International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 15 (2018) pp. 12124-12133

A Novel Face Recognition System Based on Subclass Kernel Nonparametric Discriminant Analysis

Arbia Soula

Signal, Images and Information Technologies (LR-SITI-ENIT), University of Tunis El Manar, National School of Engineers of Tunis (ENIT), Street of Béchir Salem Belkhiria Campus universitaire, BP 37, 1002, Le Bélvédère, 1002, Tunis, Tunisia. arbiasoula1@gmail.com

Riadh Ksantini

University of Windsor, Windsor, ON, Canada
SUPCOM: Sécurite Numérique
ksantini@uwindsor.ca

Salma Ben Said

National Institute of Applied Sciences and Technology (INSAT), 676 INSAT North urban center BP, Cedex 1080,Tunis, Tunisia, Signal, Images and Information Technologies (LR-SITI-ENIT)
salma_ben_said@yahoo.fr

Zied Lachiri

Electrical Engineering Departement, Research laboratory SITI, ENIT, University of Tunis El Manar, National School of Engineers of Tunis (ENIT), Street of Béchir Salem Belkhiria Campus universitaire, BP 37, 1002, Le Bélvédère, 1002, Tunis, Tunisia.
lachiri.z@gmail.com

Abstract

This paper presents a Novel Face Recognition System Based on Subclass Kernel Nonparametric Discriminant Analysis (SKNDA), which incorporates subclass information in the KNDA optimization process. It jointly minimizes the data dispersion within and between subclasses to improve classification accuracy. Moreover, SKNDA has the advantage of reducing data dimension in a discriminative way and performing a discriminant eigenspace representation, where some near-global variations of the data are incorporated in the kernel space, to handle heteroscedastic, non-normal and non-linearly separable face classes. In order to make adequate face recognition, we integrate the Gabor features and the ordinal measures to derive the facial features, which are encoded in local regions, as visual primitives. The different ordinal measures are extracted from Gabor filtering responses. Then, the statistical distributions of these primitives in diverse face image blocks are concatenated, which generates a feature vector whose dimension is reduced using PCA. Finally, the latter is employed as a feature input for the proposed SKNDA. An extensive comparison of the SKNDA model to relevant existing kernel classifiers is performed on real world datasets to show the advantages of our proposed method. In particular, the experiments on face recognition have clearly shown the superiority of the SKNDA over other methods.

Keywords: discriminant models; machine learning; nonparametric discriminant analysis; kernel; features extraction; Gabor filter; biometrics; face recognition; ordinal filter; ordinal measures.

INTRODUCTION

Past attempts performed on face recognition and its applications are highly recognized. Among these applications are access control, identity verification and video surveillance [1]. However, face recognition methodology still possesses many challenging problems in practical applications due to the large intra-class facial variations, caused by illumination, expression, pose, aging, occlusion and the small inter-class similarity [2]. Broadly defined, then, it is the ability to determine a subject's identity based on his facial characteristics.

The performance of face recognition system depends on the way feature vectors are extracted, so that they display higher relevant information, in order to be classified into appropriate classes.

To deal with these challenges, many promising research works have been conducted, which can be classified into two main classes [3], namely, appearance-based analysis and local feature description. Appearance-based analysis attempts to find a set of basis images from a training set and to represent a global appearance of a human face as a linear combination of these basis images with their projections in the subspace, using a given optimization criteria, as the case for Principal Component Analysis (PCA) [4], Independent Component Analysis ICA [5] and Fisher's Linear Discriminant (FLD) [6], which have been much recognized among the most prevailing and successful face representation methods [7]. In addition, many researchers have suggested extracting facial features vectors by utilizing spatial-frequency techniques, such as, Discrete Cosine Transform (DCT) [8] and Fourier transform [9]. Using these techniques, face images are transformed to the frequency domain and only the coefficients in the low-frequency band are maintained for face representation. In contrast to the subspace-based methods, these methods do not need a training process to learn the basic images.

Recently, there is a growing interest to develop face recognition systems based on local features as they are more robust to the intrinsic facial variations. In the local-based face representation, each local facial region is represented by a feature vector. Thus, only detailed traits within this specific area are encoded. Many representative methods based on local analysis, such as, Local Binary Patterns (LBP) [10], Local Matching Gabor (LMG) [11]and Fisher linear discriminant

Model (EFM) [12] have been proposed. However, such methods did not lead to good results given the high dimensional feature space, the low amount of information used and the variability of lighting, poses and facial expressions etc....

Thereby, several approaches have been developed to surmount the problems of face recognition based on filtering, dimension reduction and classification of facial features vectors. We can mention among them: Su et al [13] introduce a face recognition method which combines both global and local discriminative features, by keeping the low-frequency coefficients of Fourier transform and the local features extracted by Gabor wavelets. Subsequently, multiple Fisher's linear discriminant (FLD) classifiers are separately applied to the Fourier features and each patch of Gabor features. Finally, all these classifiers are integrated to form a hierarchical ensemble classifier. Alshebani et al [14] suggested a hybrid feature extraction technique for face recognition utilizing local representation techniques, the features extraction step is based on the local features of the principle holistic regions such as eye, nose and mouth and then the maximum intensity of these regions are calculated using Gabor filters. Finally, in the classification step, the Nearest Neighbours method (KNN) is employed to calculate the distances between the three regions feature vectors and the corresponding stored vectors. Salh and Mustafa [15] developed a face recognition method based on PCA and LDA for data dimension reduction, and then the support vector machine (SVM) is used in the classification stage. More precisely, this approach combines the two main techniques PCA and LDA to extract low-dimensional features from a high dimensional space, while separating images that are from different classes and gather images that are from the same class. Finally, the SVM algorithm is explored for robust decision. However, all the above mentioned face recognition methods are based on non-flexible linear classifiers. Indeed, these approaches are unsatisfactory for different face recognition problems, where data classes are non-normal, heteroscedastic and non-linearly separable.

