Kernel Group Sparse Representation based Classifier for Multimodal Biometrics

Gaurav Goswami, Richa Singh, Mayank Vatsa, and Angshul Majumdar
IIIT-Delhi, India
Email: {gauravgs, rsingh, mayank, angshul}@iiitd.ac.in

***Abstract*—Classification is an important pattern recognition paradigm with a multitude of applications in popular research problems. Utilizing multiple data representations to improve the accuracy of classification has been explored in literature. However, approaches such as combining classifiers using majority voting and score level fusion do not utilize the underlying structure of the data which is available at the representation stage itself. In this paper, we propose a kernelization based extension to the group sparse representation classifier which can utilize multiple representations of input data to improve classification performance. By using a kernel, these representations are processed in a higher dimensional space where they are more separable, without substantially increasing computational costs. The proposed algorithm selects the ideal kernel to use along with its parameters automatically as part of the training process. We evaluate the proposed algorithm on three challenging biometric problems namely, cross distance face recognition, RGB-D face recognition, and multimodal biometrics to showcase its efficacy. Experimentally, we observe that the proposed algorithm can efficiently combine multiple data representations to further improve classification performance.**

I. Introduction

Classification is a fundamental and quintessential task in pattern recognition and machine learning. It is also an integral aspect of many research problems such as object recognition and biometrics. Research to improve upon the accuracy and efficiency of classification algorithms has been a prominent research focus. One of the popular techniques to improve classifier accuracy is to train multiple classifiers to solve a single problem and then combine their decisions, i.e., to design a classifier ensemble. Usually methods such as score or rank level fusion or majority voting are utilized to combine classifiers in these scenarios. Combining classifiers is also important when each data point can be represented in multiple feature spaces and information from each of these representations contributes towards making a more accurate final decision. Features are an integral part of the classification process as the classifier attempts to fit a decision boundary to the data points in the feature space.

Sparse representation is a technique to extract features that relies on representing an unseen data point as a sparse combination of seen data points. Sparse representation based approaches have seen a wide variety of successful application. Sparse representation has been used in natural image sharpening [1], image restoration [2], face recognition [3], and medical image analysis [4]. Joint and group sparse based representations have been used for multi-modal biometrics [5], object and video face recognition [6], image super-resolution [7], and enhancement of low resolution images [8].

Fig. 1: A few RGB-D face images where the group sparse classifier [9] is unable to perform accurate recognition but the proposed kernel group sparse classifier is able to perform accurately. The second and fourth column present the depth maps corresponding to the color images presented in the first and third columns.

In our previous research, we have demonstrated that utilizing a classifier that can inherently handle multiple features and exploit the underlying structure of the problem can offer performance improvement over other techniques [9]. The group sparse classifier represents each test sample as a combination of its representations in individual feature spaces and classifies based on the residual error for each class. The quality of classification depends on the intra-class stability and inter-class differentiability of these individual representations. Projection into a higher dimensional space using an appropriate kernel function is a popular technique to achieve better separability of data. If each feature space is projected using a suitable kernel, the quality of the combined representation should improve. It is our assertion that the accuracy of the group sparse classifier can be further improved using kernelization without decreasing computational efficiency. Even though Gao *et al.* [10] have discussed kernelization in a sparse representation based approach with good results, they have not considered a multi-feature or multi-modal scenario. In such case, the algorithm would have to rely on score level or decision level fusion rules. In this paper, we explore the applicability of kernelization to the group sparse representation based classifier to combine multiple features. We evaluate the proposed algorithm on a variety of biometrics problems such as recognition in surveillance images in a cross-distance matching scenario, face

recognition for RGB-D images obtained using Kinect sensor (as shown in Fig. 1), and multi-modal biometrics using face, iris, and fingerprint. The rest of this paper is organized as follows: section 2 presents the proposed algorithm, section 3 discusses the biometric problems where we apply the proposed classifier, section 4 contains the experimental results and analysis, and section 5 concludes the paper.

II. Proposed KGSRC Algorithm

In this section, we first present the formulations of the traditional SRC for image based classification, followed by the proposed KGSRC algorithm. Traditional Sparse Representation Classification (SRC) assumes that the gallery images of a particular subject approximately form a linear basis for a new probe image belonging to the same subject. Formally, if I_{probe} is the probe image belonging to the k^{th} subject then,

$$I_{probe} = \alpha_{k,1}g_{k,1} + \alpha_{k,2}g_{k,2} + \cdots + \alpha_{k,n}g_{k,n} + \epsilon \quad (1)$$

where, $g_{k,i}$ denotes the i^{th} gallery image of the k^{th} subject and ϵ denotes the approximation error. In a biometrics based identification protocol, gallery images are provided for each subject. The identity of the probe image needs to be determined from among the subjects in the gallery, in a one-to-many matching scenario. The coefficients $\alpha_{k,i}$ in equation 1 need to be computed. Since the correct subject for a probe image is unknown, SRC represents it as a linear combination of all the gallery images,

$$I_{probe} = G\alpha + \epsilon \quad (2)$$

$$G = \Big[\underbrace{g_{1,1}|\ldots|g_{1,n}}_{g_1}|\underbrace{g_{2,1}|\ldots|g_{2,n}}_{g_2}|\ldots\underbrace{g_{c,1}|\ldots|g_{c,n}}_{g_c}\Big] \text{ and}$$

$$\alpha = \Big[\underbrace{\alpha_{1,1}|\ldots|\alpha_{1,n}}_{\alpha_1}|\underbrace{\alpha_{2,1}|\ldots|\alpha_{2,n}}_{\alpha_2}|\ldots\underbrace{\alpha_{c,1}|\ldots|\alpha_{c,n}}_{\alpha_c}\Big]$$

