



Kernel based locality – Sensitive discriminative sparse representation for face recognition

Ben-Bright Benuwa, Benjamin Ghansah*, Ernest K. Ansah

School of Computer Science, Data Link Institute, P.O. Box 2481, Tema, Ghana



ARTICLE INFO

Article history:

Received 12 March 2019

Revised 12 November 2019

Accepted 25 November 2019

Editor: Dr. B. Gyampoh

Keywords:

Face recognition

Sparse representation

Locality information

Locality sensitive adaptor

Kernel learning

Dictionary learning

ABSTRACT

Kernel-based Sparse Representation (SR) has impacted positively on the classification performance in image recognition and has eradicated the problems attributed to the non-linear distribution of face images and its implementation with Dictionary Learning (DL). However, the locality construction of image data containing more discriminative information, which is crucial for classification has not been fully examined by the current Kernel Sparse representation-based approaches. Furthermore, similar coding outcomes between test samples and neighbouring training data, restrained in the kernel space are not being fully realized from the image features with similar image groupings to effectively capture the embedded discriminative information. To handle these issues, we propose a novel DL method, Kernel Locality-Sensitive Discriminative SR (K-LSDSR) for face recognition. In the proposed K-LSDSR, a discriminative loss function for the groupings based on sparse coefficients is introduced into a locality-sensitive DL (LSDL). After solving the optimized dictionary, the sparse coefficients for the testing image feature samples are obtained, and then the classification results for face recognition is realized by reducing the error between the original and reassembled samples. Experimental results have shown that the proposed K-LSDSR significantly improves the performance of face recognition accuracies compared with competing methods and is robust to various diverse environments in image recognition.

© 2019 Published by Elsevier B.V. on behalf of African Institute of Mathematical Sciences / Next Einstein Initiative.

This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Introduction

SR is mostly applied to signal reconstruction [55], experiments have also shown that it performs well with video analysis and image classification [41,52]. Again, the implementation of SR for face recognition in the past decades has received a lot of attention in the area of pattern recognition and Computer Vision. For a specified signal to be thoroughly represented, a good dictionary must be learned from training samples. This shows that the quality of the dictionary is very crucial for efficient SR. The dictionary could be realized either by utilizing all the training samples as the dictionary to code the test samples, as in the case of the locality constrained linear coding [38] or a learned dictionary for SR is adopted for each training sample in the set (e.g. K-means, Singular Value Decomposition (K-SVD) algorithm, Fisher Discriminative Dictionary

* Corresponding author.

E-mail addresses: ben@datalink.edu.gh (B. Ghansah), founder@datalink.edu.gh (E.K. Ansah).

Learning (FDDL)). Training samples are utilized as the dictionary for all the methods that adopts the first strategy. Even though, they exhibit good performance with regards to classification, the dictionary might not be that effective to exemplify the samples well. This is mainly attributed to noisy information that may have accompanied the initial training samples and may not fully capture the Discrimination information embedded in the training samples. The latter grouping is also not appropriate for recognition, since it only ensures that the dictionary is best expressed in the training samples with strict SR. Even though numerous approaches including the LSDL [27,41,43] that incorporated a Locality Regularization Constraint into the structure of the DL which guarantees that the over complete dictionary (OCD) Learned is representative enough to resolve the above stated issues, there was still the need for methods that will help resolve issues of noise, occlusion, aging, variations in resolution, different facial expressions of objects, and illumination changes in universal environments.

More so, the problem with the traditional SR methods is that, they cannot produce the same or similar results when the input features are from the same categorization. The elementary approach to building the dictionary is by exhausting all the training samples which may result in the size of the dictionary being huge, and is subsequently unfavorable to the sparse solver [43]. A lot of techniques such as the Method of Optimal Direction (MOD) [11] which updates all the atoms simultaneously with fixed sparse codes using Least Square Methods and an orthogonal matching pursuit for sparse coding, the K-SVD algorithm [1] for studying a handful sized dictionary for SR from the perspective of the training data have recently been suggested. As has been discussed in [28], in an attempt to resolve this issue, they recommended further iteratively updating the K-SVD (which focuses on only the representational power or efficiency of the dictionary but not its discrimination abilities) trained dictionary grounded on the result of a linear classifier, by realizing a Dictionary that may be effective for classification with representational power. Further studies literarily along the same direction includes [24,25], which used a more complicated objective function in optimizing the dictionary for the training phase so as to gain some discriminative power for the dictionary. There are other discriminative DL methods that use discriminative and reconstructive error to learn the dictionary. A structured based FDDL algorithm was proposed in [48], that improves the performance of pattern classification whose dictionary has a relation with the class labels and further exploited the discriminability in both representative coefficient and its residual in [49]. A discriminative kernel SR based classification approach was introduced in [26], that uses a discriminative dictionary and a residual for an effective classification results. An over complete discriminative DL approach based on K-SVD was also proposed in [54] for an optimal classification results, by introducing a classification error term into the objective function. Recently, the discriminative structured DL with hierarchical group sparsity that reduces the linear predictive error and discriminability of sparse codes (SC) using hierarchical group sparsity was introduced in [45]. The discriminative structured DL for image classification that combines into the objective function, the discriminative properties associated with reconstruction error, characteristic error and the classification error terms was also proposed in [39]. In the same vein of improving the discriminability of dictionary learning algorithms, the structure adaptive dictionary learning for SR based classification was proposed in [6], the learning of dictionary discriminatively for group sparse representation was discussed in [2,33], and discriminative DL for face recognition with low ranked regularization by [20]. However, these methods do not take into consideration the locality structure of data in spite of the fact that they have been effectively applied in the area of image classification. One way to take advantage of data locality and Kernel Sparse Representation based Classification (KSRC) [50], the Locality-constrained Linear Coding for Image Classification (LLC) built on the theories of [47] for image classification was proposed in [38]. LLC could achieve a smaller reconstruction error with multiple code reconstruction and local smoothness with respect to sparsity. More recently, there has been an advancement in sparse DL for video semantic analysis as proposed in [52], a video semantic analysis based on Locality-sensitive Discriminant SR and Weighted KNN (LSDSR-WKNN). This approach gave a better category discrimination on the SR of video semantic concepts as well as robust face recognition. The kernelized locality sensitive adaptor was also discussed in [34], that utilize a group structure information in the training data. In the face of having improved group discrimination on the SR of video semantic models, the LSDSR-WKNN method also unsuccessful in its bid to completely explore the probable information of the given samples.

In recent years, several kernel approaches has been proposed to improve the effectiveness of face recognition algorithms. The Multi kernel learning (locality constraint) for face recognition, that considered local structure of data was discussed in [56]. In this approach, a projection matrix is learned in conjunction with a kernel weight to generate a low dimensional subspace by minimizing the within class and maximizing the between class reconstruction errors. A virtual kernel based sparse dictionary for face recognition is proposed in [12]. This approach was introduced to improve on the computational efficiency of kernel based sparse representation approaches. A self-weighted kernel method for clustering and classification was also proposed in [18]. This approach aid graph based clustering and semi-supervised classification. In [34], a robust face recognition approach based kernelized group sparse representation was engineered. The approach uses group structure information to in the training set and measures the local similarity information existing amongst the training and the test sets in the kernel space. Other approaches to exploit the nonlinear nature of face images, videos and to capture local discriminative information were also discussed in [2,7,8,22,46,52]. Though these approaches yield good face image and video recognition results, with embedded structure information exploited, they are not able to capture effectively and wholly, the discriminative information that could be hidden in the structure of face images for effective recognition results.

