \in]x_0, x_1[\\ 0 & otherwise. \end{cases} \qquad (7)$$

With $\quad p, q, x_0 < x_1 \in \Re \quad$ and $\quad x_c = \dfrac{px_1 + qx_0}{p + q}.$

We have shown (Ben Amar et al., 2005) that all derivatives of beta function are admissible wavelets. We can easily generate different mother shearlets by simply modifying beta function parameters (x_0, x_1, q, p). Below we summarize the algorithm of SN training:

Algorithm 1: SN learning

Input: image f

Output: reconstructed image f_{rec}

1. Select a shearlet $\left\{\psi_{j,l,k}\right\}$ as activation function of the shearlet network:

- Choose the mother shearlet.
- Build a library of shearlets forming a tight frame.
- Set as a stop-learning condition (number of shearlets) and iterate the following steps:

2. Calculate the weights by direct projection of the image on the shearlet as $w_i = < f, \psi_{j,l,k} >$.

3. Calculate the output of the network f_{rec}.

4. If the number of shearlets is reached, the learning stops; otherwise another shearlet is selected and we return to **2.**

C. SMRSN Algorithm

As indicated in Sec. III.A., the problem stated by the formula (2) is non-convex but the problem can be relaxed by applying the multi-stage convex optimization (5). The explicit method that we use in this paper to find the optimal solution of the recognition step is organized as follows.

First, at every stage, we compute:

$$W_x = w_{init} \times X$$
$$W_y = w_{init} \times y$$

then we compute W_s as follows: (8)

$$W_s = W_x \times \lambda^{(l)} \times \sum \min\left(\sum\left(|W_y|, 2\right), \theta\right) \backslash W_y$$

And the weight solution is:

$$w = W_x' \times W_s \qquad\qquad (9)$$

Where, according to the notation above, y is a normalized test face, X is a matrix representing a gallery of faces and $\lambda^{(l)}$ is a regularization. The operator $\backslash$ is the matrix left division used here to solve the problem as $W_s X = y$.

The initial value of the weight w_{init} is chosen using the logistic function that is a classical and well established model in statistical model to minimize the residual error (Hastie et al., 2003). This is defined as

$$f(x) = \frac{c}{1 + a\exp^{-bx}} \qquad\qquad (10)$$

where a, b, c are positive parameters. The plot is shown in the figure below:
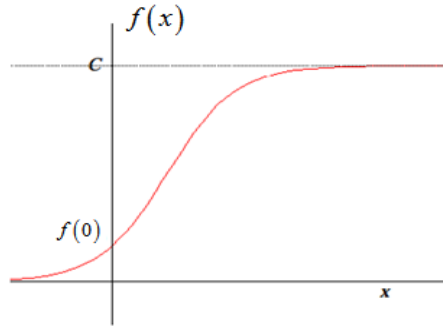

Fig. 5.   Graph of logistic function.

According to this models, the best choice of w_{init} is:

$$w_{init} = 1/(1 + 1/\exp(-\mu e_{init}^2 + \mu\delta)) \qquad\qquad (11)$$

where μ and δ are positive scalars to be determined (see subsection D.) and e_{init} is the initial residual given by:

$$e_{init} = (y - mean(X))^2 \qquad\qquad (12)$$

Recall that X, in our setting, are the aligned gallery faces (a $n \times d$ matrix) and y a normalized test face (a $n \times 1$ matrix). Note that, after optimization, we update the residual e and then the weight w_i.

Regarding the choice of the parameter $\lambda_j^{(\ell)}$, we apply the formula (Zhang, 2010):

$$\lambda_j^{(\ell)} = \lambda I\left(\left|w_j^{(\ell)}\right| \le \theta\right) \tag{13}$$

where $j = 1, ..., d$ and

$$\lambda = \tau\sigma\sqrt{\ln(d)/n} \;\; ; \;\; \theta = \mu\lambda \tag{14}$$

with $\tau = 1, 2, 4, 8, 16, ...$ and $\mu = 0.5, 1, 2, ...$ (we choose $\sigma = 1e-4$)

Below we summarize our SMRSN algorithm. As above, where X represents the reconstructed gallery faces after extraction of the features by training SN, y is the reconstruct test face with the features extracted after projection of the real test face on the tight frame of shearlets produced by the gallery faces.

Algorithm 2: SMRSN

Input: - y : normalized test face f : $y = f / norm(f, 2)$

      - X : aligned gallery faces: $X = X / \sqrt{\sum X * X}$

      - $Iter$: max of iteration

Output: w ; $Identity(y)$

     1. Compute the residual e_{init} , refer to formula (12)

     2. Compute the parameters μ and δ , refer to formulas (16) and (17)

     3. Compute w_{init} , refer to formula (11)

     4. Initialize $\lambda_j^{(0)} = \lambda$ and $\theta = \mu\lambda_j^{(0)}$, refer to formula (14)

     5. For $\ell = 1, 2, ...$

         For $j = 1, ..., Iter$

           - Compute the SMRSN using formulas (8):

$$W_s = SMRSN(X, y, w_{init}, \lambda_j^{(\ell-1)}, \theta)$$

           - Compute w_i using the formula (9)

           - Update the residual e : $e = (Xw_i - y)^2$

           - $e_{init} = e$ (used for w_{init})

           - Update μ and δ

           - Update w_{init} , refer to formula (11)

         End

           - Update $\lambda_j^{(\ell)}$ and θ , refer to the formula (13) and (14)

           - $w_{init} = w_i$

         End

$$- \quad y_{rec} = X * w_i$$

$$- \quad w = w_i$$

6. For $k = 1, \ldots,$ Classnum

$$\left| \quad error(k) = \left\| w^{1/2}(y - X_k w_k) \right\|_2^2 \right.$$

End

7. $Identity(y) = \arg\min(error)$

In the description above, Classnum denotes the number of classes of X. If $Classnum = d$ then we have the case of single sample per person (SSPP). In our experiments we have chosen to iterate ℓ three times; the value of $Iter$ can be chosen equal to ℓ.