According to the core assumption of SRC, only the gallery images from the correct subject should form a basis for representing the probe. Therefore, the vector α should be sparse; it should have non-zero values corresponding only to the correct subject and zero values for other subjects. Thus equation 2 denotes a linear inverse problem which has a sparse solution. In [11], the coefficient α is solved by utilizing a L_1-norm minimization,

$$\min_{\alpha} ||I_{probe} - G\alpha||_2^2 + \lambda||\alpha||_1 \quad (3)$$

While the first term in equation 3 ensures that the approximation error in the solution is low, the second L_1 norm term enforces a sparsity criterion so that sparse solutions are favored during optimization.

As an extension to SRC, the Group Sparse Representation Classifier (GSRC) are proposed [12], [13]. Recently GSRC has also been applied for the problem of multi-modal multi-feature biometrics recognition [9]. It can be adapted for multi-spectral and cross-distance problems of RGB-D face recognition and video surveillance as well. There are two data sources in the case of RGB-D images, the color image and the depth image. In the case of cross-distance recognition, multiple feature extractors can act as different data sources to obtain different feature spaces. In the GSRC framework, we assume that the SRC model holds for both of these sources, i.e. the probe image from that source can be expressed as a linear combination of the gallery images of the correct subject from the same source. In other words, we assume that gallery depth images of a subject can be used to linearly reconstruct the probe depth images of the same subject with minimal error and same in the case of color images of a subject. Similarly, the case can be considered for two different features. For simplicity of expression, we assume two sources $s1$ and $s2$ for the following discussion. The concept of the GSRC and KGSRC algorithms is easily extensible to more than two sources as well. Considering this assumption, a probe of the s^{th} source can be represented as follows:

$$I^s_{probe} = (G)^s\alpha^{(s)} + \epsilon \quad \forall s \in \{s1, s2\} \quad (4)$$

Here, G^s denotes the gallery of the s^{th} source and α^s denotes the associated coefficients. ϵ denotes the residual reconstruction error. It is possible to solve each of these sources using the SRC approach and combine them at a later stage using some fusion rule. However, such an approach does not take advantage of the intrinsic structure of the problem. A better approach would be to combine both the sources into a single framework. This can be succinctly represented as,

$$\begin{bmatrix} I^{s1}_{probe} \\ \\ I^{s2}_{probe} \end{bmatrix} = \begin{bmatrix} G^{s1} & 0 \\ \\ 0 & G^{s2} \end{bmatrix} \begin{bmatrix} \alpha^{s1} \\ \\ \alpha^{s2} \end{bmatrix} + \epsilon \quad (5)$$

The coefficient vector (for simplicity shown as a row vector) for each source can be expanded as: $\alpha^s = \left[\alpha^s_1, \ldots, \alpha^s_k, \ldots, \alpha^s_c\right]$, where α^s_k denotes the coefficients corresponding to the k^{th} subject for the s^{th} source. Assuming that the test sample belongs to the k^{th} subject, coefficients from this subject should be non-zero. Considering both the sources,

$$\alpha = \Big[\underbrace{\alpha^{s1}_1, \ldots, \alpha^{s1}_c}_{\alpha^{s1}}, \underbrace{\alpha^{s2}_1, \ldots, \alpha^{s2}_c}_{\alpha^{s2}}\Big]^T \quad (6)$$

When the SRC assumptions hold true for individual modalities, the α^s_k's for each s (source) will have non-zero values. α then becomes a matrix of coefficients where each row consists of the coefficients for the feature vectors from a particular source. In such a case, α has a group sparse structure where the non-zero elements occur corresponding to the indices of the k^{th} subject. This leads to a group sparse representation where the grouping happens simply based on the indices. Therefore equation 5 can be solved using group sparsity promoting $L_{2,1}$ norm:

$$\min_{\alpha} ||I_{probe} - G\alpha||_2^2 + \lambda||\alpha||_{2,1} \tag{7}$$

$$I_{probe} = \begin{bmatrix} I_{probe}^{s1} \\ \\ I_{probe}^{s2} \end{bmatrix}, G = \begin{bmatrix} G^{s1} & 0 \\ \\ 0 & G^{s2} \end{bmatrix},$$

$$and \quad \alpha = \begin{bmatrix} \alpha^{s1} \\ \\ \alpha^{s2} \end{bmatrix}$$