Based on the foregoing, this paper seeks to improve the discriminative power of SR features and further exploit the embedded non-linear information of similar image categories by encoding SC of samples from the same category. Thus proposes a kernel based locality – sensitive discriminative sparse representation for face recognition (K-LSDSR), which is more effective and has better numeric stability with the adaptation of the L2-norm. The K-LSDSR algorithm aids the modification

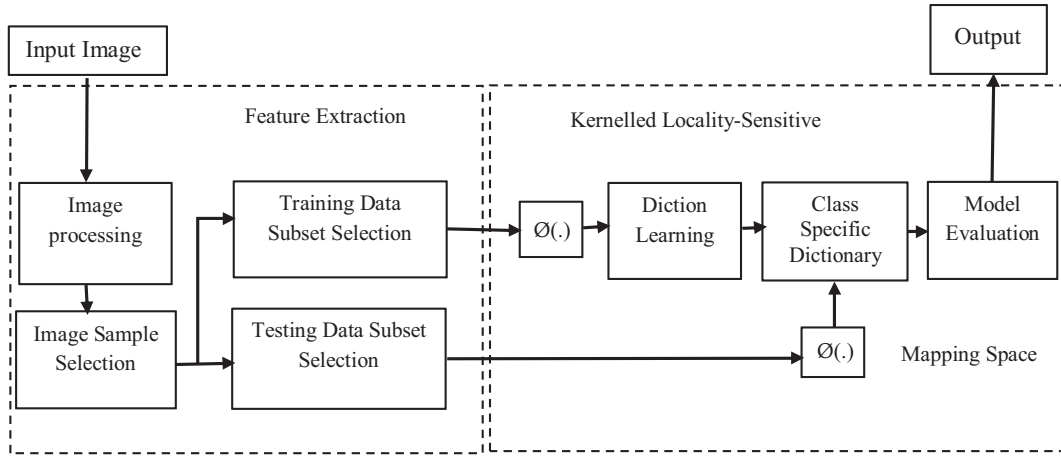


Fig. 1. Framework of the proposed K-LSDSR technique.

of the dictionary adaptively by means of the changes amongst the reconstructed dictionary atoms and the training samples with the kernelled locality adaptor mapped into high dimensional feature space. More so, the implementation of the kernel at both DL and Sparse Coding stages as well as the locality adaptor increases the efficiency of the algorithm for the locality and similarity of the dictionary atoms. The discriminative information of the dictionary can be explored by the proposed technique by introducing a discriminant loss function at the dictionary learning stage. This approach has the propensity in obtaining more discriminative information thereby enhancing the classification ability of sparse representation features. The dictionaries learned could realistically represent facial images more, and thus give a better representation of the samples, associating the sparse coefficients. The framework of the proposed K-LSDSR is depicted by Fig. 1.

The main contributions of this paper are as enumerated below:

- (1) A discriminative loss function is utilized, giving an optimal dictionary for addressing SR issues and optimized sparse coefficients for reconstructing the dictionary, so as to enhance the discriminative classification of image features. This further results in an efficient reconstruction of the dictionary, particularly when the size gets large and enforces the capture of dependencies in image features belonging to the same category.
- (2) The implementation of Kernel into the objective function enables the structure and the non-linear information hidden in the test and training data is best utilized and hence more discriminative SR obtained. More so, it enables efficient reconstruction of samples most especially when the size gets large. This enables higher recognition performance by exploring the nonlinear discriminative information embedded in image data.
- (3) A locality-sensitive discriminative dictionary learning and SR based on kernel sparsity is mainly made for face recognition that embeds image features originally. Consequently the features are sparsely encoded with image features for better preservation of the dictionary, hence an improvement in the accuracy of image recognition concept. This therefore yields an improvement in the accuracy of face recognition.

The remainder of the paper is systematized into the following categorizations; Section 2 details a review of related works, Section 3 discusses the proposed algorithms, Experimental results are presented in Section 4. Finally, Section 5 outlines the main conclusions and recommendations.

Related works

Sparse representation algorithms

Generally, SR for signal reconstruction and all the techniques involved including but not limited to decoding, sampling, compression, and transmission is one of the utmost essential ideologies as regards signal processing deductions. Additionally, a signal sampled could impeccably be reconstructed from an arrangement of samples, if only the rate surpasses two times the maximum frequency of the main signal [3]. It utilizes an OCD to linearly reconstruct a data case for a given dictionary. Assuming that example is emanating from space R^d , and thus every example integrated to a matrix $X \in R^{d \times m}$. Thus for a sample to be closely exemplified by a linear amalgamation of dictionary, D and for the magnitude of the samples being huge as compared with its dimensions in D , then $m > d$, hence D is an OCD. Generally, from the matrix stated above, if $d < m$, and $y \in R^d$ then we define the linear system of equations as $y = Z\beta$.

The sparsest representation solution can be acquired by solving $y = Z\beta$ with the l_0 -norm constraint [9]. Thus expression $y = Z\beta$ could be transformed into the optimization problem as stated in Eq. (1) below:

$$\hat{\beta} = \operatorname{argmin} \|\beta\|_0 \quad \text{s.t. } y = Z\beta \quad (1)$$

Now since the issue of noise cannot be entirely avoided as a result of real data containing some minute amount of noise, the formulation of Eq. (1) is transmuted into Eq. (2) below;

$$y = Z\beta + s \quad (2)$$

where $s \in \mathbb{R}^d$ denotes the noise and is constrained as $\|s\|_2 \leq \varepsilon$. More so, Eq. (2) could also be transformed with the introduction of noise as stated in the optimization problems of equations (3) below:

$$\begin{aligned} \hat{\beta} &= \operatorname{argmin} \|\beta\|_0 \quad \text{s.t.} \quad \|y - Z\beta\|_2^2 \leq \varepsilon \\ \text{Or} \\ \hat{\beta} &= \operatorname{argmin} \|y - Z\beta\|_2^2 \quad \text{s.t.} \quad \|\beta\|_0 \leq \varepsilon \end{aligned} \quad (3)$$

Additionally, the Lagrange multiplier could be adopted and applied on Eq. (3) using λ as the constant. Thus Eq. (3) is transmuted into Eq. (4) below.

$$\hat{\beta} = L(\beta, \lambda) = \operatorname{argmin} \|y - Z\beta\|_2^2 + \lambda \|\beta\|_0 \quad (4)$$

where λ refers to the Lagrange multiplier associated with $\|\beta\|_0$.

The sparse representation of the original sample is denoted as the sparse linear integration of the over-complete dictionary. The objective function of representation is as stated in formula (1). SR remained extensively utilized in computational intelligence, ML, computer vision and pattern recognition, and many more. Implementation SR algorithms encompasses pursuing the sparsest linear groupings of basis functions from an OCD as stated in formula (1) and is known to be essentially NP-hard with numerical instability [9]. In the quest to approximate a desired solution, some greedy algorithms have been proposed. However, the approximated solutions of these algorithms are suboptimal despite being easy and simple to implement [41]. This has led to the convex relaxation of formula (4) recently, by replacing the nonconvex L_0 - norm with the convex L_1 - norm as indicated in formula (5) below;

$$\min_x \|x\|_1 \quad \text{s.t.} \quad a = Dx. \quad (5)$$

According to Eq. (4), since the L_0 - norm minimization could adequately guarantee a distinctive SR solution of x , the SR solution can be solved with the L_1 -norm minimization problem of Eq. (5). In situations where a may have some little dense noise, the following optimization problems referred to as Lasso as discussed in literature with an introduction of a regularization parameter [35].

$$\min_x \|a - Dx\|_2^2 + \lambda \|x\|_1, \quad (6)$$

where the sparsity of x is controlled by the regularization parameter λ . In recent times, a lot of approaches have adopted different techniques to solve the problem of Eq. (6) as discussed in the literature [40]. In order to improve the over-fitting issues, the L_2 - norm regularization was introduced. However, the solutions of both l_0 -norm and L_1 -norm minimizations are strictly sparse while that of L_2 -norm, is not but limitedly-sparse and thus believe inhibits the qualities of discriminability (Z. Zhang, Wang, Zhu, Liu, & Chen). Thus in our approach, we adopted the L_2 -normalization to make our approach more stable.

Kernel sparse representation

Mapping the features to a higher dimensional space may be helpful for more efficient classification. Suppose $\vartheta(\cdot)$ is a mapping function from \mathbb{R}^N to a higher dimensional feature space. To avoid the explicit high-dimensional mapping procedure, mercer kernels could be helpful. A mercer kernel can be written as $K(x, y) = \langle \vartheta(x), \vartheta(y) \rangle$. The dot product of the two high dimensional mapping $\vartheta(x)$ and $\vartheta(y)$ can be directly replaced by the function value of (x, y) . The common mercer kernels include the linear kernel $K(x, y) = \langle x, y \rangle$, which equals to non-mapping; the Gaussian kernels $K(x, y) = \exp(-\frac{\|x-y\|_2^2}{c})$; the polynomial kernels $K(x, y) = (\langle x, y \rangle + c)^d$ (c and d are parameters) and the sigmoid kernels $K(x, y) = \tanh(a(x^T y) + r)$ (a and r are parameters) [10,16,21]. Any algorithms involved dot products can be replaced by a chosen kernel without explicit mapping to a higher dimensional space. Multiple Kernel Learning for SR based for classification taking advantage of non-linear kernel SRC for its efficient representation in high dimensional feature space was proposed in [32]. In this approach the SC are firstly updated while the kernel collaborative coefficients are fixed and vice versa by its implementation of the two methods training approaches to learn the kernel weights and the sparse code and are repeated until a stall criterion is met. Supposed the class of a training sample was predicted by the use of a training matrix \mathbf{Y} then the SC will always be zero throughout with the exception of a single 1 corresponding to the training sample under consideration and might not help in computing the optimal kernel. Hence to avoid the aforementioned, the associated column of \mathbf{Y} is set to 0 before determining the sparse code. This deduction is done as follows:

$$\hat{x}_i = \operatorname{argmin}_x k(x_i; \tilde{Y}_i, y_i) + \lambda \|x_i\|_1, \quad (7)$$

where $\tilde{Y}_i = [y_1, \dots, y_{i-1}, 0, y_{i+1}, \dots, y_N]$, and 0 is a D -dimensional with zeros as entries, x_i is the sparse vector and \mathbf{X} is the matrix composed of $\mathbf{X} = [X_1, \dots, X_N] \in \mathbb{R}^{N \times N}$. The linear amalgamation of M base kernels is as stated below in learning

the optimal kernel k

$$k(y_i, y_j) = \sum_{m=1}^M \eta_m k_m(y_i, y_j), \tag{8}$$

where η_m is the weight of the m th base kernel and $\sum_{m=1}^M \eta_m k_m = 1$. Impressive result have been reported in [32] that learns the optimum kernel function k and resulted in a extreme training classification while sidestepping the problem of over-fitting.

