

Research Article

# Image Clustering Using Exponential Regularized Discriminant Analysis

Kwang Jun Pak\*

Faculty of Applied Mathematics, Kim Chaek University of Technology, Pyongyang, the Democratic People's Republic of Korea

## Abstract

When clustering images, the images are typically sampled as nonlinear manifolds. In this case, local learning-based image clustering models are used. Several proposed clustering models are based on linear discriminant analysis (LDA). In image clustering based on linear discriminant analysis (LDA), the problem of small-sample-size (SSS) is presented when the dimensionality of image data is larger than the number of samples. To solve this problem, various image clustering models based on local learning have been introduced. In the proposed clustering models, we added tuning parameters to deal with the small-sample-size (SSS) problem arising in linear discriminant analysis (LDA). In this paper, we propose an exponential regularized discriminant clustering model as an image clustering model based on local learning. In the proposed local exponentially regularized discriminant clustering (LERDC) model, the local scattering matrices of the regularized discriminant model are projected into the exponential domain to address the SSS problem of LDA. Compared with previous clustering methods based on local learning,  $k$ -nearest neighbors and regularization parameter  $\lambda$  in the local exponentially regularized discriminant clustering model are the tuning parameters for clustering. The experiments are concluded that the clustering performance of the proposed LERDC model is comparable to that of the clustering methods based on previous local learning.

## Keywords

Image Clustering, Discriminant Analysis, Local Learning

## 1. Introduction

Cluster analysis is often used to process and classify images by computer as well as classification of data in various scientific fields [1-3].

Image clustering is a method to optimally partition images so that image data points belonging to the same group are similar to each other and not to image data points belonging to other groups.

In many fields, data are generally high dimensional, so to

process high-dimensional data, the dimensionality of data must be reduced.

Linear discriminant analysis (LDA) is a well-known method for dimensionality reduction.

In [4], a model selection algorithm for PCA+LDA and RLDA (regularized linear discriminant analysis) is proposed.

Some researchers have used repeated K-means to obtain clustering labels in LDA and LDA to select the discriminant

\*Corresponding author: pkj75123@star-co.net.kp (Kwang Jun Pak)



subspace in K-means clustering [5, 6].

In [7], a discriminant K-means (Diskmeans) algorithm is proposed to solve the trace maximization problem by a simple iterative procedure.

However, LDA cannot be directly used on high-dimensional image datasets because of the small-sample-size (SSS) problem.

Therefore, in [8], discriminant information in the data matrix was used for linear discriminant analysis.

Suppose that  $X$  is an image dataset consisting of  $n$  images from  $c$  different classes.

Let us denote the within-class and between-class scattering matrices by  $S_w$  and  $S_b$ , respectively.

We can obtain the optimal scaled cluster assignment matrix  $A^*$  using LDA by solving the maximization problem [8] as follows

$$A^* = \arg \max_A \text{tr}((S_w)^{-1} S_b). \quad (1)$$

In [9], LDA was modified by using the total scattering matrix  $S_t (= S_w + S_b)$  instead of  $S_w$  in (1).

In image clustering by LDA, if the number of images in an image dataset is less than the image feature dimension then the SSS problem often occurs.

Methods to deal with the SSS problem in LDA at the global level have been proposed in [10-12].

The regularized discriminant analysis (RDA) is presented in [10], where they added the identity matrix  $I$  with regularization parameter  $\lambda > 0$  to make  $S_t$  non-singular such as the following

$$A^* = \arg \max_A \text{tr}((S_t + \lambda I)^{-1} S_b). \quad (2)$$

In RDA,  $S_t$  and  $S_b$  are evaluated on a global level for the entire image dataset  $X$ . In [13-18] using the manifold assumption, researchers have developed various local learning based clustering methods to efficiently learn nonlinear manifolds in image data using local neighborhood information.

Ncut (normalized cutting) [16] and K-way Ncut [19] are well-known clustering methods that perform data clustering through local learning based on similarity between neighboring data points.

In [17], based on RDA [10], an image clustering model using LDMGI (local discriminant models and global integration) is proposed.

In LDMGI, the local image matrix  $X_i$  is made using  $k$ -nearest neighbor images of each image  $x_i$  in the image dataset  $X$ .