The condition $w_{init} = w_i$ is one of the novelties of the SMRSN algorithm, with respect to common MTFL algorithms. Our approach uses a classification strategy that is similar to RRC (Yang et al., 2013) and SRC (Wright et al., 2009) but has some fundamental differences. Note step 7 of our algorithm:

$$Identity(y) = \arg\min(error) \tag{18}$$

Where

$$error(k) = \left\| w^{1/2}(y - X_k w_k) \right\|_2^2 \tag{19}$$

and X_k is the training samples of a class k.

A crucial difference between our SMRSN method and RRC (or SRC) is the optimization stage where we use a multi-task regularization step. That is, we update the residual e at *every stage* and we update $\lambda_j^{(\ell)}$ and θ once we update μ, δ and w_{init}. This way we share the previous value w_i as an input weight to the next task l.

D. Selection of μ and δ parameters

As we can see in Fig. 5, the logistic curve has a single inflection point which separates the curve into two equal regions of opposite concavity. The inflection point has coordinates $\left(\dfrac{\ln a}{b}, \dfrac{c}{2} \right)$:
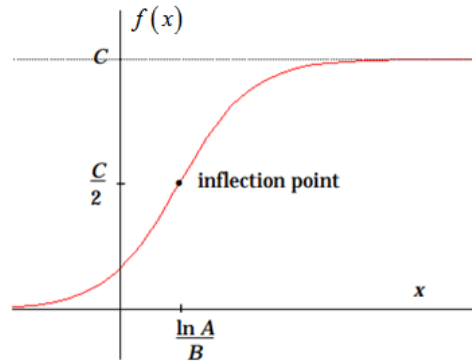

Fig. 6. Graph of logistic function with inflection point.

Our choice for the weight is $w_{init} = 1/(1+1/\exp(-\mu e_{init}^2 + \mu\delta))$. If we set $x = e_{init}^2$ then

$$
\begin{aligned}
w_{init}(x) &= \frac{1}{(1+1/\exp(-\mu x + \mu\delta))} \\
&= \frac{1}{\left(1+\exp\left(\mu x - \mu\delta\right)\right)} \\
&= \frac{1}{\left(1+\exp\left(-\mu\delta\right)\exp\left(\mu x\right)\right)}
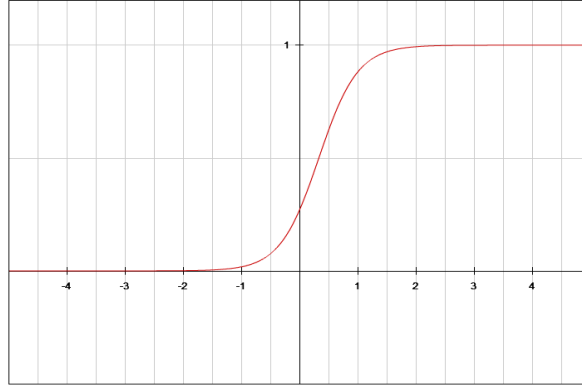\end{aligned}
\tag{15}
$$

Fig. 7. Example of a graph: the logistic function $1/\left(1+\exp\left(-3x+1\right)\right)$.

By comparing with the formula (10) we derive that $c = 1$, $a = \exp\left(-\mu\delta\right)$ and $b = -\mu$.

The symmetry of the curve about this point must occur halfway up the curve at height $w_{init} = \dfrac{1}{2}$. The coordinates of the inflection point will be $\left(\delta, \dfrac{1}{2}\right)$.

In our experiments we choose μ as:

$$
\mu = \frac{0.6}{\delta}
\tag{16}
$$

As in (Yang et al., 2013), to make the model robust to occlusions, we set δ as:

$$
\delta = \varphi\left(e_{init}\right)_l
\tag{17}
$$

where $\varphi\left(e_{init}\right)_k$ is the kth largest element of the set $\left\{e_{init}^2\left(j\right), j = 1, 2, ..., n\right\}$.

Note that δ is the parameter of demarcation point and μ is the parameter which controls the decreasing rate of function w_{init}.

E. Why the choice of the logistic function?

Our goal consists formally in classifying a test sample y (column vector) into one of the classes defined by the training samples X (matrix). Thus the classification problem can be solved by estimating a regression weight

W with $WX = y$, where the diagonal values $W_{i,i}$ contain the weights assigned to pixel i of test sample y. Note that we do not treat all pixels equally. If, for example, there is an occluded test sample y, then we will try to set a small value of the $W_{i,i}$ for the corresponding outlier (occluded) pixel, while we set a big value if there is an inlier (not occluded) pixel. This way we can control their effect in the representation of y in term of the training samples. In other words, if the model WX is a good representation of X then the occluded pixels will have a big coding residual e and the pixel with a big residual e should have a small weight $W_{i,i}$ (Yang et al., 2013). This implies that W must be inversely proportional to the residual e and, as shown in (Hastie et al., 2003), this leads naturally to the choice of the logistic function.