Locality-sensitive sparse representation

Test examples in SR are usually denoted by the atoms of the dictionary that may not be neighbouring it for signal reconstructions. This may render it incongruous for holding data locality, and thus primes to unsuitable recognition rates. In the LSDL approach, the locality constraint was introduced into both the sparse coding and the DL stages, hence a well representation of the test sample by the neighbouring dictionary atoms. The optimization function below denotes the formulation of the LSDL approach:

$$\begin{aligned} \min_{D, X} & || A - DX ||_F^2 + \lambda \sum_{i=1}^N || p_i \odot x_i ||_2^2, \\ \text{s.t.} & \mathbf{1}^T x_i = 1 \quad \forall i = 1, \dots, N \end{aligned} \tag{9}$$

where $p_i \in \mathbb{R}^{K \times 1}$ is locality adaptor, \odot is the elementwise multiplicative symbol and the regularization parameter that measures the locality constraint is λ . The SR centered classification is boosted by the introduction of a discriminative loss function to the structure of the proposed approach, by utilizing class information since the LSDL approach does not take into account class information, thus to advance the exactitude of image data representation or video analysis. In image feature based sparse representation, image features from the same image category may not be able to achieve the same coding results. It assumed however that image samples from the same category could be encoded as having similar sparse coefficients for face recognition. This is done to enhance the power of discrimination for sparse based classification approaches. To address the above challenge, a locality-sensitive discriminative sparse representative (LSDSR) algorithm was proposed in [53] based on Eq. (9). The approach incorporated a discriminative loss function based on sparse coefficients, into the objective function of the LSDL technique, which achieve an optimal dictionary with the power of discrimination enhanced. A discriminating loss function using class information to enhance the SR based classification is designed in this study, since class information is not cogitated by the LSDL technique. This was done in order to improve the accuracy of video analysis. It has been reported in [37] that LLC give promising classification results if resultant sparse coefficients are utilized as features for testing and training. A lot of locality preserving techniques such as discussed in [23,42], for preserving the local structure of data in dictionary learning that results in a close form solution during the learning process and an integration of the of the penalty term into the structure of the DL has successfully been implemented for locality DL. In [19], a dual layer locality preserving technique based on KSVD was proposed for object recognition. Despite the successes chalked by the approach in improving the discriminability of the learned dictionary, it is challenged with the issues of dimensionality reduction and fails to exploit fully, the discriminative information essential for classification. Besides, the discrimination ability of the learned dictionary is a key to an effective classification results and aims at classifying an input query sample correctly. This has motivated us to propose efficient classification techniques by adopting kernel learning and locality sensitive sparsity coding of the sparse coefficients for the purposes of enhancing the power of discrimination for video semantic analysis and classification of video semantic contents.

Proposed method

The proposed algorithm for K-LSDSR is detailed in this section by incorporating kernel and a discriminative lost function centred on SR into the objective function of locality-sensitive discriminative SR approach. The proposed approach could realize optimum dictionary and further augment the power of discriminability of SC. Image Features from representation might not be capable of achieving the same coding results, hence an assumption is made that samples are coded as related sparse coefficients in the SR based data recognition. This is done with the objective of improving the discrimination capabilities of SR based classifiers and also dealing with the capture of hidden nonlinear information. As a result, data similarity in the kernel feature space is measured, by exploring the nonlinear relationship of data better.

Assume that $A = [a_1, a_2, \dots, a_n] = [A_1, A_2, \dots, A_k] \in \mathbb{R}^{d \times n}$ denotes n training samples from k classes where the column vector a_i is the sample i ($i = 1, \dots, n$), and the submatrix A_j consisting of column vectors from class j ($j = 1, \dots, k$). If there are M atoms each in the dictionary $D = [D_1, D_2, \dots, D_M] \in \mathbb{R}^{d \times M}$, an OCD for SR with ($M \leq n$), then coding coefficients of A over D could be denoted by $X = [x_1, x_2, \dots, x_L] \in \mathbb{R}^{M \times L}$. Now considering mapping features into a high dimensional feature space which is very imperative for classification, our K-LDSR framework is modelled as:

$$\begin{aligned} \min_{D, X} & || K(A) - K(D)X ||_F^2 + \lambda \frac{N}{2} || K(P_i) \odot x_i ||_2^2, \\ \text{s.t.} & \mathbf{1}^T x_i = 1 \quad \forall i = 1, \dots, N \end{aligned} \tag{10}$$

where we have the samples A , the dictionary D and the locality-sensitive adaptor P_i being mapped into a high dimensional feature space with the function $\vartheta(\cdot)$, then $K(A) = \vartheta(A) \in R^{q \times n}$, $K(D) = \vartheta(D) \in R^{q \times M}$, and $K(P_i) = \vartheta(P_i) \in R^{K \times 1}$ could be substituted into Eq. (10). Hence our K-LSDSR framework is reformulated as stated below;

$$\begin{aligned} \min_{D, X} & \|\vartheta(A) - \vartheta(D)X\|_F^2 + \lambda \sum_{i=1}^N \|\vartheta(P_i) \odot x_i\|_2^2. \\ \text{s.t.} & 1^T x_i = 1 \quad \forall i = 1, \dots, N \end{aligned} \quad (11)$$

Here the locality-sensitive adaptor P_i utilizes the l2-norm function and its k th element in P_i is denoted by $P_{ij} = \|\vartheta a_i - \vartheta d_j\|$ and the symbol \odot represents element-wise multiplication, the shift variant constant $1^T x_i$ enforces the coding result of x to stay the same although the source of the data coordinate may be budgeted as indicated in [51] and the regularization parameter controlling the reconstruction error and the sparsity is λ . The dictionary in the high dimensional feature space, $\vartheta(D) = \vartheta(A)V$ according to [31], where $V \in R^{n \times m}$ is the representative matrix.

Considering cases where sample features from similar categories could be having comparable SC, a discriminative loss function built on sparse coefficients for the purposes of augmenting the discrimination capabilities of input signals or samples in SR. More so, a discriminative loss function is introduced. This is aimed at minimizing the within-class scatter of SC and concurrently maximizing the between-class scatter of SC based on the Fisher criterion. In addition, samples from same category are accordingly compressed and the ones from different categories are detached. The discriminative loss function centred on sparse coefficients is explained below:

$$g(x_i) = \sum_{i=1}^k (\|x_i - M_i\|_2^2 + \eta \|x_i\|_2^2), \quad (12)$$

where $\sum_{i=1}^k (\|x_i - M_i\|_2^2 + \eta \|x_i\|_2^2)$ is the within class-similar term, M_i is the mean vector of the representation coefficient x_i belonging to the class i . The term $\|x_i\|_2^2$ combined with $\|X\|_2^2$ could make Eq. (11) more stable based on theorem of [57]. With η set to 1 for simplicity, Eq. (11) can therefore be reformulated as

$$\begin{aligned} \min_{V, X} & \|\vartheta(A) - \vartheta(A)VX\|_F^2 + \lambda_1 \sum_{i=1}^N \|\vartheta(P_i) \odot x_i\|_2^2 + \lambda_2 \sum_{i=1}^N g(x_i). \\ \text{s.t.} & 1^T x_i = 1 \quad \forall i = 1, \dots, N \end{aligned} \quad (13)$$

The proposed method as stated in Eq. (13) is implemented by enforcing data locality in high dimensional feature space. The locality adaptor is utilized to determine the kernel distance amongst the test sample $\vartheta(A)$ and every column of D . Note that $D = [D_1, D_2, \dots, D_M] = [\vartheta(a_1), \vartheta(a_2), \dots, \vartheta(a_n)]$ represents all the training samples in the kernel feature space. The vector P , the dissimilarity vector in Eq. (13) is implemented to suppress the corresponding weight and also penalizes the distance existing amongst the test sample and every training sample in the kernel feature space. Furthermore, it should however be made know that the resulting coefficients in our K-LSDSR formulation may not be fully sparse with regards to l2 - norm, but is seen as sparse because the representation solutions only have few significant values with most being zero. The test samples and their neighbouring training samples in kernel space are encoded when the problem of Eq. (12) is being minimized and the resulting coefficients X still sparsified because as P_{ij} gets large, x_{ij} shrink to be zero. Most coefficients therefore get to zero, with just some few having significant values.