Filani and Adebayo [16] represent appearance based methods for face recognition based on linear and nonlinear algorithms. Firstly, the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are used as linear dimensional reduction projection techniques to encode pattern information based on second order dependencies and for nonlinear methods, the Kernel Principal Components Analysis (KPCA) and Kernel Fisher Analysis (KFA) are employed to handle relationships between three or more pixels in face images. Finaly, Mahalinobis Cosine (MAHCOS) metric is utilized to determine the resemblance measure among two face images. For this reason, Yin et al [17] introduced a face recognition technique based on RBF kernel SVM and particle swarm optimization (PSO-SVM). It consists in searching the best position with respect to the corresponding optimal solution for an optimization problem in the virtual search space. This method employs PCA to extract the features of face images, which are used to train and test the PSO-SVM model. PSO can efficiently find optimal solutions in large search spaces and it is utilized to simultaneously optimize the parameters of SVM. KSVM considers only boundary data points (i.e., support vectors) to build the model, regardless the spread of the remaining data [18], which could affect the classification accuracy and face recognition performance.

More recently, Soula et al [19][20] presented a novel face recognition method based on Gabor and ordinal wavelets for facial feature extraction, and on flexible non-linear Kernel Fisher Discriminant Analysis (KFD) and Kernel Nonparametric Discriminant Analysis (KNDA), respectively, for dimension reduction and classification. In fact, KFD consists in performing the traditional Fisher's Linear Discriminant Analysis (LDA) in the kernel space, which minimizes the intra-class variances of feature classes and maximizes the distance between their means. On the other side, KNDA improves upon KFD by incorporating the paramount nearby global characteristics of the data to handle heteroscedastic and non-normal face classes. Nevertheless, KNDA could suffer from performance degradation under special circumstances that may appear in real-world cases, such as, light directions of imaging, differences of facial expression, pose and lighting variations and high variability in facial features, which would result in the existence of face class subclasses or clusters, and an ever increase of data dispersion.

Therefore, in this paper, we solve mainly this problem by building a novel face recognition system based on Subclass Kernel Nonparametric Discriminant Analysis (SKNDA). In fact, SKNDA is advantageous since it minimizes the dispersion of the samples within each subclass and between subclasses, while capturing correctly the structural information between class boundaries, in order to improve classification performance and system discriminatory power. More precisely, we modify the standard KNDA objective function by replacing the standard within scatter matrix with a within and between subclass dispersion matrix, so that the resulted objective function takes advantage of the subclass information in its optimization process. Also, SKNDA is based on kernelization to realize flexible non-linear separation between face classes, thereby handling nonlinearly separable data. Moreover, our novelty lies in the efficiency of utilizing facial feature extraction method, which integrates the benefit of combining distinctiveness of Gabor features with robustness of ordinal measures, as a relevant solution to simultaneously handle intra-person variations and inter-person similarity face images. As a matter of fact, it consists of different steps. First, to ameliorate the local details of face texture, multichannel Gabor wavelets are applied on the input face image. Second, to derive ordinal measures from the Gabor feature images, various ordinal feature analysis techniques are applied to the Gabor feature images. Third, the different ordinal measures are jointly encoded in local regions as visual primitives. Fourth, the spatial histograms of these primitives are concatenated into a feature vector whose dimension is reduced using PCA. Finally, the novel SKNDA is further used to reduce dimension and classify feature vectors. The introduced Novel face recognition system investigates the effectiveness of the SKNDA method, which discriminates effectively between face data classes, even if they are heteroscedastic, non-normally distributed and non-linearly separable, by incorporating the neighborhood of the decision boundary of the data subclass structure, in the kernel space.

The rest of the paper is organized as follows: In section 2, we describe in details the main steps of our proposed face recognition system, including the feature extraction and the face recognition steps. Moreover, we provide detailed mathematical derivation of the novel SKNDA. In section 3, we evaluate and compare the SKNDA technique to other relevant state-of-the-art kernel classifiers, on real datasets and face datasets. Finally, in the last section, we provide concluding remarks and perspectives.

FACE RECOGNITION SYSTEM

In this section, we describe in details the derivation of the SKNDA and the face recognition method. The latter includes two main steps: Feature extraction and face classification based on novel SKNDA. Finally, we provide the face recognition method algorithm.

Facial Feature Extraction

Feature extraction is one of the most important phases in the process of face recognition. It consists in finding a specific representation of the data that can highlight relevant information, which would help in overcoming the human facial complications, such as, the light directions of imaging, differences of facial expression, variation of pose and aging.

Indeed, in our method we utilize a local feature analysis technique, namely, Gabor Ordinal Measures (GOM) for face feature representation [21], which inherits the advantages of combining distinctiveness of Gabor features with robustness of some kinds of ordinal wavelets, as a hopeful solution to reduce intra-person similarities and maximize dissimilarity between persons. Thus, the 2D Gabor filters help to produce prominent local discriminating features that are appropriate for face recognition. The Gabor filters are defined as follows [22]:

$$\psi_{\mu,v}(z) = \frac{\|k_{\mu,v}\|^2}{\sigma^2} e^{\left(\frac{\|k_{\mu,v}\|^2 \|z\|^2}{2\sigma^2}\right)} \left[e^{ik_{\mu,v}z} - e^{-\frac{\sigma^2}{2}}\right]. \quad (1)$$

Where $\mathcal{V} \in \{0, \dots, 4\}$ and $\mu \in \{0, \dots, 7\}$ are the scale and orientation of the Gabor wavelets, respectively, and $z = (x, y)$ represents the spatial position. The wave vector $k_{\mu,v} = k_v e^{i\emptyset_\mu}$ has a magnitude $k_v = \frac{k_{max}}{\lambda^v}$, where λ is the frequency ratio between filters and $\emptyset_\mu = \frac{\pi\mu}{8}, \emptyset_\mu \in [0, \pi]$.