V.   EXPERIMENTAL RESULTS

We have run extensive numerical tests to validate and illustrate the performance of our algorithm. For our numerical tests, we have used several face databases. We have used the ORL, AR (Martinez, 1998), GT (Georgia Tech Face Database), FEI (Thomaz and Giraldi, 2010) databases in controlled environments; the LFW database (Huang et al., 2007) in uncontrolled environments; the Lab2 (Xu et al., 2011) database was used for different illumination environments; the FERET (Phillips et al., 2000), FRGC v1 (Phillips et al., 2005), FEI databases were be used to test the experiments of single-sample-per-person (SSPP), one of the most challenging problem in FR.

We have compared our results against a variety of standard FR methods, including SVM and NN (nearest neighbors) methods, and state-of-the-art FR methods, including BHDT (Cevikalp, 2010), MetaFace (Yang et al., 2010), RKR (Yang et al., 2012), RRC (Yang et al., 2013) and CRC (Zhang et al., 2011) used for comparison. For the comparisons, we have used the codes provided by the authors. Regarding CRC (Zhang et al., 2011) and MetaFace (Yang et al., 2010), there is a dimensionality reduction step to set. In this case, we have determined the parameters that give the highest recognition accuracy.

A.   Controlled Environment: AR, ORL, GT and FEI Databases

As a first group of experiments, we have tested FR in controlled environment.

The AR dataset contains 50 males and 50 females where faces contain only illumination and expression changes (Martinez, 1998). For each subject, we have randomly selected seven images from Session 1 for training, and other seven images from Session 2 for testing. The face images were resized to 32×27.

The ORL database contains 10 different images of each of 40 distinct subjects (400 images). For some subjects, the images were taken at different times, varying the lighting, facial expressions and facial details (glasses/no glasses). For each subject, we have randomly selected five images for training and other five images for testing. The face images were resized to 32×27. Some samples from the AR and ORL databases are shown in Fig. 8.
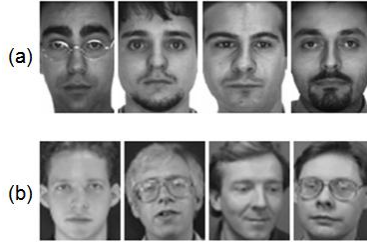
Fig. 8.   Samples from (a) AR database, (b) ORL database.

The recognition accuracy on the AR and ORL database is reported in Table I. Our method (SMRSN) has the best score, showing a significant improvement in FR rate compared with the other methods considered.

TABLE I.　　RECOGNITION ACCURACY ON THE AR & ORL DATABASE.

Method	AR	ORL
NN	0.7010	-
SVM	0.8729	0.8700
BHDT	0.5714	0.8000
MetaFace	0.8814	0.8350
RKR	0.9329	0.8100
RRC	0.9257	0.8850
CRC	0.9071	0.9000
SMRSN	**0.9500**	**0.9250**

The Georgia Tech (GT) Face Database contains 750 color images of 50 subjects (15 images per subject), as shown in Fig. 9. These images have large variations in pose and expression and some illumination changes. Images were converted to gray scale and cropped and resized to 27×18 for the first test and 32×27 for the second one. The first eight images of all subjects were used in the training (400 images) and the remaining seven images for testing (350 images). The testing results are reported in Table II.

Fig. 9.   One subject from GT database.

Also in this case, on images with large variations in pose and expression and some illumination changes, Table II shows our SMRSN approach outperforms all other methods considered.

TABLE II.　　RECOGNITION ACCURACY ON THE GT DATABASE.

Method	27×18	32×27
NN	-	-
SVM	0.5886	0.5714
BHDT	0.4314	0.4314
MetaFace	0.5800	0.6343
RKR	0.5171	0.5000
RRC	0.6400	0.6714
CRC	0.6685	0.6743
SMRSN	**0.7000**	**0.6829**

The FEI face database is a Brazilian face database that contains a set of face images taken between June 2005 and March 2006 at the Artificial Intelligence Laboratory of FEI in Brazil. There are 14 images for each of 200

individuals, for a total of 2800 images. All images are in color and taken against a white homogenous background in an upright frontal position with profile rotation of up to about 180 degrees. Scale might vary about 10% and the original size of each image is 640x480 pixels. All faces are represented by students and staff at FEI, between 19 and 40 years old, with distinct appearance, hairstyle, and adorns. The number of male and female subjects are exactly the same and equal to 100. Figure 6 shows some examples of image variations from the FEI face database of one subject. In our test, we have used the first nine images of all subjects for training (1800 images) and the remaining five images for testing (1000 images). The images were cropped and resized to 27×18 for the first experiment and resized to 32×27 for the second.

Fig. 10. One subject from FEI database.

Using this protocol with FEI database, the data reported in Table III show that SMRSN gives the second best performance with size 27×18 after Metaface (Yang et al., 2010), but achieves the best performance with 32×27 size.

TABLE III.   RECOGNITION ACCURACY ON THE FEI DATABASE.

Method	27×18	32×27
NN	-	-
SVM	0.5880	0.6050
BHDT	0.4110	0.4440
MetaFace	**0.7090**	**0.7160**
RKR	0.5640	0.6240
RRC	0.4610	0.4890
CRC	0.6300	0.7040
SMRSN	**0.6430**	0.6930

B. Lab2 Database

As a second group of experiments, we have tested FR under different illumination conditions.