Here, α^{s1} and α^{s2} are the vectors for each sample from the respective sources $s1$ and $s2$. It should be noted that the GSRC formulation does not suffer from the limitations of block sparse classification (BSC) [14], [15], [16]. Here, the probe image is not fit to all subspaces simultaneously (as is done by the BSC) and instead fit into subspaces spanned by the gallery images of the same source, i.e., gallery images of the subjects for the particular source (color or depth). The GSRC retains the versatility of the SRC and improves upon it by utilizing the problem structure. The broad distinction between SRC and GSRC is that the former groups according to the subject labels whereas the later groups according to the indices. However, the classification is based on the same principle as SRC. The representative sample for each subject for all the sources is computed as:

$$I_{rep}(k) = \begin{bmatrix} G_k^{s1}\alpha_k^{s1} \\ \\ G_k^{s2}\alpha_k^{s2} \end{bmatrix} \tag{8}$$

The test sample is assigned to the subject (*class*) having the minimum residual error between the test vector and the subject representative. Multimodal biometrics features might not be linearly separable in the feature space itself. This would lead to decreased performance due to the unavailability of an ideal decision boundary to separate samples belonging to the individual subjects. However, by utilizing kernelization, these features are projected to a higher dimensional space where they may be linearly separable, thereby resulting in better decision boundaries and better performance. For each source, we can have the equivalent non-linear representation:

$$\phi(I_{probe}^s) = \phi(G^s)\alpha^s + \epsilon, \quad \forall s \in \{s1, s2\} \tag{9}$$

Equation 9 can be kernelized by pre-multiplication of $\phi(I_{probe}^s)$

$$\phi(G^s)^T\phi(I_{probe}^s) = \phi(G^s)^T\phi(G^s)\alpha^s + \epsilon, \quad \forall s \in \{s1, s2\} \tag{10}$$

The kernel κ is defined as:

$$\kappa(x_i, x_j) = \left\langle \phi(x_i), \phi(x_j) \right\rangle \tag{11}$$

where $\left\langle .... \right\rangle$ represents the inner product. In subsequent equations, we use the notation $\kappa(x)$ to denote $\kappa(G^s, x)$. Using equation (11), equation (10) can be written as:

$$\kappa(I_{probe}^s) = \kappa(G^s)\alpha^s + \epsilon \tag{12}$$

As in GSRC, equation (12) can be compactly represented as,

$$\kappa(I_{probe}) = \kappa(G)\alpha + \epsilon \tag{13}$$

where, $\kappa(I_{probe}) = \begin{bmatrix} \kappa\left(I_{probe}^{s1}\right) \\ \\ \kappa\left(I_{probe}^{s2}\right) \end{bmatrix}$ *and* $\kappa(G) = \begin{bmatrix} \kappa(G^{s1}) & 0 \\ \\ 0 & \kappa(G^{s2}) \end{bmatrix}$

Similar to the GSRC approach, equation 13 can be solved using $L_{2,1}$ minimization which promotes group sparsity:

$$\min_{\alpha} ||\kappa(I_{probe}) - \kappa(G)\alpha||_2^2 + \lambda||\alpha||_{2,1} \tag{14}$$

Once α is obtained, the representation for each subject (class) is computed using Equation 8 and the probe image is assigned to the subject/class with the minimum residual error. An important problem in a kernel based approach is the selection of the kernel function and its parameters since these selections can greatly influence the final performance of the algorithm. Usually, in the case of existing machine learning approaches such as Support Vector Machines (SVMs) that utilize kernels, the kernel and its parameters are chosen empirically. There are certain techniques such as grid search, where a subset of possible parameter values for each kernel choice can be used for training and then optimized on the basis of maximum performance on a separate validation set. However, this requires enough training data for both the training and validation sets. The validation set is kept separate from the training set since optimizing training performance alone can lead to overfitting and reduction in performance on the test set.

We propose a method to select the kernel and its parameters using just one training set that is also able to preserve robustness and avoid overfitting, thereby avoiding the requirement for a large amount of data to create a separate validation set. Using the proposed kernel/parameter selection, the KGSRC algorithm selects the optimal kernel and its parameters automatically. This helps in focusing only on selecting the appropriate feature sources and training data, which become the parameters of the algorithm while a large set of possible kernels and associated parameters can be set as the search space for the kernel selection algorithm. To preserve robustness while performing parameter selection using the training set, we propose a metric to quantize the goodness of separation created by the transformation enacted by a kernel function. Given a list of kernel functions, K and a set of associated parameters θ_K, we compute the following metric for each possible combination for the training data:

$$\Psi_{K,\theta_{Ki}} = \frac{\Sigma_{a=1}^{C}\Sigma_{b=1}^{C}\chi^2\left(\mu_{Ki}(a),\mu_{Ki}(b)\right)}{\Sigma_{z=1}^{C}\Sigma_{x=0}^{n_z}\Sigma_{y=0}^{n_z}\chi^2(K_{\theta_i}(\gamma_{zx}),K_{\theta_i}(\gamma_{zy}))} \quad (15)$$