In practice, the Gabor wavelet decomposition and representation of the face image is the convolution of the image I with a family of Gabor kernels $\psi_{\mu,v}(z)$, defined as:

$$G_{\mu,v}(z) = I(z) \times \psi_{\mu,v}(z). \quad (2)$$

So, the Gabor wavelet response produced for a given scale and orientation in Equation (2), is a complex number, given by the following equation [23]:

$$G_{\mu,v}(z) = A_{\mu,v}(z) \,.\, e^{i\theta(z)} \quad (3)$$

where A and θ define the magnitude response and the phase of Gabor kernel at each image position z, respectively.

The complex Gabor filter is a powerful descriptor. In fact, it is a strong tool to characterize the image texture. Therefore, it can extract the local structure corresponding to specific spatial frequency, spatial locality and selective orientation, which are demonstrated to be discriminative and robust to expression changes and illumination.

The complex Gabor filtering can also be described with real part and imaginary part. Thus, we can obtain four features for each face image, which are phase, magnitude, real and imaginary Gabor feature images.

As for ordinal or multi-lobe differential filtering for ordinal feature extraction, they provide a richer representation of facial features and are well-conditioned to uniform noise. In essence, ordinal features have the advantage of describing the neighboring relationship not only in image space, but also in various orientations and scales of Gabor responses.

From a mathematical perspective, multi-lobe differential filter (MLDF) is formed by many positive and negative lobes, which allow the arrangement of dissociated image regions in intensity level and feature level. In the intensity level, only the qualitative relationship between the average intensity values of two image regions is considered, whereas in the feature level, qualitative details on the image features are calculated. Thus, the MLDF has the benefit of invariance to monotonic illumination variation and robustness against noise. The ordinal values are distinctively “0” or “1” as the filtering results are, respectively, negative or positive. The MLDF can be presented with Gaussian Kernels as follows:

$$MLDF = c_p \sum_{i=1}^{N_P} \frac{1}{\sqrt{2\pi\delta_{pi}}} e^{\left[\frac{-(x-\omega_{pi})^2}{2\delta_{pi}^2}\right]} c_n \sum_{j=1}^{N_n} \frac{1}{\sqrt{2\pi\delta_{nj}}} e^{\left[\frac{-(x-\omega_{nj})^2}{2\delta_{nj}^2}\right]} \quad (4)$$

where ω is the central position, δ is the scale of 2D Gaussian filter, N_n is the number of negative lobes and N_p is the number of positive lobes and c_n and c_p are two constants. Different ordinal feature analysis methods are applied to Gabor feature images, such as, Gabor magnitude, phase, real and imaginary, in order to capture the robust ordinal features in diverse directions.

Face image can be analyzed on two levels: local intensity level and local feature level. As for local intensity variation, it is insubstantial as facial skin would have the same intensity of reflection ratio. In response to such a limited role, ordinal measures derived from features level become more powerful as they have a manifest discriminatory power in face recognition. Consequently, the use of Gabor filter aims at getting a more discriminative feature as well as ameliorating the local details of face texture. As a pleasing result, the integration of feature Gabor images with ordinal filters leads to a better recognition rate.

As a result, the ordinal measures derived from different components of Gabor images significantly expand the feature vector space of a face image. Thus, it is essential to integrate many binary codes in GOM feature images to find a discriminant texture parameter and minimize the length of GOM feature.

The Gabor Ordinal Measures (GOM) for face feature representation could be described by the following algorithm:

Algorithm1. Face features extraction based on Gabor and Ordinal filters

1. We utilize a series of 2D Gabor filters constituted by five frequencies and eight orientations on every pixel of a face, so each type of Gabor feature images includes 40 samples and has the same size as the original face image, using the equations (1), (2) and (3) .

2. We employ four tri-lobe and four di-lobe ordinal filters using equation (4), with orientation values equal to 0º, 45º, 90º and 135º on Gabor response images in order to obtain ordinal measures .

3. We combine every eight Gabor Ordinal Measures for every pixel, given MLDF and scale into a visual code as the following model:

$$GOM-MAP-\sigma_v^i(x,y)=[GOM-\sigma_{0,v}^i(x,y),GOM-\sigma_{1,v}^i(x,y),\ldots,GOM-\sigma_{7,v}^i(x,y)]$$

Where σ ϵ{*p: phase, m: magnitude, i: imaginary, r: real*} and $GOM-\sigma_v^i\,(x,y)$ is the texture primitive obtained at position (x,y) for the $i-th$ ordinal measure at scale v.

4. Spatial histograms are derived from each visual primitive. Then, they are joined to form the statistical distributions of these primitives in a face image.

5. We employ the Principal Component Analysis (PCA) to minimize feature vectors dimensions and maintain the appropriate information.

The Novel Subclass Kernel Nonparametric Discriminant Analysis (SKNDA)

In this section, we describe the derivation of the novel Subclass Kernel Nonparametric Discriminant Analysis (SKNDA), which is a discriminative dimensionality reduction and classification tool. The SKNDA assumes the existence of subclasses in the classes related to, e.g., different facial expressions, different viewing angles, pose and lighting variations and high variability in facial features, and minimizes the feature vectors dispersion between subclasses and within each subclass, in the kernel space. Hence, we modify the standard KNDA objective function by plugging in a within and between subclass dispersion matrix $S_{WK}^{subclass}$, so that the resulted objective function takes advantage of the subclass information in its optimization process. Also, SKNDA is based on the principle that the normal vectors on the decision boundary are the most informative for discrimination. Thus, it attempts to extract nonparametric discriminating features with the computation of the between-class scatter matrix S_{BK} on a local basis in the neighborhood of the decision boundary, as a promising solution to enhance classification performance.

Let us assume that we have $C_b, b=1,2,\ldots,L$ classes constituting an input space of $N=\sum_{b=1}^{L} nC_b$ samples $\{x_d\}_{d=1}^{N}\ in\ \mathbb{R}^M$, where each class C_b is formed by nC_b samples , $C_b=\{x_1^b,x_2^b,\ldots,x_{nC_b}^b\}$. Hence, the SKNDA classifier has two stages: it describes the nonlinear mapping function and transforms the data samples into a large dimensional feature space Ƒ, where linear classification can be realized to learn nonlinear relations with a linear classifier.