To evaluate the clustering performance of images in the local image matrix  $X_i$ , a local clustering model was considered.

The local clustering model with all the local image matrices  $X_i$ ,  $i=1, 2, \dots, n$ , is set as the following optimization problem.

$$A^* = \arg \min_{A_{(i)}} \sum_{i=1}^n \text{tr}[A_{(i)}^T L_i A_{(i)}], \quad (3)$$

$$L_i = H_k (\tilde{X}_i^T \tilde{X}_i + \lambda I)^{-1} H_k, \quad (4)$$

where  $A_{(i)}$ ,  $L_i$ ,  $\tilde{X}_i$  are the local scaled cluster assignment matrix, local discriminant model and centering matrix for the local image matrix  $X_i$ , respectively. And  $\tilde{X}_i$  is expressed by

$$\tilde{X}_i = X_i H_k, H_k = I - (1/k) \mathbf{1}_k \mathbf{1}_k^T \in \mathbb{R}^{k \times k}.$$

To obtain the solution of the cluster assignment matrix, we used the eigenvalue decomposition method, which obtains a more discretized binary cluster assignment matrix for each image data  $x_i$ ,  $i=1, 2, \dots, n$  of the image dataset  $X$ .

Next, comparing (2) and (4), we can see that the regularization parameter  $\lambda$  added in (2) to solve the SSS problem of LDA on the image dataset  $X$  at the global level also exists in the local discriminant model  $L_i$ , such as (4), to deal with the SSS problem of LDA on the local image data  $X_i$  at the local level.

In the LDMGI model proposed in [17], the value of  $\lambda$  was chosen reasonably for  $\{10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6, 10^8\}$  and the number of images  $k$  nearest to image  $x_i$  was set to 5.

In [20], an EDA (exponential discriminant analysis) is proposed to map scattering matrices into exponential domain in LDA.

As shown in [20], the discriminant information of LDA is not lost even in EDA.

Some researchers have proposed exponential local discriminant embedding (ELDE) [21] and exponential discriminant analysis method [22] based on local learning for face recognition problems.

In this paper, we consider a regularized exponential discriminant model based on local learning.

In LERDC (local exponential regularized discriminant clustering) model, the local scattering matrices of regularized discriminant are mapped onto the exponential domain.

In the proposed LERDC model, the local image matrix  $X_i$  for each image  $x_i$  was constructed including  $k$  ( $=5$ ) nearest neighbor images and a local exponential regularized discriminant model (LERDM) was proposed to perform clustering for images in the local image matrix  $X_i$ .

The whole image clustering result was obtained by summing all the LERDMs for all local image matrices  $X_i$ ,  $i = 1, 2, \dots, n$ .

In LERDM for image  $x_i$ ,  $\tilde{\tilde{X}}_i$  is the image complement of  $\tilde{X}_i$ , and  $k$  and  $\lambda$  are the clustering parameters in the proposed local exponential regularized discriminant clustering model.

The clustering performance of the proposed LERDC model is comparable to known clustering models (Ncut [19], SEC [18] and LDMGI [17]).

Experimental results using 10 image databases showed that the proposed LERDC model can compare clustering performance with the LDMGI model.

The proposed LERDC model is based on EDA.

The proposed LERDC model is evaluated to be able to compare clustering performance with the currently known local learning-based clustering methods.

The rest of this paper is organized as follows.

In Section 2 of the paper, we give the mathematical formulation of the LERDC model.

In Section 3, the proposed LERDC model is compared with the existing clustering method. The experimental results are presented in Section 4. The conclusions for the proposed LERDC are given in Section 5.

## 2. LERDC Model

In this section, we describe the proposed local exponential regularized discriminant clustering (LERDC) model for image clustering.

Let  $\mathbb{S}$  be a set of  $n$  images and  $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{m \times n}$  be the image data matrix, where  $x_i \in \mathbb{R}^m$ ,  $i = 1, 2, \dots, n$  is the  $i$ th image data and  $m$  is the feature dimension of image data  $x_i$ .