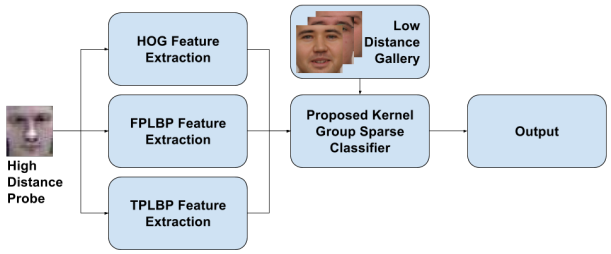
Here, θ_{Ki} denotes the i^{th} parameter setting for the kernel K, $K_{\theta_i}(\cdot)$ is the kernel function when using θ_{Ki} as the parameter setting, $\chi^2(x,y)$ denotes the χ^2 distance between x and y, C is the number of classes, n_z is the number of training samples that belong to a particular class z, γ_{zx} denotes the x^{th} training sample belonging to the z^{th} class, and $\mu_{Ki}(a)$ denotes the mean transformed vectors of class a after kernelization using the kernel K with parameter setting θ_{Ki}. The class mean for a class c is computed as follows:

$$\mu_{Ki}(c) = \frac{\Sigma_{x=0}^{x=n_c} K_{\theta_i}(\gamma_{cx})}{n_c} \quad (16)$$

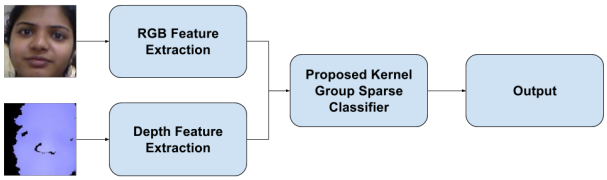
Essentially, equation 15 computes the joint kernel-parameter combination score Ψ which is a measure of how well a particular kernel-parameter function can optimize the trade-off between inter-class distance (numerator in equation 15) and intra-class distance (denominator in equation 15). By using the aggregated statistic for intra-class distance across all data points in training and using only the means to compute the inter-class distance, we help make the selection robust towards overfitting to the training data. Our assertion is that a kernel-parameter combination with a high Ψ score is one that is able to maximize this metric consistently over the entirety of the training data and should also perform well on the test data. The underlying assumption is that the test data is not drastically different from the training data which is required inherently for good performance of a sparse representation based approach even without the involvement of kernel selection. In all the discussed results, we evaluate the performance of the proposed algorithm with the proposed kernel selection approach being used to determine both the kernel and its parameters.

III. Case Studies

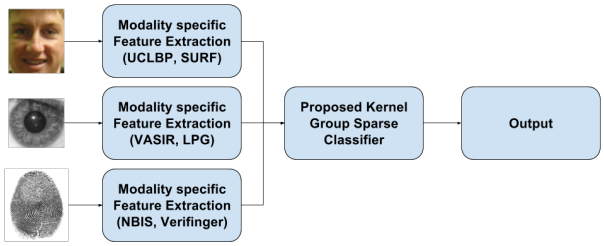
In this paper, we apply the proposed KGSRC algorithm to three challenging problems in biometrics recognition: (1) cross-distance face recognition, (2) RGB-D face recognition, and (3) multi-modal biometric recognition, to assess its performance. The flow diagrams for these problems is presented in Fig. 2. For a given probe sample, individual representations are extracted using different feature extractors, modalities, or spectrums as applicable. A sparse representation for the probe image is computed over the features extracted from the gallery images using the proposed KGSRC algorithm. The residual error for each subject is compared and the probe is assigned to the subject/class with the minimum residual error. In each case, face detection is performed using the Viola Jones detector [17] before processing an image during either training or testing and only the cropped face image is used. To populate our kernel list we have used the linear, spline, wave, sigmoid, radial basis function, polynomial, multi-quadratic, rational quadratic, and laplacian kernel classes along with correspondingly appropriate parameter ranges.



(a) Flow diagram for cross-distance face recognition using the proposed KGSRC algorithm.



(b) Flow diagram for RGB-D face recognition using the proposed KGSRC algorithm.



(c) Flow diagram for multi-modal biometric recognition using the proposed KGSRC algorithm.

Fig. 2: Flow diagrams for each of the three biometric problems that we use to evaluate the proposed KGSRC algorithm.

A. Cross-distance Face Recognition

Cross-distance face recognition is a challenging problem which occurs frequently when surveillance footage needs to be used for face recognition. In such a case, the face image is captured at a high distance from a fixed camera position that inherently imparts the problems of pose and illumination variation along with low-resolution. These images are then usually matched with mugshot images which are of high quality. To use the KGSRC algorithm in this scenario, we first perform face detection and crop the face image only from the background. We then utilize three different feature extractors, Three Patch Local Binary Patterns (TPLBP) [18], Four Patch Local Binary Patterns (FPLBP) [19], and HOG [20], and perform feature level fusion using the KGSRC algorithm. The training data is used as gallery for each subject, to learn the ideal kernel parameters, and to learn the sparse representation coefficients. The same process of extracting multiple features is employed for a given probe image at runtime and the KGSRC algorithm computes a ranked list of subjects ordered by increasing residual error.