Let the function φ maps the classes $C_b, b=1,2,\ldots,L$ to high dimensional feature class $Ƒ_b=\{\varphi(x_i^b)\}_{i=1}^{nC_b}$, $b=1,2,\ldots,L$, respectively. However, if Ƒ is very high dimensional, this will be impossible to perform mapping directly. In this regard, the kernel trick [24] is employed to calculate the dot products of the higher-dimensional data instead of the samples themselves. It is defined by this formula:

$$K(x_i,x_j)=\langle\varphi(x_i),\varphi(x_j)\rangle,\quad \forall\, i,j\in\{1,2,\ldots,N\}. \qquad (5)$$

Thus, the decision function is described as follows:

$$y(x_i,\omega)=\sum_{i=1}^{N} f_i^x\omega_i+\omega_{0,} \qquad (6)$$

where $(f_1^x,f_2^x,\ldots,f_N^x):\mathcal{X}\to Ƒ$ represent a non-linear mapping function from the input space to a output feature space for the input variable x , $f_i^x=K(x,x_i)$, $\forall\, i\in\{1,2,\ldots,N\}$ and $\{\omega_i\}_{i=1}^{N}$ are the weights to be estimated.

The objective of the SKNDA is to find an adequate subspace in order to optimize the projection direction by maximizing the between-class distance of the kernel feature vectors classes, while minimizing the feature vectors dispersion between subclasses and within each subclass. Thus, it consists of estimating the projection direction ω that maximizes the following objective function:

$$J(\omega)=\frac{\omega^T S_{BK}\,\omega}{\omega^T S_{WK}^{subclass}\,\omega}, \qquad (7)$$

where $S_{WK}^{subclass}$ is the within and between subclasses dispersion scatter matrix, which is defined as follows:

$$S_{WK}^{subclass}=\overbrace{\sum_{b=1}^{L}\sum_{k=1}^{N_{subC_b}}\sum_{j=1}^{nC_{bk}} p_{bk}\left(x_j^{Kbk}-\mu_{bk}^K\right)\left(x_j^{Kbk}-\mu_{bk}^K\right)^T}^{\text{Within Subclasses Matrix}} + \underbrace{\sum_{b=1}^{L}\sum_{k=1}^{N_{subC_b}}\left(\mu_{bk}^K-\mu_b^K\right)\left(\mu_{bk}^K-\mu_b^K\right)^T}_{\text{Between Subclasses Matrix}} \qquad (8)$$

Where N_{subC_b} is the number of subclasses of class C_b, *b=1,2,…,L*, nC_{bk} is the samples number of subclass k of class

C_b , $p_{bk} = \frac{nC_{bk}}{nC_b}$, $\left(x_j^{Kbk}\right)_d = K\left(x_j^{bk}, x_d\right)$ $(\forall\, d \in \{1,2,\dots,N\})$ are the components of the kernel vector associated to the data sample x_j^{Kbk} , $(\mu_b^K)_d = \frac{\sum_{i=1}^{nC_b} K(x_i^b, x_d)}{nC_b}$ and $(\mu_{bk}^K)_d = \frac{\sum_{j=1}^{nC_{bk}} K\left(x_j^{bk}, x_d\right)}{nC_{bk}}$ $(\forall\, d \in \{1,2,\dots,N\})$ are the kernel mean vectors of class C_b and its subclass k, respectively. Therefore, the within and between subclasses dispersion scatter matrix can be rewritten as follows:

$$S_{WK}^{subclass} = \sum_{b=1}^{L} \sum_{k=1}^{N_{subC_b}} K_{bk}(I - 1_{nC_{bk}})\, {K_{bk}}^T + \sum_{b=1}^{L} \sum_{k=1}^{N_{subC_b}} (\mu_{bk}^K - \mu_b^K)(\mu_{bk}^K - \mu_b^K)^T, \qquad (9)$$

where, K_{bk} is the kernel matrix of the subclass k of the class C_b , with $(K_{bk})_{jd} = K\left(x_j^{bk}, x_d\right)$, $(j,d) \in \{1,2,\dots,nC_{bk}\} \times \{1,2,\dots,N\}$, and $1_{nC_{bk}}$ is the matrix with all entries $\frac{1}{nC_{bk}}$. The between-scatter matrix S_{BK} is defined as follows:

$$S_{BK} = \frac{1}{N} \sum_{b=1}^{L} \sum_{c=1, c\neq b}^{L} \sum_{k=1}^{N_{subC_b}} \sum_{j=1}^{nC_{bc}} \Psi_j^{bk}\left(C_b, C_c\right) L_b(x_j^{Kbk}) L_b(x_j^{Kbk})^T. \qquad (10)$$

Where Ψ_j^{bk} are the weighting functions to nullify the effects of samples that are far from the boundary [25]. It is described as follows:

$$\Psi_j^{bk}(C_b, C_c) = \frac{min\left\{d\left(x_j^{Kbk},\ M(\kappa - NN\left(x_j^{Kbk}\right)^b)\right)^\gamma, d\left(x_j^{Kbk},\ M(\kappa - NN\left(x_j^{Kbk}\right)^c)\right)^\gamma\right\}}{d\left(x_j^{Kbk},\ M(\kappa - NN\left(x_j^{Kbk}\right)^b)\right)^\gamma + d\left(x_j^{Kbk},\ M(\kappa - NN\left(x_j^{Kbk}\right)^c)\right)^\gamma}. \qquad (11)$$