For image data matrix  $X$ , the total scattering matrix  $S_t$  and the between-class scattering matrix  $S_b$  are

$$S_t = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T = \tilde{X}\tilde{X}^T,$$

$$S_b = \sum_{i=1}^d n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T = \tilde{X}\tilde{Z}\tilde{Z}^T \tilde{X}^T,$$

$$\tilde{X} = XH_n, H_n = I - (1/n)1_n 1_n^T \in \mathbb{R}^{n \times n},$$

$$1_n = (1, 1, \dots, 1)^T \in \mathbb{R}^n, H_n = H_n^T = H_n H_n$$

where  $\tilde{X}$  is the centralization matrix of the image data matrix  $X$ ,  $\bar{x}_i$  is the average of images belonging to the  $i$ th class, and  $\bar{x}$  is the average of  $n$  images belonging to the image dataset  $\mathbb{S}$ . Image clustering is to partition  $n$  images into  $d$  clusters  $\{D_j, j=1, 2, \dots, d\}$  of similar images. Let  $Y = [y_1, y_2, \dots, y_n]^T \in \{0, 1\}^{n \times d}$  be the cluster assignment matrix, where  $y_i \in \{0, 1\}^{d \times 1}$  is a cluster assignment column vector whose image  $x_i$  belongs to the  $j$ th cluster, while the  $j$ th element of  $y_i$  is 1 and otherwise 0.

By [19], the scaled cluster assignment matrix  $Z \in \mathbb{R}^{n \times d}$  is given by

$$Z = [z_1, z_2, \dots, z_n]^T = Y(Y^T Y)^{-1/2}, \quad (5)$$

where  $z_i$  is the scaled cluster assignment vector as follows

$$z_j = \left[ \underbrace{0, \dots, 0}_{\sum_{i=1}^{j-1} n_i}, \underbrace{\frac{1}{\sqrt{n_j}}, \dots, \frac{1}{\sqrt{n_j}}}_{n_j}, \underbrace{0, \dots, 0}_{\sum_{i=j+1}^d n_i} \right]^T \quad (6)$$

Here  $n_j$  is the number of images in the  $j$ th cluster  $\{D_j, j=1, 2, \dots, d\}$ .

In (5), we see that  $Y^T Y$  is a diagonal matrix and

$$Z^T Z = (Y^T Y)^{-1/2} Y^T Y (Y^T Y)^{-1/2} = I \quad (7)$$

From [6], the optimal scaled cluster assignment matrix  $Z^*$  can be write as

$$Z^* = \arg \max_Z \text{tr}((\exp(S_t + \lambda I))^{-1} \exp(S_b)). \quad (8)$$

Using (7), the maximization problem of (8) can be replaced by the minimization problem as follows

$$Z^* = \arg \min_Z \text{tr}\{Z^T Z - ((\exp(S_t + \lambda I))^{-1} \exp(S_b))\}. \quad (9)$$

The manifold learning algorithms [13, 14] and clustering algorithms [16-18] demonstrated that using local geometric properties is beneficial to recover the intrinsic manifold structure of data in manifold learning. Considering that the structure of the local manifold in [14] is mostly linear, it can be considered that the local image cluster  $\mathbb{S}_k(x_i)$  for each image  $x_i$  consists of  $k$  images including image  $x_i$  and its nearest  $k-1$  neighbor images.

Let the index set of images belonging to the set of local images  $\mathbb{S}_k(x_i)$  for image  $x_i$  be  $F_i = \{i_0, i_1, i_2, \dots, i_{k-1}\}$ ,  $i_0 = i$ , and the local image data matrix of  $\mathbb{S}_k(x_i)$  be  $X_i = (x_{i_0}, x_{i_1}, \dots, x_{i_{k-1}})$ .