B. RGB-D Face Recognition

RGB-D face recognition focuses on recognizing faces when they are captured in the form of a color image (RGB) along with a corresponding depth map (D). The color image and depth map pair is called as one RGB-D image. The methods of acquisition of the depth map varies from device to device but the Microsoft Kinect device uses an infrared laser projector combined with a monochrome CMOS sensor. Two public databases [21], [22] containing RGB-D face images captured using the Kinect sesnor are used in this research. Given a RGB-D face image, registration of the faces is performed to increase the correspondence between the pixels of the depth map and the color image. Then, face detection is performed using the color face image and the same bounding box is used for the depth map as well. The color face image and the depth map act as independent feature sources and features can be extracted from both. In this research, we have used TPBLB and HOG as features for RGB-D faces. Each subject's gallery images are used to train the proposed KGSRC algorithm and learn the corresponding sparse representation. Given a probe image, features are extracted from both the color image and the depth map separately and matched with the gallery using the KGSRC algorithm and a ranked list of subjects is obtained by sorting them by ascending order of the residual error.

C. Multi-modal Biometric Recognition

In multi-modal biometric recognition, multiple biometric modalities are used to determine the identity of a given individual. The most popular biometric modalities are face, fingerprint, and iris. In this scenario, each modality acts as an independent feature source. Appropriate pre-processing is applied to the image from each modality and then as shown in Fig. 2c features are extracted using a feature extractor relevant to the modality (e.g. NBIS and Verifinger features for fingerprints). Then, the corresponding features are used for training the KGSRC algorithm. A given probe image may be identified on the basis of any combination of modalities (e.g. only face, only iris, or face and iris or all the modalities) since the algorithm can inherently handle the missing data problem. The residual error is computed for each subject and a ranked list of candidate matches is produced based on assigning the highest rank to the subject with the least residual error.

IV. Experimental Results

The following subsections present experimental results pertaining to each of the case studies including database and protocol details for each of the problems pertaining to each of the problems:

A. Cross-distance Face Recognition in Surveillance

The performance of the proposed algorithm is evaluated on the publicly available SCFace database [23]. It contains 4,160 human face images for 130 subjects in both visible and near-infrared spectrums. The images are captured in uncontrolled indoor environments using five video surveillance cameras of varying quality and at variable distances. The database

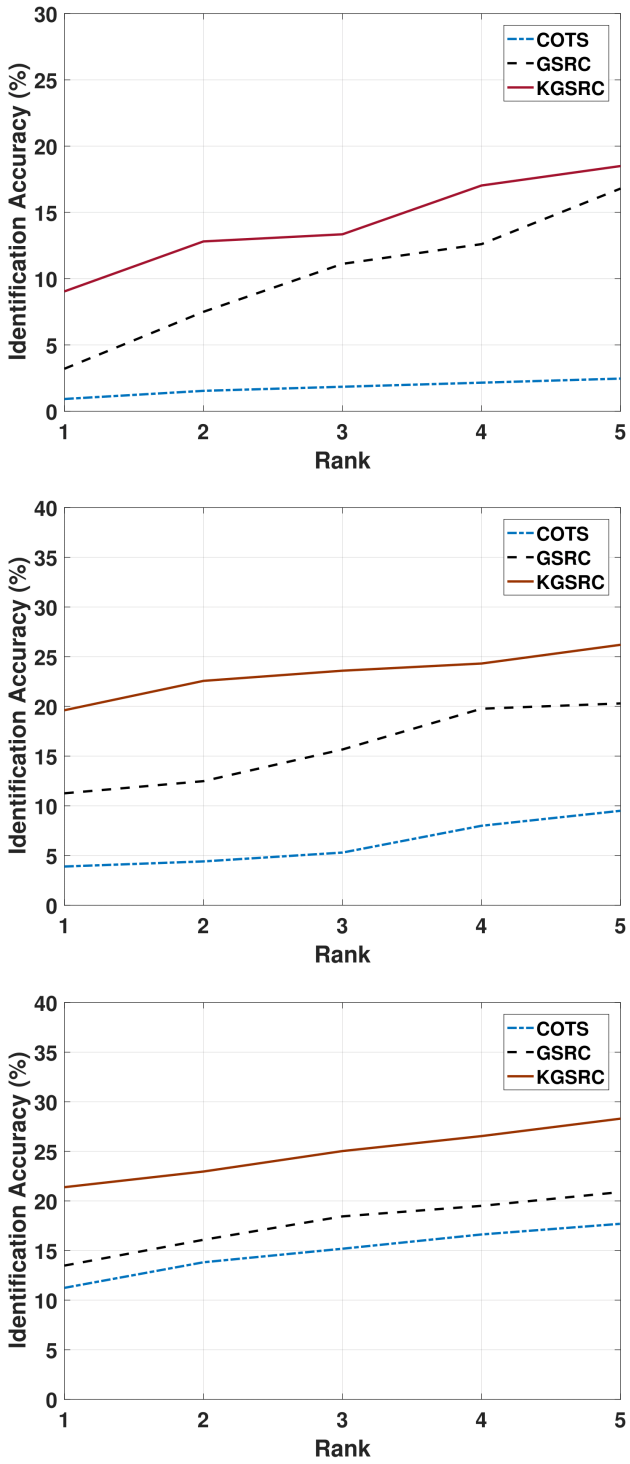
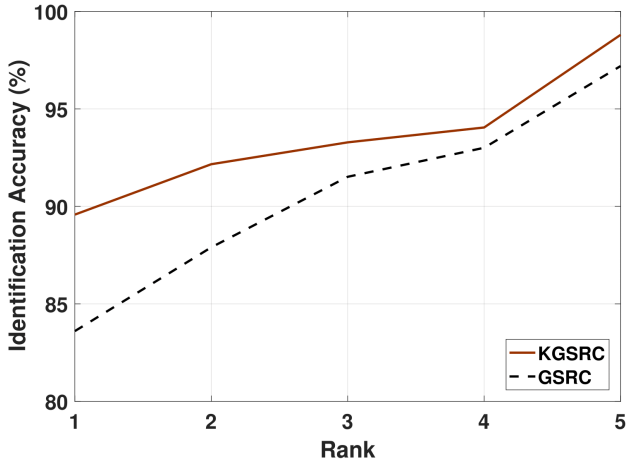
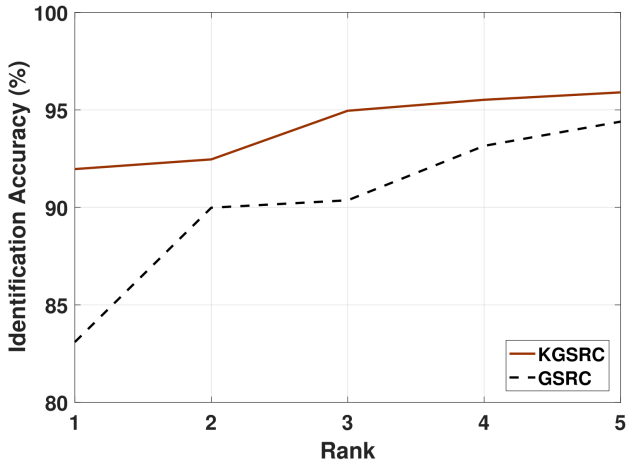