The proposed K-LSDSR approach integrates kernel, data locality and sparsity in obtaining sparse coefficients, hence the capability of learning discriminative sparse representation coefficients for classification. A kernelled exponential locality adaptor was implemented in our proposed method as explained in Eq. (14) below:

$$P_{ij} = \sqrt{\exp\left(\frac{d_k(a_i, a_j)}{\sigma}\right)} \quad (14)$$

where σ is a constant and $d_k(a_i, a_j)$ is the kernel Euclidean distance induced by a kernel space, k defined as

$$d_k(a_i, a_j) = \sqrt{\vartheta(a_i) - \vartheta(a_j), \vartheta(a_i) - \vartheta(a_j)} = \sqrt{k(a_i, a_i) - 2k(a_i, a_j) + k(a_j, a_j)} \quad (15)$$

The P_{ij} increases exponentially with an increase $d_k(a_i, a_j)$ hence yields a huge P_{ij} when the samples a_i and a_j are far apart, hence a smaller a_{ij} .

Optimization

The objective function in Eq. (13) could be sub-divided into two sub-divisions by updating X with D fixed and updating D with X fixed since it is not convex as a whole. To find the desired dictionary D and the sparse coefficient X , an alternative optimization is implemented iteratively adopting the theories of [5,13,17,48].

Updating X with V fixed, sparse coding stage

When the coefficient matrix X is being updated with the representative matrix V being fixed, the objective function of Eq. (13) is reduced to the coding problem below:

$$\langle X \rangle = \min_X || \vartheta(A) - \vartheta(A) VX ||_F^2 + \lambda_1 \sum_{i=1}^N ||\vartheta(P_i) \ominus x_i ||_2^2 + \lambda_2 \sum_{i=1}^N g(x_i). \tag{16}$$

The X in Eq. (16) could be addressed on class bases. Thus x_i corresponding to class i could be derived as:

$$\langle x_i \rangle = \min_{x_i} || \vartheta(A_i) - \vartheta(A) V x_i ||_2^2 + \lambda_1 \sum_{i=1}^N ||\vartheta(P_i) \ominus x_i ||_2^2 + \lambda_2 \sum_{i=1}^N g(x_i). \tag{17}$$

We let

$$J(x_i) = \min_{x_i} || \vartheta(A_i) - \vartheta(A) V x_i ||_2^2 + \lambda_1 \sum_{i=1}^N ||\vartheta(P_i) \ominus x_i ||_2^2$$

and simplify it as

$$= \min_{x_i} tr(x_i^T W x_i - 2 x_i^T Q + S), \tag{18}$$

where $W = V^T k(A, A) V + k(a_i, a_i)$, $Q = V^T k(A_i, A) + k(a_i, a_j)$, and $S = k(A_i, A_i) + k(a_j, a_j)$.

Using the analysis in [14] Eq. (18) is equivalent to Eq. (19) as stated below:

$$J(x_i) = \min_{x_i} || \tilde{A}_i - \tilde{D} x_i ||_F^2, \tag{19}$$

where $\tilde{A}_i = \Sigma^{-1/2} U^T Q$, $\tilde{D} = \Sigma^{1/2} U^T$ and SVD of W is $U \Sigma U^T$. Thence Eq. (17) could be rewritten as:

$$\langle x_i \rangle = \min_{x_i} || \tilde{A}_i - \tilde{D} x_i ||_2^2 + \lambda_1 \sum_{i=1}^N ||x_i ||_2^2 + \lambda_2 \sum_{i=1}^N g(x_i). \tag{20}$$

The convex problem in Eq. (20) above could be solved by the iterative projection method [29] and each column of X could be updated as $x_i = (i = 1, \dots, k)$.

Updating V with the sparse coefficient matrix X fixed, the dictionary update stage

During the dictionary update stage, the objective function in Eq. (13) is reduced to Eq. (16) as stated below:

$$V = \min_V || \vartheta(A) - \vartheta(A) VX ||_F^2 + \lambda_1 \sum_{i=1}^N ||\vartheta(P_i) \ominus x_i ||_2^2. \tag{21}$$

Let $\Gamma(V) = \min_V || \vartheta(A) - \vartheta(A) VX ||_F^2 + \lambda_1 \sum_{i=1}^N ||\vartheta(P_i) \ominus x_i ||_2^2$ then

$$\Gamma(V) = tr(\vartheta(A)(I_n - VX - X^T V^T + V X X^T V^T) \vartheta(A) V) k(a_i, a_j)$$

We take derivative of $\Gamma(V)$ over V and equate it to zero ($\nabla \Gamma(V) = 0$), the update of the equation is as stated in Eq. (22) below:

$$V = X^T (X X^T)^{-1} \tag{22}$$

Classification scheme

For a given test sample y, the representation coefficients on dictionary D is given as:

$$x = \arg \min_x ||y - Dx ||_2^2 + \lambda ||x ||_2. \tag{23}$$

The test sample can finally be classified after obtaining V, hence y and D in Eq. (23) could be replaced with $\phi(y) \in R^d$ and $\phi(A) V$ respectively as stated in Eq. (24) below:

$$x = \arg \min_x ||\phi(y) - \phi(A) V x ||_2^2 + \lambda ||x ||_2, \tag{24}$$

where λ is a scalar constant. Let $T(x) = \min_x ||\phi(y) - \phi(A) V x ||_2^2$ and $T(x)$ is easily simplified as

$$\min_x tr(x^T W x - 2 x^T Q + S), \tag{25}$$

where $W = V^T k(A, A) V$, $Q = V^T k(y, A)$, and $S = k(y, y)$.

Using the analysis in [14], Eq. (25) is equivalent to Eq. (26) as stated below:

$$T(x) = \min_{x_i} || \tilde{y} - \tilde{D} x_i ||_2^2 \tag{26}$$

Algorithm 1

Kernelled locality –sensitive discriminative sparse representation.

Input: Training samples X , Initial dictionary D and parameters, λ_1 and λ_2 Output: The learned dictionary D

-
- (a) Initialize X and V . Based on Eq. (28), the representative coefficient X and the representative matrix V could be initialized using the kernel KSVD algorithm [36].
- (b) Optimize the solution by updating the sparse coding matrix X and the representation matrix V combined with Eq. (28). Update X by fixing V and updating each $x_i = (i = 1, \dots, k)$ according to Eq. (20) on class bases using the IPM [29].
- (c) Fix X and update V using Eq. (22)
- (d) Iteratively return to step (b) until convergence.
- (e) Get the representative coefficient x of y based on Eq. (26) combined with (28)
- (f) Identify y : classify y using D and X based on Eq. (27)
-

where $\tilde{y} = \Sigma^{-1/2}U^TQ$, $\tilde{D} = \Sigma^{1/2}U^T$ and SVD of W is $U\Sigma U^T$. Then Eq. (24) could be rewritten as:

$$X = \min_x \left\| \tilde{y} - \tilde{D}X \right\|_2^2 + \lambda_1 \|X\|_2^2. \quad (27)$$

We get the sparse coding efficient associated with test sample y by solving Eq. (26) and (27), after which the residual $r(y) = \|y - D_j x_j\|_2^2$ is calculated. Where x_j is the sparse coding coefficient on the i^{th} class and the test sample y is lastly classified into the class that has the least residual (Algorithm 1).

$$K(x, y) = \sum_{j=1}^J P_j K_j(x, y) \quad (28)$$

$$P_j \geq 0, \quad \sum_{j=1}^J P_j = 1$$

Eq. (28) above indicates J features for each sample and could be infused by a kernel combination of J kernels where sub-kernel K_j , corresponds to feature j . Hence Eqs. (20) and (26) are comparable to (28) for the proposed K-LSDSR algorithm.