The total scattering matrix and the between-class scattering matrix for the local image data matrix  $X_i$  can be defined as follows

$$S_{t(i)} = \tilde{X}_i \tilde{X}_i^T, S_{b(i)} = \tilde{X}_i Z_{(i)} Z_{(i)}^T \tilde{X}_i^T,$$

where  $Z_{(i)}$  is the local scaled cluster assignment matrix obtained from matrix  $Z$  as follows

$$Z_{(i)} = G_{(i)}^T Z. \quad (10)$$

In (10) elements of  $G_{(i)} \in \{0, 1\}^{n \times k}$  is the selection matrix with its elements  $(G_{(i)})_{pq} = 1$  if  $p = F_i(q)$ ;  $(G_{(i)})_{pq} = 0$ , otherwise and  $Z_{(i)} \in \mathbb{R}^{k \times d}$ .  $p$  is the index of image data and  $q$  is the index of the

image data which belong to the set of local images  $\mathfrak{S}_k(x_i)$ .

The  $j$ th column of the matrix  $Z_{(i)}$  is equal to

$$Z_{(i)}^j = \left[ \underbrace{0, \dots, 0}_{\sum_{r=1}^{j-1} m_{i_r}}, \underbrace{\frac{1}{\sqrt{n_j}}, \dots, \frac{1}{\sqrt{n_j}}}_{m_{i_j}}, \underbrace{0, \dots, 0}_{\sum_{r=j+1}^d m_{i_r}} \right]^T \quad (11)$$

where  $m_{i_r}, r=1, 2, \dots, d$  be the number of images belonging to the  $r$ th cluster in the set of local images  $\mathfrak{S}_k(x_i)$ . As in [17], imposing a regular term  $\text{tr}(Z_{(i)}^T H_k Z_{(i)})$  to control the power of LERDC, the objective function for the  $i$ th local image matrix  $X_i$  can be written as follows

$$\arg \min_{Z_{(i)}} \text{tr}\{Z_{(i)}^T H_k Z_{(i)} - ((\exp(S_{t(i)} + \lambda I))^{-1} \exp(S_{b(i)}))\}. \quad (12)$$

Using define and property of matrix exponential, we have

$$\arg \min_{Z_{(i)}} \text{tr}\{Z_{(i)}^T H_k Z_{(i)} - \exp(-(S_{t(i)} + \lambda I) + S_{b(i)})\}. \quad (13)$$

Also, substituting values for  $S_{t(i)}$  and  $S_{b(i)}$  in (13), we obtain

$$\arg \min_{Z_{(i)}} \text{tr}\{Z_{(i)}^T H_k Z_{(i)} - \exp(-(\tilde{X}_i \tilde{X}_i^T + \lambda I) + \tilde{X}_i Z_{(i)} Z_{(i)}^T \tilde{X}_i^T)\}. \quad (14)$$

It has been shown in [23] that for any square matrix  $U$ , its exponential  $\exp(U)$  is finite and there is also an inverse

$$(\exp(U))^{-1} = \exp(-U).$$

Using Jordan matrix decomposition in [24], the square matrix  $U$  is decomposed into

$$U = W J W^{-1},$$

where  $U$  and  $J$  are similar matrices,  $J$  is a matrix of Jordan canonical form and  $W^{-1}$  is the matrix inverse of  $W$ . We can write from the above relation:  $J = W^{-1} U W$ . Thus, we can see that

$$\exp(J) = \exp(W^{-1} U W) = W^{-1} \exp(U) W.$$

Lemma [22]: Let  $U = \tilde{X}_i \tilde{X}_i^T + \lambda I$ ,  $V = \tilde{X}_i Z_{(i)} Z_{(i)}^T \tilde{X}_i^T$  and  $\bar{U}, \bar{V}$  are the image complements of matrices  $U$  and  $V$ ,

respectively. Then we have  $-\overline{U + V} = -\bar{U} \bar{V}$

By the Lemma, (14) can be written as follows

$$\arg \min_{Z_{(i)}} \text{tr}\left\{Z_{(i)}^T H_k Z_{(i)} - \exp\left(-\left(\tilde{X}_i \tilde{X}_i^T + \lambda I\right) \tilde{X}_i Z_{(i)} Z_{(i)}^T \tilde{X}_i^T\right)\right\}, \quad (15)$$

where  $\bar{X}_i$  is the image complement of  $\tilde{X}_i$  obtained using [25].