The Lab2 database (Xu et al., 2011) contains visible light images and near-infrared images of 50 subjects. Each subject provides twenty visible light face images (1000 images) and the same number of near-infrared face images. These images were acquired under four different illumination conditions (4 sessions). The face images also have variation in facial expression and pose. We have randomly selected 5~15 samples from the first three sessions for training and 5 samples from the fourth session for test. The face images were resized to 32×27.
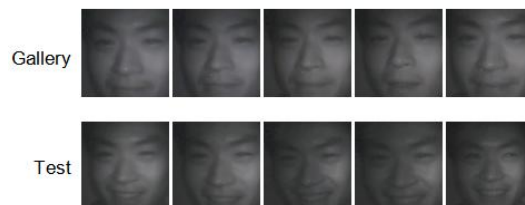

Fig. 11. One subject from near-infrared Lab2 database.

The recognition accuracy of near-infrared faces is reported in Table IV. Our SMRSN method shows superior performance with respect to all the other methods considered using 5, 10 and 15 images.

Method	5	10	15
NN	-	-	-
SVM	0.6880	0.7880	0.8480
BHDT	0.5880	0.7360	0.8200
MetaFace	0.7320	0.7920	0.7600
RKR	0.7200	0.8000	0.8640
RRC	0.7360	0.8320	0.8440
CRC	0.6680	0.8040	0.8480
SMRSN	**0.7400**	**0.8360**	**0.8880**

Fig. 12. One subject from visible light Lab2 database.

Similarly in Table V, showing the recognition accuracy of visible light faces, we can see SMRSN achieves the highest recognition rate on all three tests (5, 10, 15 images) and outperforms the other methods considered.

TABLE V.   RECOGNITION ACCURACY ON VISIBLE LIGHT IMAGES FROM THE LAB 2 DATABASE.

Method	5	10	15
NN	-	-	-
SVM	0.2440	0.4480	0.6560
BHDT	0.1560	0.3320	0.5920
MetaFace	0.3000	0.4280	0.5560
RKR	0.2680	0.4200	0.6040
RRC	0.2160	0.4360	0.6560
CRC	0.3000	0.4720	0.6880
SMRSN	**0.3320**	**0.4800**	**0.7160**

C. Uncontrolled Environment: LFW Database

As a third group of experiments, we have tested FR in uncontrolled environment.

The LFW database (Huang et al., 2007) contains images of 5,749 different individuals in uncontrolled environment (see samples in Fig. 13). LFW-a is a version of LFW after alignment using commercial face alignment software (Wolf et al., 2010). We obtain a dataset with 158 subjects from LFW-a. For each subject, we have randomly selected 2~5 samples for training and another 2 samples for test. The images were firstly cropped to 121×121 and then resized to 32×32 (Zhu et al., 2012).

Fig. 13. Four subjects from LFW database.

The FR rates on the LFW dataset are reported in Table VI. Also in this case, the result shows that SMRSN outperforms the other methods for all tests (2, 3, 4 and 5 images); there is only one instance where RKR achieves the same performance as SMRSN.

TABLE VI.   RECOGNITION ACCURACY ON THE LFW DATABASE.

Method	2	3	4	5
NN	0,1100	0.1320	0.1470	0.1620
SVM	0.2152	0.2468	0.3038	0.3544
BHDT	0.0791	0.1203	0.1361	0.1772
MetaFace	0.1582	0.2152	0.2405	0.2563
RKR	**0.3038**	0.3607	0.4113	0.4525
RRC	0.2690	0.3449	0.3956	0.4462
CRC	0.1899	0.2595	0.3322	0.3607
SMRSN	**0.3038**	**0.3734**	**0.4146**	**0.4684**

D. Single-Subject- Per-Person: FERET, FRGC v1and FEI Databases

In this fourth group of experiments, in addition to computing of the recognition accuracy, we also plot the Receiver Operator Characteristic (ROC) curves for our SMRSN side-by-side with other state-of-the-art methods. Here we use the FERET, FRGC v1 and FEI databases.

The FERET dataset contains a large number of subjects (single image per subject) in the gallery and probe sets with differences in illumination, facial expression variations and aging effects (Phillips et al., 2000). The frontal faces in the FERET database are divided into five sets, including the subsets *fa*, consisting of 1196 images, used as gallery set containing one image per person and *fb*, consisting of 1195 images, taken with different expressions. For our SSPP experiments, we have randomly selected 100~200 images from fa for training and randomly selected 100~200 images from fb for testing. The images were resized to 32×27. The FR rates on the FERET dataset are reported in Table VII and the ROC curves are plotted in Fig. 14. The SMRSN achieves the best recognition accuracy compared to the others methods.

TABLE VII.   RECOGNITION ACCURACY ON THE FERET DATABASE.