Experimental results and analysis*Database selection and face recognition*

Experiments are performed on natural images to demonstrate the efficient performance of our proposed K-LSDSR method is discussed in this section. We scrutinize and relate the performance of our DL algorithm with most distinctive approaches such as the K-SVD [1], SRC [15], FDDL [48], Kernel Locality-Sensitive Group Sparse Representation (KLSGSR) [34], Kernel Sparse Representation based Classification (KSRC), Locality Sensitive Discriminative Sparse Representation (LSDSR) [52] and LSDL [41] algorithms. Out of the many collections of datasets for objects categorization and classification, four of them are being used in our experimental set up or evaluations. The public datasets being used are the ORL face dataset for face recognition, Yale B face dataset for face recognition, FERET face dataset for face recognition, and AR face dataset for face recognition and gender discrimination. Despite the better recognition performances chalked by deep machine learning with its implementation on datasets recently, the proposed approach did not adopt it simply because of its limitations with regards to its super hardware and large training samples requirements [4,44]. The proposed approach considers the two different locality adaptors (L2 norm and Exponential) on face recognitions.

Parameter selection

Several parameters including λ , the classification parameter, λ_1 , a positive weighing parameter for the locality sensitive constraint and λ_2 for the discriminative constraint were utilized by the proposed K-LSDSR and the classification error scheme, with varied values assigned depending on the dataset being used. More so, the classification parameter, λ made a little or no effect on the output of the experimental results hence was set to 0.01 in all experiments. The recognition rates of the proposed K-LSDSR approach (for the optimization phase) with varying values of λ_1 and λ_2 on AR face dataset are as depicted by Fig. 2. It could be seen from the Fig. 2 that, the optimum results were obtained for recognition when λ_1 is 0.01 and λ_2 is 0.6 respectively with λ_2 and λ_1 values being fixed for each dataset by fivefold cross validation. Throughout all experiments, the Principal Component Analysis (PCA) was used to reduce feature dimensions before classification. And the parameter values used for the various experiments for each dataset are shown in Table 1.

Table 1 indicates the various parameter selections with their respective datasets. Based on the aforementioned values, an experiment was conducted to validate the performance of our proposed algorithms and the findings are detailed below.

Experimental results

There are a lot of datasets utilized for classification of objects. Out of these collections, three of them are being considered in our experiments. The experimental outcomes are realized by an N-fold cross-validation specifically 5-fold. The training

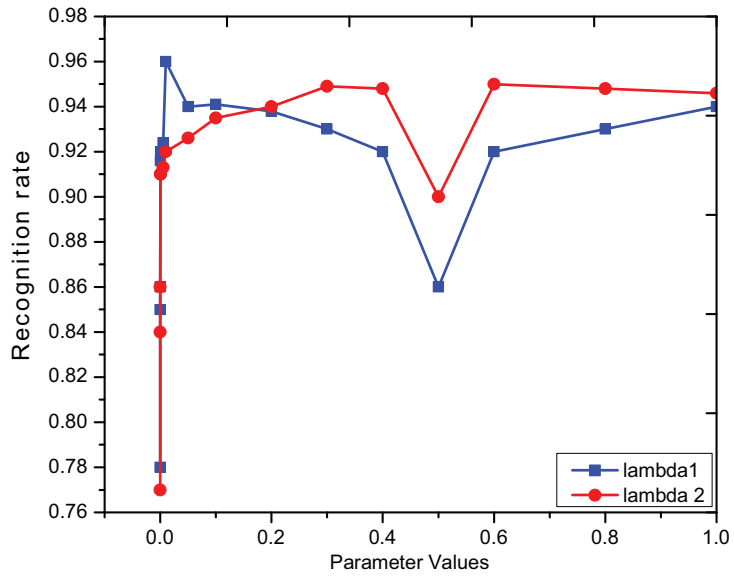


Fig. 2. Recognition rate Vrs variations of parameter values.



Fig. 3. (a). Sample of face images for training as indicated in the first row and that of the testing second row for ORL face database. (b) Sample of face images for training and testing indicated respectively in the first and second rows for ORL face database with 10% corruption.

Table 1
Parameter values implemented on the various data sets.

Dataset	Parameter values used	
	λ_1	λ_2
Yale B face	0.6	0.01
AR face	0.05	0.005
ORL	0.05	0.01
FERET	0.01	0.4

Table 2
Average recognition rate (%) of various algorithms with varying corrupted images on ORL database randomly averaging over five splits.

Occlusions	0	10	20	30	40	50
KSRC	90.4	79.9	63.0	53.9	38.4	26.9
KSVD	93.2	90.4	81.6	76.5	67.9	57.8
FDDL	96.7	86.9	74.7	62.8	50.2	37.3
LSDL	91.6	83.7	70.5	62.6	47.6	68.4
LSDSR	96.1	92.9	87.0	77.4	70.1	69.9
KLSGSR	96.0	93.4	87.2	77.9	69.8	70.9
K-LSDSR	96.3	95.1	87.6	78.3	70.9	73.8

Table 3
The outcome of the face recognition (%) of the various methods on the Yale B database.

Dim	50	80	120	200	300
KSRC	93.6	94.7	95.3	96.5	97.9
K-SVD	95.2	95.9	96.5	97.0	97.6
LSDL	95.6	96.1	96.7	97.2	97.9
FDDL	96.2	96.6	97.3	97.4	98.1
LSDSR	96.4	96.7	97.3	97.5	97.9
KLSGSR	96.4	97.0	97.6	97.7	98.1
K-DLSR	96.8	97.3	97.7	97.8	98.5

samples are randomly selected from each class and the rest utilized as testing samples which are later interchanged to yield better results in all cases. The first to be analysed is **ORL face repository** which contains four hundred (400) face images gotten from forty (40) different entities with each comprising of ten (10) facial images [30]. More so, some of these images were snapped at diverse periods with different facial expressions and specifics, under varying light intensity, images captured homogeneously against dark and grey background with the entities in an upright frontal position. Every image was resized using Matlab codes to 32×32 image matrixes to ensure sufficient data in the estimation of Gaussian models and also for covariance matrices to be positive definite. The images were randomly corrupted manually by an unrelated block image at random locations. The recognition accuracies at dissimilar or various occlusion levels are shown in [Table 2](#).

It could be seen from [Table 2](#) that, our algorithm consistently accomplishes the best optimal result under increasing occlusion levels as compared with the other state-of-the-art approaches in all cases. However, the FDDL achieves the greatest result (96.7%) in the absence of noise or corruption which goes to suggest the effectiveness of low ranked approximation in the existence of noise and performs better owing to the adaptation of the loss function exhibiting the traits of Fisher criterion on the coefficients. The method that came close to the proposed K-LSDSR is the KLSGSR recording average recognition accuracy of 70.9% against occlusion rate of 50, which also yielded a good result due to the implementation of the group sparse representation. This was followed closely by the LSDSR approach also recording 69.9% recognition accuracy.

One of the imperative concerns of DL for SR is the size of the dictionary since it could greatly affect the performance of the model under discussion. Thus dictionary of varied sizes is exploited using the proposed K-LSDSR on the ORL face dataset. Random faces of cropped images are used. The training images of each subject are set from 20 to 200 consisting of 16 to 32 images per subject. The recognition rates of the proposed K-LSDSR is compared with other existing approaches in different dictionary sizes as plotted in [Fig. 4](#). From the figure, it could be realized that K-LSDSR achieves a better recognition rate of over 98% and outperform other state-of-the-art approaches even when the size of the dictionary is small.