Fig. 3: Identification results on the SCFace database at three different distances of the probe image: (a) Distance 1 = 4.2 metres, (b) Distance 2 = 2.6 metres, and (c) Distance 3 = 1.0 metres. The matching is performed between a high distance probe and a low distance (high resolution) gallery. We observe that the KGSRC algorithm outperforms both the COTS and the GSRC algorithm. For both the KGSRC and GSRC algorithms, the best feature combination is showcased in the CMC.



(a) Identification results on the EURECOM [21] database.



(b) Identification results on the IIIT-D RGB-D [22] database.

Fig. 4: Identification results on the two RGB-D databases when compared to GSRC. For both the KGSRC and GSRC algorithms, the best feature combination is showcased in the CMC. We see that kernelization using the proposed algorithm improves the performance.

also provides five good quality close-distance face images per subject for the purpose of enrollment/gallery data. These mugshots are captured using a high-quality photo camera. In order to conduct our experiments, we utilize the most frontal three good quality faces per subject as gallery. Then, we perform three experiments for each of the three different distances in the database: distance 1 is furthest from the sensor, whereas distance 3 is the closest to the sensor. All the images available for a subject for each distance are used as probe. Therefore, there are a total of 390 gallery and 650 probe images per distance. Distances 1, 2, and 3, denote distances of 4.20, 2.60, and 1.00 metres from the camera, respectively. Other capture conditions in the database are also reflective of the real-world scenario, i.e., the camera is placed slightly above the subject's head. The database encompasses a highly challenging recognition scenario. We compare the

	Algorithm	Identification Accuracy (%)	
		EURECOM	IIIT-D
Existing	Goswami *et al.* ∗[22]	98.5 ± 1.6	95.3 ± 1.7
	3D-PCA ∗	94.1 ± 2.7	83.4 ± 2.1
	FPLBP ∗	94.3 ± 1.4	85.0 ± 0.7
	HOG ∗	89.5 ± 0.8	75.1 ± 0.7
	SIFT ∗	83.8 ± 2.1	50.1 ± 1.4
	Sparse ∗	84.8 ± 1.7	87.2 ± 1.9
	HOG + GSRC	95.4 ± 1.7	89.9 ± 2.8
	TPLBP + GSRC	97.2 ± 1.2	94.4 ± 1.5
Proposed	HOG + KGSRC	96.7 ± 1.5	91.8 ± 2.8
	TPLBP + KGSRC	**98.8 ± 1.2**	**95.9 ± 1.6**

TABLE II: Identification results at rank 5 on both EURECOM and IIIT-D RGB-D databases. We use the existing protocol and partitions [22] to enable direct comparison. ∗ Results have been taken from [22].

performance of the proposed algorithm to a state-of-the-art Commercial Off the Shelf (COTS) face recognition system, as well as with traditional GSRC approach. The results, in the form of rank 5 identification accuracy and CMC curves, are presented in Table I and Fig. 3, respectively.

We observe that the proposed KGSRC algorithm offers substantial improvements over the COTS algorithm, achieving a best case performance of 18.5%, 26.2%, and 28.3% rank 5 identification accuracy at distances 1, 2, and 3, respectively. In comparison, the COTS approach is only able to achieve 2.5%, 9.5%, and 17.7% accuracy at distances 1, 2, and 3 respectively using the same protocol. As opposed to the non-kernelized version (GSRC), the proposed KGSRC algorithm offers improvement in the range of 1.5% - 7.4% for different distances. We also observe that a total of 11, 38, and 48 probe images that are misclassified by the GSRC are identified correctly by the proposed algorithm at distances 1, 2, and 3, respectively.