Method	100	150	200
SVM	0.7700	0.7333	0.7150
BHDT	0.5000	0.4200	0.3350
MetaFace	0.8900	0.8933	0.8950
RKR	0.8900	0.8533	0.8500
RRC	0.9400	0.9133	0.9150
CRC	0.9000	0.8667	0.9000
SMRSN	**0.9600**	**0.9400**	**0.9600**

(a)   (b)   (c)
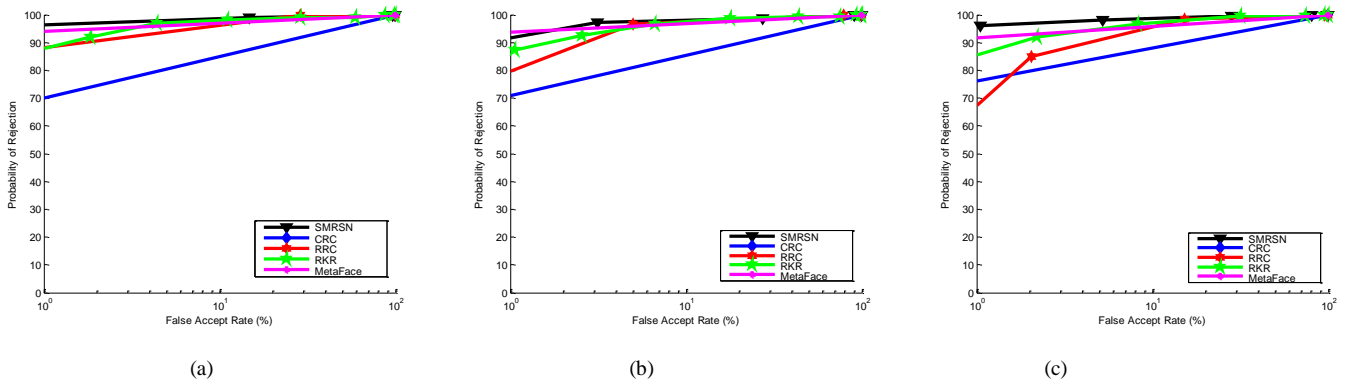
Fig. 14. ROC curves using FERET database with. (a) 100 images. (b) 150 images. (c) 200 images.

The FRGC v1 database contains faces acquired under uncontrolled conditions (PHILLIPS ET AL., 2005). Experiment 1 contains a single controlled gallery image and a probe with one controlled still image per subject

(183 training images, 152 gallery images, and 608 probe images). Experiment 2 considers the identification of a person given a gallery with four controlled still images per subject (732 training images, 608 gallery images, and 2432 probe images). Finally, experiment 3 considers a gallery with one controlled still image per subject and multiple uncontrolled probe images per subject (366 training images, 152 gallery images, and 608 probe images) (SCHWARTZ ET AL., 2012). We randomly selected 152 images for training and 152 images for test. The images were cropped and resized to 27×18 for the first experiment and resized to 32×27 for the second. The recognition accuracy on the FRGC v1 database with two image sizes is reported in Table VIII and the ROC curves are shown in Fig. 15. Also in this case, SMRSN shows superior performance with respect to all the other methods considered.

TABLE VIII.   RECOGNITION ACCURACY ON THE FRGC v1 DATABASE.

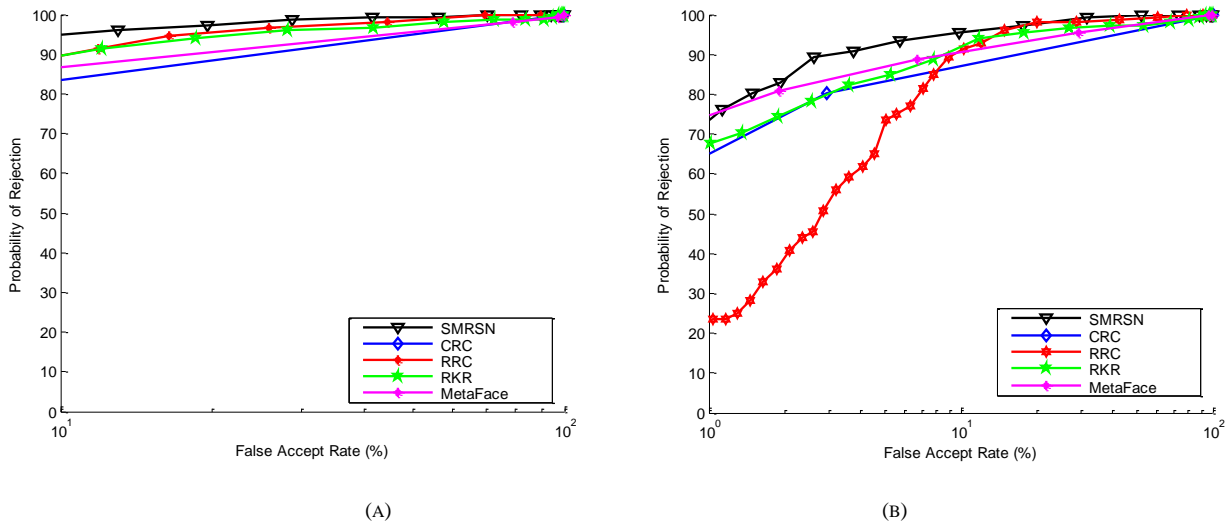| Method	27×18	32×27
NN	-	-
SVM	0.6053	0.6974
BHDT	0.2697	0.2829
MetaFace	0.6842	0.7171
RKR	0.6316	0.6316
RRC	0.7500	0.7697
CRC	0.6513	0.7434
SMRSN	**0.7763**	**0.8158**

Fig. 15. ROC curves using FRGC v1 database with. (a) size 27×18. (b) size 32×27.