The second is the **Yale B Face Database** that contains a total of 2432 frontal images of 38 subjects snapped under capricious light intensity environments were normalized to a size of 32×32 and the database was randomly split into two halves. One set comprising of 32 images for each subject utilized as the dictionary and the remaining set used for testing. The recognition rates for KSRC, K-SVD, LSDL, FDDL, LSDSR, KLSGSR and K-LSDSR achieved similar results in all dimensions with a rate difference of less than 0.6% as shown in the [Table 3](#): It could be seen from the table that, the proposed K-LSDSR approach consistently recorded the highest recognition results of 96.8%, 97.3%, 97.7%, 97.8% and 98.5% with varying

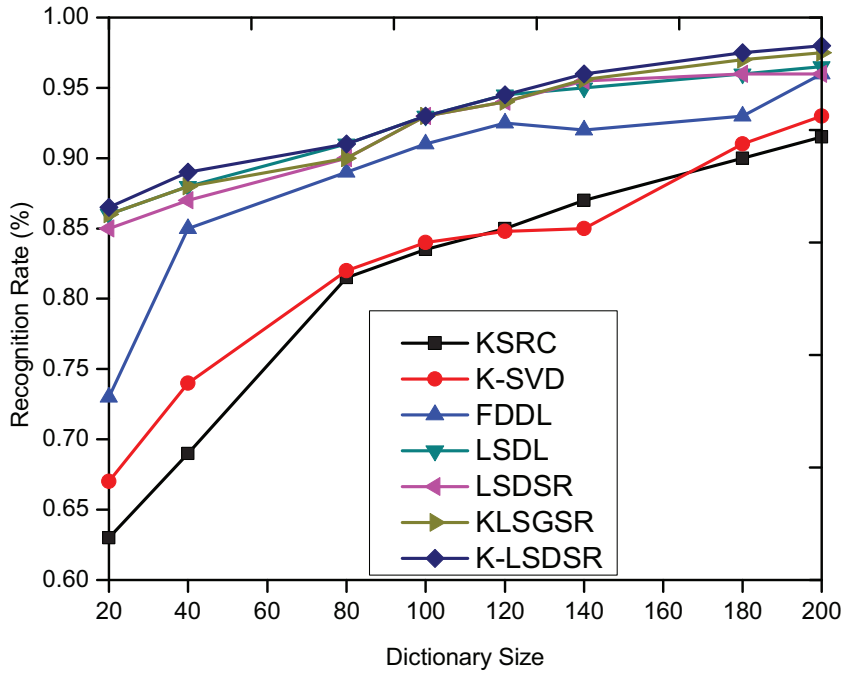


Fig. 4. Experimental results indicating recognition accuracies of K-LSDSR against other approaches with varying dictionary size on ORL face dataset with 10% occlusion.

Table 4

The results of the face recognition (%) of the various methods on the AR database.

Dim	30	60	120	200	300
KSRC	83.7	83.3	89.5	92.6	93.4
K-SVD	90.6	92.5	91.3	95.5	96.7
LSDL	90.9	91.7	93.2	96.1	97.3
FDDL	89.8	91.9	94.4	95.6	96.8
LSDSR	90.4	92.7	93.3	96.5	97.9
KLSGSR	90.9	93.0	94.6	97.1	98.1
K-LSDSR	92.3	93.6	95.2	97.8	98.7

dimensions of 50, 80, 120, 200 and 300 respectively as compared with the other methods at all levels. It was closely followed KLSGSR and LSDSR in that order. This was achieved under complex illumines conditions of the various face images.

The third to be analysed is the **AR face repository** which contains 1200 face images taken from 100 different entities or subjects with each encompassing I_2 normalized facial images of each subject out of which 50 male entities and 50 female entities are chosen. More so, the first 25 images of females and 25 males were utilized for training and the rest for testing. Every image was resized using Matlab to 32×32 image matrixes to guarantee adequate data in the estimation of Gaussian models and also for covariance matrices to be positive definite. It could be seen from **Table 4** that the K-LSDSR achieves the best results consistently in all cases when compared with the other competing methods with high dimensionality. The KSLGSR and the K-LSDSR also had the best results in terms of recognition rates higher than the other competing methods which go to say that, the Kernelled and Locality adaptor has a higher impact in facial classification. Experimental results as indicated in **Table 4** shows the recognition accuracy against varying dimensions on AR face dataset of the various algorithms used with respect to this paper. It could be seen from the results that, recognition rate increases for the K-LSDSR as the dimensions are increase from 30 to 300. This goes to suggest that, kernel approaches work best with an increase in dimension of the face images with respect to the AR database. The results of the proposed K-LSDSR approach was followed closely by KLSGSR and LSDSR in that order.

The fourth database is the **FERET Face Database** that contains a total of 1400 frontal images of 200 subjects each with 7 different images of varying illuminations and facial expressions. The images taken under varying light intensity conditions were normalized to a size of 32×32 and the database was however spliced indiscriminately into two. One set containing images of 4 for each entity utilized as the dictionary and the remaining set used for testing. The recognition rates for Kernel Sparse Representation based classification (KSRC) algorithm [50], K-SVD, LSDL, FDDL, LSDSR, KLSGSR and K-LSDSR achieved similar results in all dimensions with rate differences are shown in the **Table 5**. The table depicts the experimental

Table 5

The results of the face recognition of the various methods on the FERET database.

Dim	40	80	120	200	300
KSRC	96.4	95.4	95.7	96.8	97.9
K-SVD	93.4	95.2	96.7	97.3	97.6
LSDL	95.3	95.9	97.2	97.9	98.3
FDDL	94.8	95.7	97.4	97.7	98.0
LSDSR	95.5	96.7	97.3	97.5	97.9
KLSGSR	95.7	97.0	97.6	97.7	98.1
K-DLSR	95.9	96.4	97.5	98.2	98.5

Table 6

The outcomes of gender classification of the various methods on the AR face database.

Method	Accuracy
KSRC	0.95
K-SVD	0.96
LSDL	0.96
FDDL	0.95
LSDSR	0.96
KLSGSR	0.97
K-LSDSR	0.98

results indicating recognition accuracy against varying dimensions on FERET face dataset of the various algorithms used with respect to this paper. K-LSDSR achieved the best results in 4 of the dimensions. KSRC recorded the highest result when feature dimension was 40 with a recognition rate of 96.4%.

Gender classification

In this session, we choose 14 images per entity for the AR face dataset comprising of 50 males and 50 female entities or subjects out of which 25 images each of female and that of the males were using for training and the lingering 25 images of each gender used for testing. The dimensions of the images were reduces to 300 based on the setting of [48]. The recognition results of the K-LSDSR approach and the other competing approaches are as listed in Table 6. The dimensionality of each image as reduced to and the training samples coded with each dictionary and then classified based on the representation or the residual error. The query atom y is classified to the sample which gives the minimal residual $r(y) = ||y - D_i x_i||_2^2$. When the minimization results of the K-LSDSR are matched with that of the others, one can see that our approach get the best minimization results as matched with the first which validates that, the implementation of the dictionary using the Kernel and a locality sensitive adaptor with a loss function is more powerful on the same number of training samples. It could be realized from the results that the K-LSDSR method realized the best classification result of 98% when the implementation is based on testing the images of the dictionary of each class, and it gets best when the implementation is done with the test samples on the entire dictionary. The above accession is due to the fact that there are only two categorizations involved in gender classification, hence each categorization has a lot of training samples, resulting in the learned dictionary of every class being represented adequately in the case of the test sample.

It could be realized that the proposed DL learns a more compacted dictionary by assignments of label information adaptively to the dictionary atoms. In classification based dictionary, dictionary size is considered an important factor based on an effective rule used in setting the size. Usually, classification size increases with an increase in dictionary size. However, the size of the dictionary must not be that little for better representations more so, when kernel is involved. The dictionary size could initially be set as the number of training samples when training samples are not adequate and the size of the dictionary should be reduced than the training samples to lessen redundancy when there are sufficiently enough training samples. Samples of male and female facial images are shown in Fig. 5.

Algorithmic complexity and computational time

The complexity time is carefully presented in this section of the manuscript. It is very important to note that, though dimensionality has effect on computational cost, the speed of the kernel matrix is only reduced by high feature dimensionality and this can be precomputed. The calculations of efficiency can also be boosted by adopting the KPCA algorithm. Thus irrespective of an increase in the feature dimensions of the images, our algorithm still does well of computational speed. The algorithmic (computational) complexity of the proposed K-LSDSR and the comparative approaches are presented in Section 4.6 of our manuscript. Consequently, to convex optimization problem is to solve an over-complete dictionary (sparse



Fig. 5. Samples of male and female facial images from the AR database.

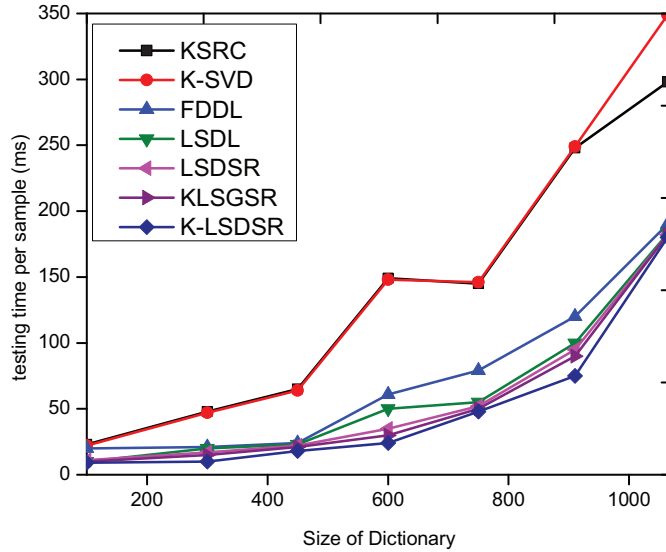


Fig. 6. The figure shows varying dictionary sizes with their testing times.

representation). Each dictionary atom is initially reconstructed by the remaining atoms and the algorithmic complexity of the reconstruction is represented as $O(M)$, where M denotes the size of the dictionary D for the sparse learning. If the dictionary D is assumed to be fixed at the sparse learning stage, then the algorithmic complexity of the optimized sparse coefficient stage is $O(N^2+N)$ where N denotes the size of the training set A . Lastly, if we assumed that the matrix X is fixed at the dictionary update stage, the algorithmic complexity of solving the learned dictionary D is $O(M^2)$. Thus the algorithmic complexity of the proposed K-LSDSR will be $O(K*(M + +N + +N^2+M^2))$ where K is the iteration number. For LSDSR, the algorithmic complexity is $O(K*(N^2+M^2+N))$. The computational complexity of LSDL is $O(K*(M^2+M + +N))$ because it has no discrimination function as part of the sparse coefficient representation.