γ is a control parameter which can range from zero to infinity, and $d\left(x_j^{Kbk},\ M(\kappa - NN(x_j^{Kbk})^b\right)$ and $d\left(x_j^{Kbk},\ M(\kappa - NN\left(x_j^{Kbk}\right)^c)\right)$ are the Euclidean distances from x_j^{Kbk} to the means of its κ nearest neighbors *($\kappa - NN$'s)* from classes C_b and C_c in the kernel space, respectively, where $M(\kappa - NN\left(x_j^{Kbk}\right)^c) = \frac{1}{\kappa}\sum_{h=1}^{\kappa}\left(\kappa - NN\left(x_j^{Kbk}\right)^c\right)(h)$ and $M(\kappa - NN\left(x_j^{Kbk}\right)^b) = \frac{1}{\kappa}\sum_{h=1}^{\kappa}\left(\kappa - NN\left(x_j^{Kbk}\right)^b\right)(h)$, with $\left(\kappa - NN\left(x_j^{Kbk}\right)^b\right)(h)$ and $\left(\kappa - NN\left(x_j^{Kbk}\right)^c\right)(h)$ are the h^{th} nearest neighbors of data sample x_j^{Kbk} from classes C_b and C_c, respectively. Here, each component of the matrix $L_b(x_j^{Kbk})$ is defined as:

$$\left(L_b(x_j^{Kbk})\right)_d = x_j^{Kbk} - \left(M(\kappa - NN\left(x_j^{Kbk}\right)^c)\right)_d,\ (j,d) \in \{1,2,\dots,nC_{bk}\} \times \{1,2,\dots,N\}. \qquad (12)$$

More precisely, κ is the free parameter which defines how many neighbors to consider. This parameter needs to be optimized for each dataset. Formula (12) represents the direction of the gradients of the respective class density functions in the kernel space [26].

Problem (7) can be resolved by obtaining the principal eigenvalues and eigenvectors of $(S_{WK}^{subclass})^{-1} S_{BK}$. However, since the higher-dimensional kernel space is of dimension N, numerical problems could cause the matrix $S_{WK}^{subclass}$ not to be invertible. Hence, the matrix $S_{WK}^{subclass}$ needs to be regularized before calculating its inverse. This is achieved by adding a small multiple β of the identity matrix I [27]. Therefore, the optimal solution ω^*, defining the optimal decision hyper-plane, consists of the eigenvector corresponding to the largest eigenvalues of $(S_{WK}^{subclass} + \beta I\,)^{-1} S_{BK}$.

Face Recognition Method Algorithm

The algorithm of the face recognition method is defined as follows:

Algorithm 2. Face Recognition Algorithm

---

1. Facial feature extraction with (Gabor Ordinal Measures) for compact and discriminative feature representation.
2. Generate the discriminative subclasses into each class using a clustering algorithm.
3. Mapping each feature vector into kernel feature space, using RBF Kernel $K(x, x_i) = e^{-\|x - x_i\|^2/\sigma}$, where σ is the "width" parameter.
4. Training phase: Train the SKNDA on the different classes of face feature vectors in the Kernel space using equations (9) and (10) in order to derive nonparametric discriminant faces features.
5. Testing phase: the trained SKNDA is used to project the test samples into different lower dimension classes.
6. The Euclidean distance is used to classify the testing samples into resulted different lower dimension classes.
7. Face Recognition.

EXPERIMENTAL RESULTS

In this section, we evaluate and compare the proposed SKNDA against relevant state-of-the-art classifiers, in order to show the advantage of incorporating the subclass information of each class and of minimizing the distribution of the samples within each subclass and between subclasses. Then, we provide a comparative evaluation of our face recognition method using SKNDA on well-known face datasets.

Datasets Used

We have used both real world datasets and faces datasets in our experiments to test the effectiveness of our proposed method SKNDA in a general context and in face recognition context. For the experiments on real world datasets, detailed description of these datasets can be found in Table 1. These datasets have been extracted from the UCI machine learning repository and selected minutely [28], so that we have a variety of dimensions and sizes. In the face recognition context, we have used face databases which are described as follows:

Yale face database [29]: This database is created by Yale University. It contains 5760 images of 10 subjects and the number of pictures per person is 576 under different viewing conditions (9 poses x 64 illumination conditions).

UMIST [30]: This Face Database consists of 564 images of 20 individuals (mixed race/gender/appearance), each individual is shown in a range of poses from profile to frontal views and the number of pictures per person from around 24 a 84.

AR [31]: The AR database consists of 1190 images of 85 individuals taken under very significant changes in illumination, facial expression and occlusions.

FRAV2D [32]: This database is composed of 109 subjects, with 32 images per person, taken under different poses and lighting conditions.

PIE [33]: This database is composed of 68 subjects, with 60 images per person, taken for different expressions and under different poses and illumination conditions.

Table 1. Description of the real data sets.

Dataset name	Number of class1 samples	Number of class2 samples	Number of classes	Number of features
Breast Cancer	233	430	2	9
diabetes	268	500	2	8
diabetes1	268	500	2	7
German	700	300	2	24
Heart Disease	134	159	2	13
liver	145	200	2	6
Thyroid	150	35	2	5
Ionosphere	225	126	2	34
Haberman's survival	225	81	2	3
Twonorm	3697	3703	2	20
Waveform	3304	1696	2	21
SPECT	212	55	2	44
Ringnorm	3736	3664	2	20

Experimental Protocol

We have organized the experimental comparative evaluation into two sets: The first set of experiments was performed on the real datasets (see Table 1 and table 2), to show the efficiency of the proposed SKNDA in terms of classification performance. The SKNDA has been evaluated and compared to three other contemporary classifiers, namely, the Kernel Support Vector Machines (KSVM) [24], the Kernel Fisher Discriminant (KFD) and the KNDA. The second set of experiments was carried out on the face datasets described above, in order to evaluate the proposed face recognition system based on SKNDA, in terms of recognition performance.

To make sure that the obtained results are not coincidental or biased, for each evaluated classification technique, we utilize a 10 fold cross validation process [34]. More precisely, each used dataset was randomly split into 10 subsets of equal size. Then, to build a model, one of the 10 subsets was removed to represent test samples, and the rest was used as the training data. Finally, the accuracy over all models of a dataset is estimated by averaging over the 10 obtained accuracy estimates.