From the FEI face database, we selected the frontal image number twelve with facial expression for training (200 images) and the image number fourteen with illumination (200 images) for testing in order to assess the impact of illumination and expression changes on the recognition. The images were cropped and resized to 27×18 for the first experiment and resized to 32×27 for the second. Also in this case, the data reported in Table IX show that SMRSN achieves the best performance. The corresponsing ROC curves are show in Fig. 16.

TABLE IX.   RECOGNITION ACCURACY ON THE FEI DATABASE.

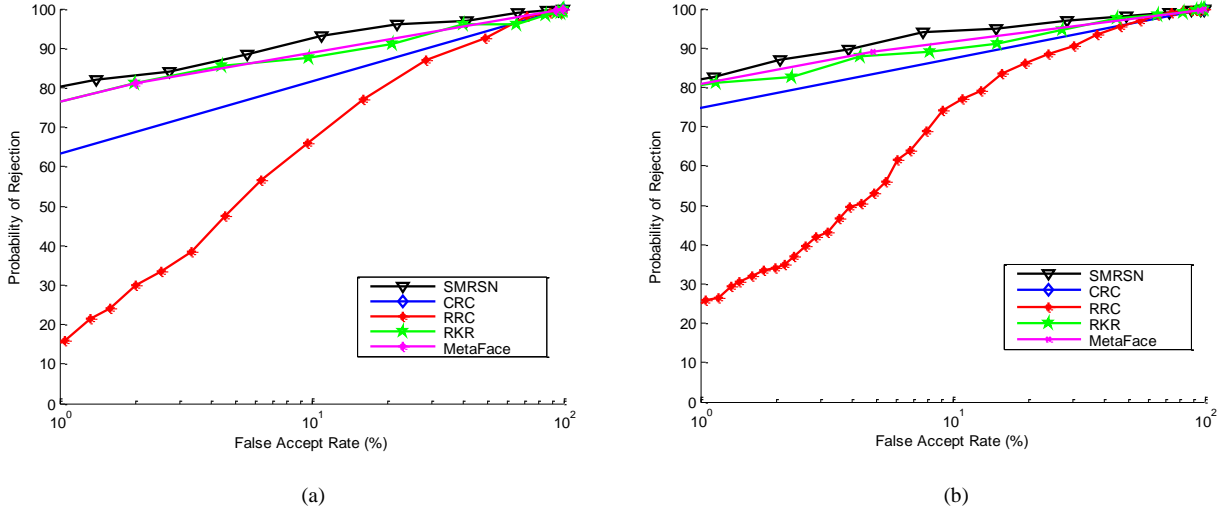| Method	27×18	32×27
NN	-	-
SVM	0.2200	0.2350
BHDT	0.1100	0.1400
MetaFace	0.6900	0.7350
RKR	0.6900	0.7600
RRC	0.4600	0.5450
CRC	0.6650	0.7200
SMRSN	**0.7350**	**0.7650**

Fig. 16. ROC curves using FEI database with. (a) size 27×18. (b) size 32×27.

E. Runing Time

We have compared the average running time of all methods we have considered using SSPP based FR experiments. For all our numerical tests, we have used Matlab version 7.0.1 with Intel core 2 duo 2.10 GHz CPU and with 2.87G RAM. Note that, in practical applications, training is usually an offline stage while recognition is usually an online step. Thus, we decided to only report the values of the recognition time in Table X. The table shows that the algorithms RKR (Yang et al., 2012) and CRC (Zhang et al., 2011) are the fastest; the running time of our SMRSN algorithm is comparable with the SVM, RRC (Yang et al., 2013) and MetaFace (Yang et al., 2010) algorithms.

TABLE X.   THE AVERGE RUNNING TIME (SECONDS).

Database	FERET			FRGC v1		FEI	
	size 32×27						
	100	150	200	size 27×18	size 32×27	size 27×18	size 32×27
NN	-	-	-	-	-	-	-
SVM	0.1522	0.6776	2.0481	0.7458	0.6957	2.0417	2.0937
BHDT	0.0047	0.0122	0.0262	0.0125	0.0116	0.0219	0.0255
MetaFace	0.4291	0.6921	0.8607	0.8753	0.8402	1.0500	0.9939
RKR	1.60e-004	3.13e-004	3.10e-004	3.09e-004	3.09e-004	3.15e-004	3.10e-004
RRC	0.6167	0.6561	0.6606	0.1600	0.6683	0.1727	0.6710
CRC	0.0028	0.0026	0.0095	0.0026	0.0026	0.0093	0.0037
SMRSN	0.0384	0.0354	0.0420	0.0361	0.0430	0.1512	0.0539