The key measuring tool used in evaluating dictionary learning algorithms is the testing time of samples. Thus the mean testing time of the proposed K-LSDSR of every test sample on the Extended YaleB face dataset is compared with the other algorithms including KSRC, K-SVD, LSDL, and FDDL. Varied dictionary sizes are set ranging from 100 to 1064 with each method tested on random faces of 10 divisions of the datasets. The means of the testing results are shown in Fig. 6. The figure shows varying dictionary sizes with their testing times. The result shows that the proposed K-LSDSR outperformed the other methods with varying dictionary size.

Analysis of the proposed K-LSDSR approach

In the proposed K-LSDSR method, the kernelized locality-sensitive adaptor is employed to increasingly control the process of dictionary learning by superimposing the capture of local discriminative nonlinear data of face data. Also, a discriminative loss function based on sparse coefficients is combined into the structure of the proposed K-LSDSR to enhance the discrimination ability of SR. In addition, the proposed algorithm results in an optimized dictionary that enhances further, the discriminative classification of SR. We chose face samples or images from the Extended Yale B dataset for a subject with 32 training images each with 192×168 pixels. An experiment was performed to determine the efficient discrimination ability of the proposed K-LSDSR, LSDL and KSRC. The experiment was carried out with the same training sets for the three

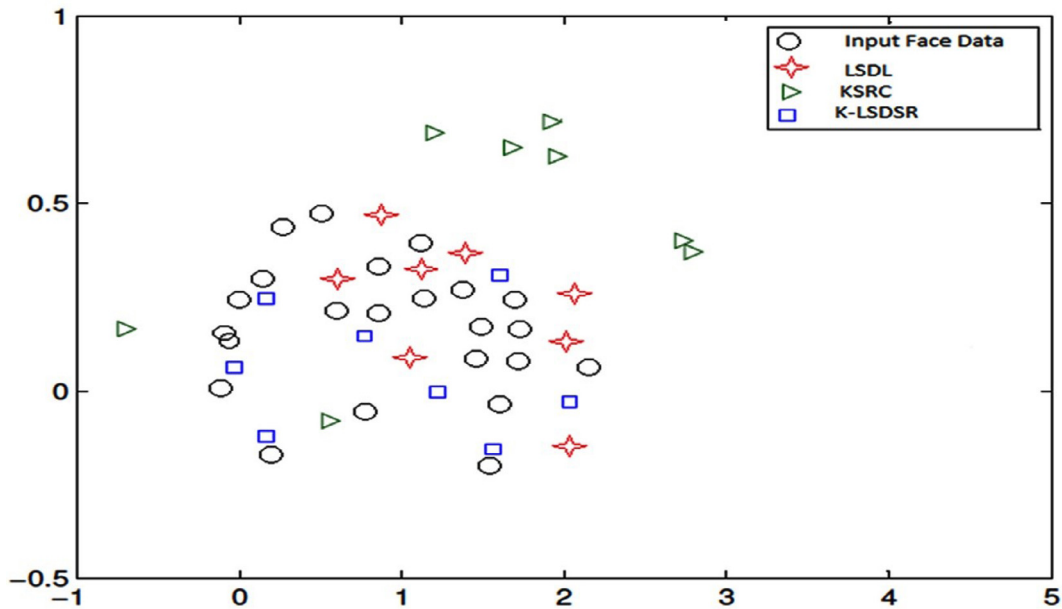


Fig. 7. Reconstructed Samples from two classes of KSRC, LSDL and K-LSDSR.

algorithms. This was done to optimize the dictionaries and obtain sparse coefficients of the testing samples. The test samples are then reconstructed by the optimized dictionaries and projected onto two dimensional subspaces by PCA with the first 300 eigenvectors. The two dimensional subspaces of KSRC, LSDL and the proposed K-LSDSR method are plotted via multidimensional scaling as shown in the sample distribution graphs of Fig. 7 by randomly choosing (8) instances. There is no clear distinction amongst the atoms for KSRC after reconstruction. Even though the integration of data locality and sparsity into the structure of LSDL resulting in a sufficient representation, there still exist a massive dispersion amongst some of the learned atoms from the training instances as could be seen from the figure. However, the proposed K-LSDSR has its boundaries more clarified with the optimized dictionary. More so, samples from the same categorization are more compact while the ones from different categories are well separated for the proposed K-LSDSR technique. Subsequently, we can conclude that the proposed K-LSDSR is expected to results in an improved recognition performance due to the integration of kernel, data locality and sparsity as well as the introduction of a loss function.

Discussions

The highest recognition rate of all the experimental results are represented in bold values as indicated by Tables 2-5. The algorithms achieve a higher recognition rate since the images are mapped into a high dimensional feature space as a result of more information being carried by high dimensionality which enhances recognition performance. The rate of recognition increases steadily as feature dimensions also increases amongst the various approaches. In addition, the proposed method demonstrated superiority over the other methods even at low feature dimensions as indicated by the various experimental results on the different datasets. This progresses in that order to the features space with high dimensionality. Furthermore, the proposed K-LSDSR approach exhibits more classification and discrimination capabilities as a result of the introduction of the Locality-Sensitive adaptor into the structure of the proposed discriminative algorithm.

Conclusion and recommendations for future work

To handle efficiently multi-featured approach to discrimination, a Kernel Locality Sensitive Discriminative Sparse Representation (K-LSDSR) for face recognition is presented in this paper. The paper recognized that, image samples from the same categorization could be encrypted as the same sparse coefficients in the course of image recognition to enhance the discrimination capability features of SR. Thus it takes into account data locality in the kernel space and hence better discriminability of SR and better utilization of non-linear information hidden in the training and the test data. The proposed K-LSDSR directly use kernel to map features from a high dimensional space, and hence enables fusion of multi features and an enhanced data demonstration due to its ability to preserve data locality. It could be seen from the experimental results that, incorporating kernel into the structure of dictionary learning coupled with Locality - Sensitive Regularization Adaptor and a Loss Function inherently improves the classification performance as compared with other Dictionary Learning approaches. Furthermore, the proposed method is more computationally effective as a result of the implementation of the l_2 - norm which increases

its stability. Thus close form solutions are derived from the sparse coding and dictionary update stages. We however proposed to incorporate weighted K-nearness neighbour into the structure of the Kernel Locality-Sensitive Discriminative Sparse Representation to enhance the power of classification.

Declaration of Competing Interest

The authors declare they have no competing interests.

CRediT authorship contribution statement

Ben-Bright Benuwa: Conceptualization, Writing - review & editing. **Benjamin Ghansah:** Resources, Software, Data curation. **Ernest K. Ansah:** Formal analysis, Methodology.