For data kernelization, we have used the Gaussian RBF kernel $K(x, x_i) = e^{-\|x-x_i\|^2/\sigma}$, where σ represents the positive "width" parameter. In fact, this kernel was proven to be robust and flexible [35].

For the SKNDA and KNDA, cross-validation and grid search are used in order to get the combination of hyperparameters, namely, σ and κ that yield the best classification performance. As far as the KFD is concerned, the tuning of σ is achieved using cross-validation. Regarding KNDA and SKNDA, we performed 10 independent runs with different nearest neighbor numbers $\kappa \in \{1,2,\dots,10\}$, respectively, and then we select the value of κ that provides best classification performance. In order to evaluate and compare two class classifiers in term of performance, the Receiver Operating Characteristic (ROC) [36] are mostly used. The ROC curve represents a powerful measure of the performance of studied classifiers. It does not depend on the number of training or testing data points, but rather it depends only on rates of correct and incorrect samples classification. The ROC curve is created by plotting the True Positive Rate (TPR) vs the False Positive Rate (FPR). To evaluate the different methods, we have also used the Area Under Curve (AUC) produced by the ROC curves [37]. Also, paired *t*-test confidence intervals (CI) estimated, over AUC values of 10 models, between the SKNDA and the KSVM, KNDA and KFD, in order to quantify the probability of the paired distributions being nearby or not, in order to favour the performance enhancement by SKNDA. The higher the confidence interval, the lower is the probability that the underlying distributions are statistically indifferent. In order to identify the clusters of each data class, we have chosen the clustering method and the validity index proposed in [38], as it performs well when clusters are highly overlapped or there is significant variation in their covariance structures. So, in order to initialize the clustering algorithm of [38], for all data sets used we set the number of clusters *Cmin* = 2 and *Cmax* = 10, with the assumption that each data set class has a minimum of 2 clusters (subclasses) to a maximum of 10 clusters. Hence, we obtain the partitions described in Table 2.

Concerning the face datasets, the classification accuracy has been used for the evaluation and comparison of each method recognition performance. It is defined by $100\,\frac{N_{CC}}{N_{test}}\%$, where N_{test} is the total number of testing samples and N_{CC} is the number of samples classified correctly.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 15 (2018) pp. 12124-12133

RESULTS AND DISCUSSION

As we can see from Table 3, SKNDA provides best performance in terms of AUC values, on all real datasets, by averaging over 10 different models. Moreover, the last row of Table 3 provides the confidence intervals (in %) obtained from the performed *t*-tests. We can clearly notice that all the confidence intervals are high (very close to 100%), which shows that SKNDA indeed provides statistically significant accuracy improvements. In fact, SKNDA has the advantage of incorporating subclass information, in order to minimize the dispersion of the samples within each subclass and between subclasses, while capturing correctly the structural information between class boundaries, thereby improving classification discriminatory power.

Unsurprisingly, KSVM and KFD provide almost the lowest AUC values. In fact, the KSVM considers only boundary data points (i.e., support vectors) to build the model. However, these points do not completely represent the overall class. Generally, solutions to boundary-based methods are only calculated based on the points near the decision boundary, regardless the spread of the remaining data [38]. Thus, solutions to boundary-based methods like KSVM can be misled by the spread of data, since these methods tend to separate the data along large spread directions. On the other side, KFD assumes that the samples of each class are generated from underlying multivariate normal distributions of common covariance matrix with different means (i.e., homoscedastic data). Hence, KFD is incapable of dealing explicitly with heteroscedastic (classes with different covariance matrices) and normally distributed data.

The KNDA provides better classification results than the KFD and KSVM. This is based on the principal that the normal vectors on the decision boundary are the most suitable for discrimination. Thus, KNDA tries to find in feature space, the normal vectors to compute the between-class scatter matrix on a local basis in the neighborhood of the decision boundary, thereby relaxing the normality and homoscedasticity assumptions of the KFD. Also, we present some individual graphical results by plotting the actual ROC curves of the SKNDA, KNDA, KFD and KSVM for German and Breast Cancer datasets. Fig. 1 and Fig. 2 show the ROC curves of the four classifiers on the German and Breast Cancer datasets. The rule-of-thumb to judge the performance of a classifier from a ROC curve is “The best classification has the largest area under curve”. We can clearly see from the Fig. 1 and Fig. 2 that SKNDA indeed leads to best ROC curves.

According to table 4, as far as the face datasets are concerned, we can see that the proposed face recognition method based on SKNDA outperforms the other recognition systems based on KFD, KSVM and KNDA, in terms of recognition accuracy. This was expected, as the SKNDA takes into account subclass information and minimizes face classes dispersion caused by special circumstances, such as, light directions of imaging, differences of facial expression, pose and lighting variations and high variability in facial features, which would result in an improvement of the face recognition system performance.

In addition, we present in table 3 the average training times, computed over 10 model runs, of the compared classifiers. From this table, we can clearly remark that SKNDA, KFD and KNDA outperform the KSVM in terms of running time. In fact, the KSVM scales with $O(N^2)$ [35], where N is the number of data samples, whereas each of the KFD, KNDA and SKNDA scales with a computational complexity of $O(N^3)$ [6], but requires only one iteration to converge. Here, the SKNDA and KNDA have a slight overhead running time, since they are based on the κ nearest neighbors estimation.

Table 2. Number of subclasses of each class for the 13 real-world datasets

Dataset name	Number of class1 samples	Number of class2 samples	Number of classes	Number of features
Breast Cancer	233	430	2	9
diabetes	268	500	2	8
diabetes1	268	500	2	7
German	700	300	2	24
Heart Disease	134	159	2	13
liver	145	200	2	6
Thyroid	150	35	2	5
Ionosphere	225	126	2	34
Haberman's survival	225	81	2	3
Twonorm	3697	3703	2	20
Waveform	3304	1696	2	21
SPECT	212	55	2	44
Ringnorm	3736	3664	2	20

Table 3 Average AUC, Confidence and Training Time (in Seconds) of each method for the 13 real-world datasets (best method in bold, second best emphasized).