B. RGB-D Face Recognition

The performance of the proposed algorithm is evaluated on two publicly available RGB-D databases: the EURECOM database [21] and the IIIT-D RGB-D database [22]. First, we present a brief overview of each of these databases and explain the experimental protocol. Next, we present identification results on these databases using the proposed algorithm and also present a comparison with existing state-of-the-art algorithms on these databases.

The EURECOM database [21] consists of RGB-D images of 52 individuals captured in two different sessions. The database consists of variations in illumination and expression, and consists of a total of 936 images. In order to compare results with the state-of-the-art algorithms, we follow the existing protocol [22]. The entirety of the data is partitioned such that data for 31 subjects is in the testing partition and the remaining 21 subjects are part of the training partition. This partitioning is performed five times for cross validation. We report the mean accuracies at each rank in the form of a CMC curve in Fig. 4a. Mean accuracy and standard deviation at rank 5 are presented in Table II.

Algorithm	Features	Rank 5 Identification Accuracy (%)		
		Distance 1	Distance 2	Distance 3
GSRC	HOG + FPLBP	16.8	17.4	12.5
	HOG + TPLBP	15.7	19.7	20.9
	TPLBP + FPLBP	13.7	13.8	15.1
	HOG + FPLBP + TPLBP	16.7	20.3	17.2
COTS		2.5	9.5	17.7
Proposed KGSRC	HOG + FPLBP	18.3	**26.2**	27.5
	HOG + TPLBP	12.9	23.8	**28.3**
	TPLBP + FPLBP	11.1	10.0	15.1
	HOG + FPLBP + TPLBP	**18.5**	25.2	27.5

TABLE I: Identification results at rank 5 on the SCFace database. Distance 1 is the furthest and distance 3 is the closest from the camera.

The IIIT-D Kinect RGB-D face database [22] consists of RGB-D images of 106 individuals (4605 images), captured in two different sessions. Similar to the case of the EURECOM database, existing protocol [22] is used to facilitate comparative results. Each partition contains training data pertaining to 42 subjects and testing data pertaining to 64 subjects. We report the mean accuracies at each rank in the form of a CMC curve in Fig. 4b. Mean accuracy and standard deviation at rank 5 are presented in Table II. Along with demonstrating the performance of the proposed algorithm with different feature extractors on the two databases, we have compared the performance with existing state-of-the-art algorithms on the two databases. These algorithms are Goswami et al. [22], 3D-Principal Component Analysis (3D-PCA), Four patch LBP (FPLBP), Histogram of Oriented Gradients (HOG), SIFT, and sparse classification.

On both the databases, the proposed KGSRC algorithm performs the best, achieving 98.8% and 95.9% rank 5 identification accuracies on the EURECOM and IIIT-D RGB-D databases, respectively. It outperforms its non-kernel counterpart by 1.6% on the IIIT-D database and 1.5% on the EURECOM database. It is also able to outperform the previous best performing algorithm by 0.3% and 0.6%, furthering the state-of-the-art on the two databases. We also observe that the KGSRC algorithm performs consistently better with TPLBP than with HOG. In the current implementation, HOG is a local feature descriptor of a smaller size (256) whereas the TPLBP descriptor provides a larger feature representation of size 3,072. This can be attributed to the nature of sparse classification algorithms that perform better if over-complete representations are provided as input.

C. Multi-modal Biometric Recognition

The WVU multi-modal database [25] consists of data pertaining to multiple biometric modalities for 270 individuals. In this research, we focus on three popular biometric modalities: iris, face, and fingerprint. Since not every modality is available for all individuals, the database incorporates the missing data problem. 231, 270, and 316 subjects have data available for the iris, fingerprint, and face modalities, respectively. Data pertaining to 108 subjects (40%) is utilized for training and data for the remaining 162 subjects (60%) is utilized for testing. Three images are used as gallery and training and the remaining images are probes.

Modality	**Features**		**WVU**
Face	Individual	UCLBP	75.4
		SURF	79.1
	Fusion	SRC	82.3
		GSRC	83.7
		KGSRC	**85.2**
Iris	Individual	Vasir	85.0
		LPG	90.5
	Fusion	SRC	92.9
		GSRC	93.5
		KGSRC	**96.1**
Finger	Individual	NBIS	85.9
		VeriFinger	90.7
	Fusion	SRC	92.6
		GSRC	93.1
		KGSRC	**95.8**
Face and Iris	Fusion	SRC	93.9
		GSRC	94.6
		KGSRC	**97.4**
Face and Finger	Fusion	SRC	94.4
		GSRC	95.3
		KGSRC	**97.1**
Iris and Finger	Fusion	SRC	95.6
		GSRC	95.9
		KGSRC	**98.2**
Face, Finger and Iris	Fusion	Sum Rule (score level)	95.0
		Context Switching [24]	95.8
		SRC	95.1
		GSRC	99.1
		KGSRC	**99.7**

TABLE III: Rank-1 identification accuracy (%) with fusion of multiple modalities and multiple features (SRC, GSRC, and KGSRC) on the WVU database.