VI. CONCLUSION AND FUTURE WORK

This paper presents a novel for robust face recognition method called Sparse Multi-Regularized Shearlet Network (SMRSN). Our approach uses two innovative ideas to improve the recognition rate and execution time of face recognition system even under uncontrolled conditions. First, we take advantage of the sparse representation properties of shearlets through a Shearlet Network (SN). SN is applied to extract the geometric features from an image in order to encode the facial geometry. The advantage of the shearlet approach, with respect to more traditional ideas (e.g., PCA, wavelets), lies in the highly sparse representation properties of shearlets that have a unique ability to sparsely encode the essential geometric content of the image. Second, we use a refined version of the multi-task learning framework to exploit the relationship among the multiple shared stages generated by changing the regularization parameter. This module uses regularization theory to control the trade-off between the fidelity to the data (gallery faces) and the smoothness of the solution (probe or test faces). At every stage of our optimization stage we use a new value of regularization parameter and the weight generated by the previous stage to share the features. We have conducted extensive numerical tests, under controlled and uncontrolled conditions, as well as in the single-sample-per-person setting, and we have compared our method against a number of standard and state-of-the-art-method. The numerical results show that our method provides outstanding recognition rates, consistently outperforming all the competing methods we have considered. We attribute the improved performance of our method to: (i) the properties of our SN stage, since it produces features that are robust with respect to moderate distortions of the image, ensuring robust recognition even with (moderate) changes in pose and illumination of the face; (ii) the application of a refined multi-task learning strategy that allow us to exploit the relationships among multiple shared stages of our recognition algorithm in such a way to obtain more robust recognition. Our innovative approach provides a significant contribution towards the development of a new generation of face recognition systems for applications in areas including law enforcement and surveillance systems where images occur with variations in pose, illumination and facial expression.

Based on the findings of this paper, we can identify a number of promising extensions and improvements of our work that will further impact the development of improved face recognition systems. Several improvements can be achieved by incorporating stronger geometric *invariances* into our feature extraction stage. In particular, one drawback of our current approach is that it may not perform well if images are rotated. This limitation can be addressed by modifying the SN framework in such a way to extract rotation-invariant features. Invariances with respect to other types of distortions can also be useful to obtain robustness to ageing effects. In this case, we expect that it will be possible to take advantage of the properties of the shearlet representation to create features that are robust with respect to diffeomorphic distortions, e.g., those smooth distortions that are frequently associated with ageing effect. Concerning the wider applicability of our method, we expect that the ideas developed in this paper have potential applicability to other problems, most notably the recognition of facial expressions. The hallmark of every facial expression system is accuracy and current systems are hampered by the same limitation of FR systems that have motivated our approach. Thanks to the ability to extract and process the essential facial landmarks, we expect that a combination of SN and an appropriate version of multi-task learning, similar to the one we have developed in this paper, will perform very competitively for facial expression recognition. This will be part of our future work.

Acknowledgment

REFERENCES

Ben Amar et al., 2005 C. Ben Amar, M. Zaied and M.A. Alimi. (2005). Beta Wavelets. Synthesis and application to lossy image compression. *Advances in Engineering Software*, 36, 459-474.

Belhumeur et al., 1997 P.N. Belhumeur, J.P. Hespanha, and D.J. Kriengman. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19, 711–720.

Bertero et al., 1988 Bertero, M., Poggio, T., & Torre, V. (1988). Ill-posed problems in early vision. *Proc. IEEE*, 76, 869–889.

Bertero, 1986 Bertero, M. (1986). Regularization methods for linear inverse problems. *Inverse Problems*, ed. C.G. Talenti, Springer, Berlin.

Borgi et al., 2013 M.A. Borgi, D. Labate, M. El'arbi and C. Ben Amar. (2013). Shearlet Network-based Sparse Coding Augmented by Facial Texture Features for Face Recognition. In: *International Conference on Image Analysis and Processing*, (pp. 611-620)

Breiman and Friedman, 1998 L. Breiman and J.H Friedman. (1998). Predicting Multivariate Responses in Multiple Linear Regression. *Royal Statistical Society Series B*.

Brown and Zidek, 1980 P.J. Brown and J.V. Zidek. (1980). Adaptive Multivariate Ridge Regression. *The Annals of Statistics*, 8, 64–74.

Caruana, 1997 R. Caruana, R. (1997). Multi-task learning. *Machine Learning*, 28, 41–75.

Cevikalp, 2010 Cevikalp H. (2010). New clustering algorithms for the support vector machine based hierarchical classification. *Pattern Recognition Letters*, 31, 1285-1291.

Daubechies et al., 2008 I. Daubechies, R. Devore, M. Fornasier and C.S. Gunturk (2008). Iteratively Re-Weighted least squares minimization for Sparse Recovery. ArXiv: 0807-0575.

Destrero et al., 2009 A. Destrero, C. De Mol, F. Odone and A. Verri. (2009). A Regularized Framework for Feature Selection in Face Detection and Authentication. *International Journal of Computer Vision*, 83, 164-177.

Donoho et al., 1998 D. L. Donoho, M. Vetterli, R. A. DeVore, & Daubechies, I. (1998). Data compression and harmonic analysis. *IEEE Trans. Inform. Th.* 44, 2435–2476.

Easley and Labate, 2012 Easley. G. R., & Labate, D. (2012.. Critically sampled wavelets with composite dilations. *IEEE Trans. Image Process*., 21, 550-561.

Easley et al., 2014 Easley. G. R., Labate, D., & Patel, V. (2014). Directional multiscale processing of images using wavelets with composite dilations. *Journal of Mathematical Imaging and Vision*, 48, 13-34.

Georgia Tech Face Database. http://www.anefian.com/face_reco.htm, 2007.

Gong et al., 2013 Gong, P., Ye, J., & Zhang, C. (2013). Multi-Stage Multi-Task Feature Learning. *Journal of Machine Learning Research*, 14, 2979-3010.

Guo and Labate, 2007 Guo, K. & Labate, D. (2007). Optimally sparse multidimensional representation using shearlets. *SIAM J Math. Anal*., 39 , 298-318.

Hastie et al., 2003 T. Hastie, R. Tibshirani and J. Friedman (2003). The Elements of Statistical Learning. Springer Series in Statistics.