Dataset name	KSVM	KFD	KNDA	SKNDA
Breast Cancer	77.9	77.8	*92.53*	**95.8**
diabetes	78.10	77.7	*79.89*	**82.2**
diabetes1	74.70	73.26	*80.37*	**82.85**
German	79.10	78.8	*91.30*	**93.51**
Heart Disease	*85.64*	83.4	85	**87.79**
liver	78.74	78.20	*80.69*	**81.2**
Thyroid	96.62	95.98	*97.3*	**98.7**
Ionosphere	85.56	83.8	*90.23*	**92.89**
Haberman's survival	84.5	83.87	*89.55*	**90.19**
Twonorm	97.5	95.3	*97.85*	**98.2**
Waveform	91.4	90	*93*	**93.78**
SPECT	87.72	84.1	*90.35*	**91.18**
Ringnorm	97.90	96.2	*98.22*	**98.67**
Avg.time	4.15	2.60	2.68	2.67
confidence	99.6	99.8	97.4	-

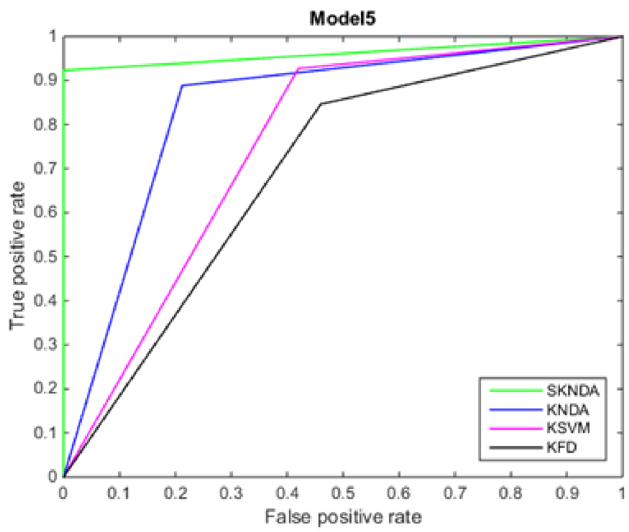

Figure 1. ROC curves for the SKNDA and the three classifiers applied on German

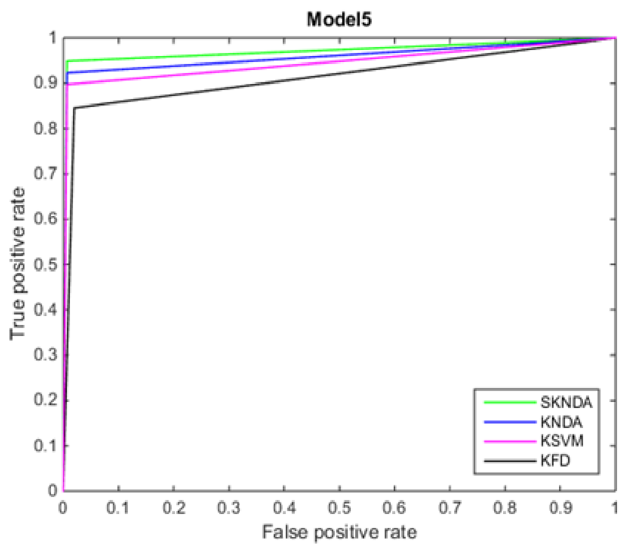

Figure 2. ROC curves for the SKNDA and the three classifiers applied on Breast Cancer.

Table 4. Classification accuracy of each classifier on the face databases (best method in bold, second best emphasized).

Database	KSVM	KFD	KNDA	SKNDA
UMIST	91.79	90.90	97.0	97.80
AR	88.6	87.81	95.20	95.73
YALE	90.2	89.54	96.09	97.7
FRAV2D	77.8	76.12	85.54	89.3
PIE	80.7	79.6	87.15	90.9

CONCLUSION

In this paper, we have proposed a novel face recognition system based on Subclass Kernel Nonparametric Discriminant Analysis (SKNDA). This latter is advantageous since it minimizes the dispersion of the samples within each subclass and between subclasses, while capturing correctly the structural information between class boundaries, in order to improve classification performance and system discriminatory power. Moreover, the SKNDA between-class scatter matrix is estimated on a local basis, using Nearest Neighbours method (KNN), to deal with non-normal and heteroscedastic data in a proper manner. Also, SKNDA is based on kernelization to realize flexible non-linear separation between face classes, thereby handling nonlinearly separable data.

As far as face recognition is concerned, our novelty lies in the efficiency of utilizing facial feature extraction method, which integrates the benefit of combining distinctiveness of Gabor features with robustness of ordinal measures, as a relevant solution to simultaneously handle intra-person variations and inter-person similarity face images.

We performed evaluation and comparison of the SKNDA to relevant state-of-the-art classification algorithms KSVM, KFD, and KNDA on real world datasets and face datasets. Experiments showed the effectiveness of the proposed novel face recognition system and the superiority of the SKNDA in terms of classification and face recognition performance.

In future work, we will investigate effectiveness of SKNDA in other recognition problems.

REFERENCES

[1] Jain, A. K., Klare, B., and Park, U., 2011, "Face recognition: Some challenges in forensics," 9th IEEE International Conference on Automatic Face and Gesture Recognition., Santa Barbara, CA, March, pp. 726-733.

[2] Jain, Anil K., and Stan Z. Li., 2011, Handbook of face recognition, New York, 2011.

[3] Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A., 2003 "Face recognition: A literature survey," Journal of ACM computing surveys (CSUR), 35(4), pp. 399-458.