The results of the evaluation are presented in Table III. We observe that the proposed KGSRC algorithm achieves the best performance among all the evaluated algorithms. We notice an improvement of 0.6% on the final fusion of all three modalities and is able to further improve upon the already near perfect accuracy achieved by the group sparse classifier. Among two modality fusion experiments, we see consistent improvement of about 2-3% when the group sparse classifier already achieves over 95% accuracy, leaving little room for improvement. Therefore, we can assert that the kernelization of the features before matching helps in improving the discriminative capability of the classifier and results in a more robust feature fusion methodology. The algorithm is able to successfully classify samples that are not

separable in the relatively lower dimensional space of the non-kernel group sparse classifier by projecting them to a higher dimensional space using the automatically determined kernel function. We also see experimental support for the efficacy of the proposed automatic kernel and parameter selection metric since the choice of kernel and parameter can greatly influence the success of any kernel based approach.

V. Conclusion

Multi-feature classification has been explored in literature and usually involves combination of feature information at match score or decision level. However, utilizing a classifier which can interpret multiple feature sources inherently and exploit the underlying data structure can improve performance and robustness. In this paper, we present the Kernel Group Sparse Representation Classifier (KGSRC) which is an extension of the GSRC framework with kernelization. Using the training data, we not only learn the coefficients for sparse coding to generate the most discriminative representation of data, but also learn the appropriate kernel function and associated parameter values automatically. Using three case studies on cross-distance face recognition, RGB-D face recognition, and multimodal biometric fusion, we observe that the proposed KGSRC algorithm outperforms existing state-of-the-art results. As a future research direction, we plan to extend the proposed algorithm for video based person recognition with learnt features [26].

Acknowledgment

The authors would like to thank the creators of the SC face database [23] and the EURECOM database [21] for providing the databases. This research is supported from a grant from Ministry of Electronics and Information Technology, India and in part by Infosys Centre of Artificial Intelligence.

References

[1] L. Xu, S. Zheng, and J. Jia, "Unnatural l0 sparse representation for natural image deblurring," in *The IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[2] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, 2013.

[3] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372–386, 2012.

[4] G. Wang, S. Zhang, H. Xie, D. N. Metaxas, and L. Gu, "A homotopy-based sparse representation for fast and accurate shape prior modeling in liver surgical planning," *Medical image analysis*, vol. 19, no. 1, pp. 176–186, 2015.

[5] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 113–126, 2014.

[6] X. T. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," *IEEE Transactions on Image Processing*, vol. 21, no. 10, pp. 4349–4360, 2012.

[7] M. Cheng, C. Wang, and J. Li, "Single-image super-resolution in rgb space via group sparse representation," *IET Image Processing*, vol. 9, no. 6, pp. 461–467, 2015.

[8] W. Shi, C. Chen, F. Jiang, D. Zhao, and W. Shen, "Group-based sparse representation for low lighting image enhancement," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 4082–4086.

[9] G. Goswami, P. Mittal, A. Majumdar, M. Vatsa, and R. Singh, "Group sparse representation based classification for multi-feature multimodal biometrics," *Information Fusion*, 2015.

[10] S. Gao, I. W. H. Tsang, and L. T. Chia, "Sparse representation with kernels," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 423–434, 2013.

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[12] A. Majumdar and R. K. Ward, "Robust classifiers for data reduced via random projections," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 5, pp. 1359–1371, 2010.

[13] ——, "Improved group sparse classifier," *Pattern Recognition Letters*, vol. 31, no. 13, pp. 1959–1964, 2010.

[14] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1873–1879.

[15] A. Majumdar and R. Ward, "Robust classifiers for data reduced via random projections," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 5, pp. 1359–1371, 2010.

[16] X.-T. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3493–3500.

[17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 511–518.

[18] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Real-Life Images workshop at the European Conference on Computer Vision*, 2008.

[19] ——, "Descriptor based methods in the wild," in *Workshop on faces in'real-life'images: Detection, alignment, and recognition*, 2008.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.

[21] R. Min, N. Kose, and J.-L. Dugelay, "Kinectfacedb: A kinect database for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 11, pp. 1534–1548, 2014.

[22] G. Goswami, M. Vatsa, and R. Singh, "RGB-D face recognition with texture and attribute features," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 10, pp. 1629–1640, 2014.

[23] M. Grgic, K. Delac, and S. Grgic, "Scface — surveillance cameras face database," *Multimedia Tools Appl.*, vol. 51, no. 3, pp. 863–879, 2011.

[24] S. Bharadwaj, H. S. Bhatt, R. Singh, M. Vatsa, and A. Noore, "QFuse: Online Learning Framework for Adaptive Biometric System," *Pattern Recognition (accepted with minor revision)*, 2015.

[25] S. Crihalmeanu, A. Ross, S. Schuckers, and L. Hornak, "A Protocol for Multibiometric Data Acquisition, Storage and Dissemination," WVU, Tech. Rep., 2007.

[26] G. Goswami, M. Vatsa, and R. Singh, "Face verification via learned representation on feature-rich video frames," *IEEE Transactions on Information Forensics and Security*, vol. (in-press), 2017.