Acknowledgement

The authors are thankful to Juliana Serwaa Andoh (Mrs) for reviewing this article and making critical inputs that improved the quality of the article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Trans. Signal Process.* 54 (11) (2006) 4311–4322.
- [2] B.-B. Benuwa, Y. Zhan, J. Liu, J. Gou, B. Ghansah, E.K. Ansah, Group sparse based locality-sensitive dictionary learning for video semantic analysis, *Multimed. Tools Appl.* (2018) 1–24.
- [3] B.B. Benuwa, B. Ghansah, D.K. Wornyo, S.A. Adabunu, A comprehensive review of particle swarm optimization, in: *Proceedings of the International Journal of Engineering Research in Africa*, 23, 2016.
- [4] B.B. Benuwa, Y.Z. Zhan, B. Ghansah, D.K. Wornyo, F. Banaseka Kataka, A review of deep machine learning, in: *Proceedings of the International Journal of Engineering Research in Africa*, 2016.
- [5] S. Cai, W. Zuo, L. Zhang, X. Feng, P. Wang, Support vector guided dictionary learning, in: *Proceedings of the European Conference on Computer Vision*, 2014.
- [6] H. Chang, M. Yang, J. Yang, Learning a structure adaptive dictionary for sparse representation based classification, *Neurocomputing* 190 (2016) 124–131.
- [7] S.-P. Chen, X.-H. Wang, Z.-Q. Peng, D.-C. Chen, H. Jun, Facial expression recognition via kernel sparse representation, *DEStech Trans. Comput. Sci. Eng.* (2019) 231–236 (ammso).
- [8] B. Ding, H. Ji, Learning kernel-based robust disturbance dictionary for face recognition, *Appl. Sci.* 9 (6) (2019) 1189.
- [9] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, *Proc. Natl. Acad. Sci.* 100 (5) (2003) 2197–2202.
- [10] M. Dyrba, M. Grothe, T. Kirste, S.J. Teipel, Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM, *Hum. Brain Mapp.* 36 (6) (2015) 2118–2131.
- [11] K. Engan, S.O. Aase, J.H. Husoy, Method of optimal directions for frame design, in: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999.
- [12] Z. Fan, D. Zhang, X. Wang, Q. Zhu, Y. Wang, Virtual dictionary based kernel sparse representation for face recognition, *Pattern Recognit.* 76 (2018) 1–13.
- [13] Z. Feng, M. Yang, L. Zhang, Y. Liu, D. Zhang, Joint discriminative dimensionality reduction and dictionary learning for face recognition, *Pattern Recognit.* 46 (8) (2013) 2134–2143.
- [14] M. Harandi, M. Salzmann, Riemannian coding and dictionary learning: kernels to the rescue, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [15] K. Huang, S. Aviyente, Sparse representation for signal classification, in: *Proceedings of the Advances in Neural Information Processing Systems*, 2006.
- [16] M. Hussain, S.K. Wajid, A. Elzaat, M. Berbar, A comparison of SVM kernel functions for breast cancer detection. Paper presented at the computer graphics, imaging and visualization (CGIV), in: *Proceedings of the Eighth International Conference on*, 2011.
- [17] Z. Jiang, Z. Lin, L.S. Davis, Learning a discriminative dictionary for sparse coding via label consistent K-SVD, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [18] Z. Kang, X. Lu, J. Yi, Z. Xu, Self-weighted multiple kernel learning for graph-based clustering and semi-supervised classification, 27 2018 2312–2318. arXiv:1806.07697.
- [19] Y.-S. Lee, C.-Y. Wang, S. Mathulapransan, J.-H. Zhao, J.-C. Wang, Locality-preserving K-SVD based joint dictionary and classifier learning for object recognition, in: *Proceedings of the ACM on Multimedia Conference*, 2016.
- [20] L. Li, S. Li, Y. Fu, Discriminative dictionary learning with low-rank regularization for face recognition, in: *Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2013.
- [21] H.-T. Lin, C.-J. Lin, A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods, *Neural Comput.* 3 (2003) 1–32.
- [22] J. Liu, J. Gou, Y. Zhan, Q. Mao, Discriminative self-adapted locality-sensitive sparse representation for video semantic analysis, *Multimed. Tools Appl.* 77 (21) (2018) 29143–29162.
- [23] W. Liu, Z. Yu, M. Yang, L. Lu, Y. Zou, Joint kernel dictionary and classifier learning for sparse coding via locality preserving K-SVD, in: *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2015.
- [24] J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman, Discriminative learned dictionaries for local image analysis, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR*, 2008.
- [25] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, F.R. Bach, Supervised dictionary learning, in: *Proceedings of the Advances in Neural Information Processing Systems*, 2009.
- [26] J. Meng, C. Jung, Class-discriminative kernel sparse representation-based classification using multi-objective optimization, *IEEE Trans. Signal Process.* 61 (18) (2013) 4416–4427.
- [27] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, Z. Zhu, Robust face recognition via occlusion dictionary learning, *Pattern Recognit.* 47 (4) (2014) 1559–1572.

- [28] D.-S. Pham, S. Venkatesh, Joint learning and dictionary construction for pattern recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR*, 2008.
- [29] L. Rosasco, A. Verri, M. Santoro, S. Mosci, S. Villa, Iterative projection methods for structured sparsity regularization, 2009 1–29.
- [30] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, 1994.
- [31] B. Schölkopf, R. Herbrich, A.J. Smola, A generalized representer theorem, in: *Proceedings of the International Conference on Computational Learning Theory*, 2001.
- [32] A. Shrivastava, V.M. Patel, R. Chellappa, Multiple kernel learning for sparse representation-based classification, *IEEE Trans. Image Process.* 23 (7) (2014) 3013–3024.
- [33] Y. Sun, Q. Liu, J. Tang, D. Tao, Learning discriminative dictionary for group sparse representation, *IEEE Trans. Image Process.* 23 (9) (2014) 3816–3828.
- [34] S. Tan, X. Sun, W. Chan, L. Qu, L. Shao, Robust face recognition with kernelized locality-sensitive group sparsity representation, *IEEE Trans. Image Process.* 26 (10) (2017) 4661–4668.
- [35] R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective, *J. R.Stat. So.* 73 (3) (2011) 273–282.
- [36] H. Van Nguyen, V.M. Patel, N.M. Nasrabadi, R. Chellappa, Kernel dictionary learning, in: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.
- [37] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained Linear Coding for image classification, in: *Proceedings of the Computer Vision and Pattern Recognition*, 2010.
- [38] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [39] P. Wang, J. Lan, Y. Zang, Z. Song, Discriminative structured dictionary learning for image classification, *Trans. Tianjin Univ.* 22 (2016) 158–163.
- [40] Q. Wang, G. Qu, Extended omp algorithm for sparse phase retrieval, *Signal Image Video Process.* 11 (8) (2017) 1397–1403.
- [41] C.-P. Wei, Y.-W. Chao, Y.-R. Yeh, Y.-C.F. Wang, Locality-sensitive dictionary learning for sparse representation based classification, *Pattern Recognit.* 46 (5) (2013) 1277–1287.
- [42] C.P. Wei, Y.W. Chao, Y.R. Yeh, Y.C.F. Wang, Locality-sensitive dictionary learning for sparse representation based classification, *Pattern Recognit.* 46 (5) (2013) 1277–1287.
- [43] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
- [44] G.-S. Xie, X.-Y. Zhang, X. Shu, S. Yan, C.-L. Liu, Task-driven feature pooling for image classification, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015.
- [45] Y. Xu, Y. Sun, Y. Quan, B. Zheng, Discriminative structured dictionary learning with hierarchical group sparsity, *Comput. Vis. Image Underst.* 136 (2015) 59–68.
- [46] S. Yadav, V.P. Vishwakarma, Extended interval type-II and kernel based sparse representation method for face recognition, *Expert Syst. Appl.* 116 (2019) 265–274.
- [47] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [48] M. Yang, L. Zhang, X. Feng, D. Zhang, Fisher discrimination dictionary learning for sparse representation, in: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011.
- [49] M. Yang, L. Zhang, X. Feng, D. Zhang, Sparse representation based fisher discrimination dictionary learning for image classification, *Int. J. Comput. Vis.* 109 (3) (2014) 209–232.
- [50] J. Yin, Z. Liu, Z. Jin, W. Yang, Kernel sparse representation based classification, *Neurocomputing* 77 (1) (2012) 120–128.
- [51] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, in: *Proceedings of the Advances in Neural Information Processing Systems*, 2009.
- [52] Y. Zhan, J. Liu, J. Gou, M. Wang, A video semantic detection method based on locality-sensitive discriminant sparse representation and weighted KNN, *J. Vis. Commun. Image Rep.* 41 (2016) 65–73.
- [53] Y. Zhan, J. Liu, J. Gou, M. Wang, A video semantic detection method based on locality-sensitive discriminant sparse representation and weighted KNN ☆, *J. Vis. Commun. Image Rep.* 41 (2016) 65–73.
- [54] Q. Zhang, B. Li, Discriminative K-SVD for dictionary learning in face recognition, in: *Proceedings of the Computer Vision and Pattern Recognition*, 2010.
- [55] Z. Zhang, Y. Xu, J. Yang, X. Li, D. Zhang, A survey of sparse representation: algorithms and applications, *IEEE Access* 3 (2015) 490–530.
- [56] Z. Zheng, H. Sun, G. Zhang, Multiple kernel locality-constrained collaborative representation-based discriminant projection for face recognition, *Neurocomputing* 318 (2018) 65–74.
- [57] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 67 (2) (2005) 301–320.