Heisele et al., 2001 B. Heisele, P. Ho, and T. Poggio. (2001). Face recognition with support vector machine: Global versus component-based approach. In: *IEEE International Conference on Computer Vision*, (pp. 688-694)

Huang et al., 2007 G.B. Huang, M. Ramesh, T. Berg and E. Learned-Miller. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *Technical Report, University of Massachusetts*, Amherst, 7-49

Jiang et al., 2008 X.D. Jiang, B. Mandal and A. Kot. (2008). Eigenfeature Regularization and Extraction in Face Recognition *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, 383-394.

Labate et al., 2005 D. Labate, W.-Q. Lim, G. Kutyniok and G. Weiss. (2005). Sparse multidimensional representation using shearlets. In: *Wavelets XI (San Diego, CA), SPIE, Bellingham, WA,* (pp. 254–262)

Lu et al., 2003 J. Lu, K.N. Plataniotis and A.N. Venetsanopoulos. (2003). Regularized Discriminant Analysis For the Small Sample Size Problem in Face Recognition. *Pattern Recognition Letters*, 24, 3079-3087.

Martinez, 1998 Martinez, A. (1998). The ar face database. *CVC Technical Report 24*.

Monson, 1996 Hayes Monson H. (1996). Statistical Digital Signal Processing and Modeling. Wiley.

Phillips et al., 2000 P.J. Phillips, H. Moon, S.A. Rizvi and P.J. Rauss. (2000). The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 1090-1104.

Phillips et al., 2005 P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek. (2005). Overview of the face recognition grand challenge. In: *IEEE International Conference in Computer Vision and Pattern Recognition*, (pp. 947-954)

Said et al., 2009 S. Said, B. Ben Amor, M. Zaied, C. Ben Amar and M. Daoudi. (2009). Fast and efficient 3D face recognition using wavelet networks. In: *IEEE International Conference on Image Processing*, (pp. 4153-4156)

Schwartz et al., 2012 W. R. Schwartz, H. Guo, J. Choi and L. S. Davis. (2012). Face Identification Using Large Feature Sets IEEE Transactions on Image Processing, 21, 2245-2255.

Thomaz and Giraldi, 2010 C. E. Thomaz and G. A. Giraldi. (2010). A new ranking method for Principal Components Analysis and its application to face image analysis. *Image and Vision Computing*, 28, 902-913.

Thrun and Pratt, 1997 Thrun, S., & Pratt, L. (1997). Learning to Learn. Kluwer Academic Publishers.

Tibshirani et al., 1996 Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Statistical Methodology),* 58, 267–288.

Turk and Pentland, 1991 Turk, T., & Pentland, A. (1991). Eigenfaces for recognition. *J. Cognitive Neuroscience,* 3, 71–86.

Wang et al., 2009 X. Wang, C. Zhang and Z. Zhang. (2009). Boosted Multi-Task Learning for Face Verification With Applications to Web Image and Video Search.In: *IEEE International Conference in Computer Vision and Pattern Recognition*, (pp. 142-149)

Wright et al., 2009 Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3, 210–227.

Wolf et al., 2010 Wolf, L., Hassner, T., & Taigman. Y., (2009). Similarity scores based on background samples. In: *Asian Conference in Computer Vision*, (pp. 88–97)

Xu et al., 2011 Xu, Y., Zhong, A., Yang, J., & Zhang, D. (2011). Bimodal biometrics based on a representation and recognition approach. *Opt. Eng.,* 50.

Yang et al., 2010 Yang, M., Zhang, L., & Zhang, D. (2010). Metaface learning for sparse representation based face recognition. In: *IEEE International Conference on Image Processing*, (pp. 1601-1604).

Yang et al., 2012 Yang, M., Zhang, L., Zhang, D., & Wang, S. (2012). Relaxed collaborative Representation for Pattern Classification. In: *IEEE International Conference in Computer Vision and Pattern Recognition*, (pp. 2224-2231).

Yang et al., 2011 Yang, M., Zhang, L., Yang, J., & Zhang, D. (2011). Robust sparse coding for face recognition. In: *IEEE International Conference of Computer Vision and Pattern Recognition* (pp. 625-632).

Yang et al., 2013 Yang, M., Zhang, L., Yang, J., & Zhang, D. (2013). Regularized Robust Coding for Face Recognition. *IEEE Trans. on Image Processing*, 22, 1753-1766.

Yi et al., 2009 SYi, S., D. Labate, D., Easley. G. R., & Krim, H. (2009). A Shearlet approach to Edge Analysis and Detection. *IEEE Trans. Image Process*, 18, 929–941.

Zhang, 2010 Zhang, T. (2010). Analysis of multi-stage convex relaxation for sparse regularization. *Journal of Machine Learning Research*, 11 (2010), 1081–1107.

Zhang et al., 2011 L. Zhang, M. Yang, M., &, Feng, X. (2011). Sparse Representation or Collaborative Representation: Which Helps Face Recognition? In: *IEEE International Conference in Computer Vision*, (pp. 471-478).

Zhang, 2012 Zhang, T. (2012). Multi-stage convex relaxation for feature selection. *Bernoulli*,

Zhu et al., 2012 Pengfei Zhu, Lei Zhang, Qinghua Hu and Simon C. K. Shiu. (2012). Multi-scale Patch Based Collaborative Representation for Face Recognition with Margin Distribution Optimization. In: *IEEE European Confernce in Computer Vision*, (pp. 822-835)