[4] Turk, M., and Pentland, A., 1991, "Eigenfaces for recognition," Journal of cognitive neuroscience, 3(1), pp.71-86.

[5] Bartlett, M. S., Movellan, J. R., and Sejnowski, T. J., 2002, "Face recognition by independent component analysis, " IEEE Transactions on neural networks Learning Systems, 13 (6), pp. 1450-1464.

[6] Belhumeur, Peter N., João P. Hespanha., and David J. Kriegman., 1997, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Transactions on pattern analysis and machine intelligence, 19(7), pp.711-720.

[7] Sun, Z, and Tieniu,T., 2009, "Ordinal measures for iris recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(12), pp. 2211-2226.

[8] Hafed, Z. M., and Martin, D. L., 2001, "Face recognition using the discrete cosine transform," International Journal of Computer Vision, 43(3) pp.167-188.

[9] Lai, J.H., Pong, C. Y., and Guo, C.F., 2011, "Face recognition using holistic Fourier invariant features," Pattern Recognition, 34 (1), pp.95-109.

[10] Pietikäinen, M., Hadid, A., Zhao, G., and Ahonen, T., 2011, "Local binary patterns for still images," Computer vision using local binary patterns. Springer London, pp. 13-47.

[11] Perez, C. A., Leonardo, A. C., and Luis, E. C., 2011 "Local matching Gabor entropy weighted face recognition," , IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG), pp.179-184.

[12] Liu, C. and Wechsler, H. 2000 "Robust coding schemes for indexing and retrieval from large face databases, ",IEEE Transactions on image processing, 9 (1), pp.132-137.

[13] Su, Y., Shan, S., Chen, X., and Gao, W., 2009 "Hierarchical ensemble of global and local classifiers for face recognition, " IEEE Transactions on Image Processing,18(8), pp. 1885-1896.

[14] Alshebani, Q., Premaratne, P. and Vial, P., 2014 "A hybrid feature extraction technique for face recognition, " 2nd International Conference on Information Technology and Science (ICITS), pp.166.170.

[15] Salh, T. A. and Nayef, M. Z., 2013 "Face Recognition System based on Wavelet, PCA-LDA and SVM, "Computer Engineering and Intelligent Systems Journal, 4 (3), pp. 26-31.

[16] Filani, Araoluwa S., and Adebayo, O. A., 2016 "Development of an Efficient Face Recognition System based on Linear and Nonlinear Algorithms," International Journal of Computer Applications, 134(7) , pp- 27-30.

[17] Yin, H., Qiao, J., Fu, P., ans Xia, X. Y., 2014 "Face feature selection with binary particle swarm optimization and support vector machine, " J. Inf. Hiding Multimed. Signal Process, 5(4), 731-739.

[18] Shivaswamy, P.K., and Tony, J., 2010 "Maximum relative margin and data-dependent regularization," Journal of Machine Learning Research, 11,pp- 747-788.

[19] Soula, A., Ksantini, R., Said, S. B., and Lachiri, Z., 2016 "A Novel Kernelized Face Recognition System," 4th International Conference on Control Engineering & Information Technology (CEIT), pp.1-5.

[20] Soula, A., Ksantini, R., Said, S. B., and Lachiri, Z., 2017 "A novel face recognition system base on nonparametric discriminant analysis," 2017 International Conference on Control, Automation and Diagnosis (ICCAD), pp. 343-348.

[21] Chai, Z., Sun, Z., Méndez-Vázquez, H., He, R., and Tan, T., 2014, "Gabor ordinal measures for face recognition, " IEEE transactions on information forensics and security, 9(1), pp.14-26.

[22] Wiskott, L., Krüger, N., Kuiger, N. and Von Der Malsburg, C., 1997, "Face recognition by elastic bunch graph matching, ", IEEE Transactions on pattern analysis and machine intelligence, 19 (7), pp. 775-779.

[23] Serrano, Á., de Diego, I. M., Conde, C. and Cabello, E., 2010, "Recent advances in face biometrics with Gabor wavelets: A review,", Pattern Recognition Letters, 31(5), pp. 372-381.

[24] Vapnik, V. N. 1999, "An overview of statistical learning theory, " IEEE transactions on neural networks, 10(5), pp. 988-999.

[25] Fukunaga, K., 2013, Introduction to statistical pattern recognition, 2nd ed, Academic press, London.

[26] Fukunaga, K., and Larry, H., 1975, "The estimation of the gradient of a density function, with applications in pattern recognition," IEEE Transactions on information theory 21(1), pp. 32-40.

[27] Mika, S., Ratsch, G., Weston, J., Scholkopf, B., and Mullers, K. R., 1999, "Fisher discriminant analysis with kernels, " Neural networks for signal processing IX, Proceedings of the IEEE signal processing society workshop.

[28] Lichman, M., 2013, UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences.

[29] Georghiades, A- Yale face database, http://cvc.yale.edu/projects/yalefaces/yalefaces.html, 1997.

[30] https://www.sheffield.ac.uk/eee/research/iel/research/face

[31] http://www2.ece.ohiostate.edu/~aleix/ARdatabase.html, 1998.

[32] http:// www.frav.es.

[33] https://www.flintbox.com/public/project/4742/.

[34] Alippi, C., and Roveri, M., 2010, "Virtual k-fold cross validation: An effective method for accuracy assessment," In Neural Networks (IJCNN), The International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, pp. 1-6.

[35] Cristianini, N. and Shawe-Taylor, J., 2000, An introduction to support vector machines and other kernel-based learning methods, Cambridge university press.

[36] Fawcett, T., 2006, "An introduction to ROC analysis", Pattern recognition letters, 27(8), pp. 861-874.

[37] Hanley, J. A. and McNeil, B. J., 1983, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases, " Radiology, 148 (3), pp.839-843.

[38] Bouguessa, M., Wang, S., and Sun, H., 2006, "An objective approach to cluster validation," Pattern Recognition Letters, 27 (13), pp. 1419-1430.