Theorem 1: Let matrix  $A$  is  $d \times n$  dimension and  $B$  is  $n \times d$  dimension. Then  $\text{tr}(AB) = \text{tr}(BA)$

Proof:

$$\begin{aligned} \text{tr}(AB) &= \sum_i [AB]_{ii} = \sum_i A_i * B_{*i} = \sum_i \sum_k a_{ik} b_{ki} = \sum_k \sum_i b_{ki} a_{ik} \\ &= \sum_k B_{k*} A_{*k} = \sum_k [BA]_{kk} = \text{tr}(BA) \end{aligned}$$

Let  $A = (\tilde{X}_i \tilde{X}_i^T + \lambda I) \tilde{X}_i Z_{(i)}$  and  $B = Z_{(i)}^T \tilde{X}_i^T$  from theorem 1. Then we show that (15) can be circularly shifted as follows

$$\arg \min_{Z_{(i)}} \text{tr}\left\{Z_{(i)}^T H_k Z_{(i)} - \exp\left(-Z_{(i)}^T \tilde{X}_i^T (\tilde{X}_i \tilde{X}_i^T + \lambda I) \tilde{X}_i Z_{(i)}\right)\right\}. \quad (16)$$

(16) is as following

$$\arg \min_{Z_{(i)}} \text{tr}\left\{Z_{(i)}^T H_k Z_{(i)} - \exp\left(-Z_{(i)}^T (\tilde{X}_i^T \tilde{X}_i + \lambda I) \tilde{X}_i Z_{(i)}\right)\right\}, \quad (17)$$

because  $A(A^T A + \lambda I) = (A A^T + \lambda I) A$  for any matrix  $A$ .

Theorem 2: The optimization problem of (17) is equivalent to the following cost function

$$\arg \min_{Z_{(i)}} \text{tr}\left\{Z_{(i)}^T \left[ H_k \exp\left(\frac{-(\tilde{X}_i^T \tilde{X}_i + \lambda I) \tilde{X}_i Z_{(i)} Z_{(i)}^T \tilde{X}_i}{\|(\tilde{X}_i^T \tilde{X}_i + \lambda I) \tilde{X}_i Z_{(i)} Z_{(i)}^T \tilde{X}_i\|}\right) H_k \right] Z_{(i)}\right\} \quad (18)$$

We can simply prove by using the proof method of theorem 6 in [22].

Using Theorem 2, we can write (18) as follows

$$\arg \min_{Z_{(i)}} \text{tr}[Z_{(i)}^T L_{ER_i} Z_{(i)}], \quad (19)$$

where

$$L_{ER_i} = H_k \exp\left(\frac{-(\tilde{X}_i^T \tilde{X}_i + \lambda I) \tilde{X}_i Z_{(i)} Z_{(i)}^T \tilde{X}_i}{\|(\tilde{X}_i^T \tilde{X}_i + \lambda I) \tilde{X}_i Z_{(i)} Z_{(i)}^T \tilde{X}_i\|}\right) H_k \quad (20)$$

is the LERDM (local exponential regularized discriminant model) for the image  $x_i$ .

In (20), the dimensions of  $\widetilde{\mathbf{X}}_i$  and  $\widetilde{\mathbf{X}}_i^T \widetilde{\mathbf{X}}_i$  are  $d \times k$  and  $k \times k$ , respectively. Also, as in [17, 18], the number of nearest neighbor images  $k$  is set to 5. Thus,  $L_{ER_i}$ ,  $i=1, 2, \dots, n$  is the  $k \times k$  matrix which is computationally not extensive.

To obtain the optimal  $Z$ , the following objective function was obtained by summing all the local exponential regularized discriminant model (LERDM)  $L_{ER_i}$ ,  $i=1, 2, \dots, n$ :

$$\arg \min_{Z(i)} \sum_{i=1}^n \text{tr} \left[ Z(i)^T L_{ER_i} Z(i) \right]. \quad (21)$$

From (10), we have

$$\arg \min_Z \text{tr} \left[ Z^T \left( \sum_{i=1}^n G(i) L_{ER_i} G(i)^T \right) Z \right].$$

Let

$$\begin{aligned} L_{ER} &= \sum_{i=1}^n G(i) L_{ER_i} G(i)^T \\ &= [G_1, G_2, \dots, G_n] \begin{bmatrix} L_{ER_1} & & \\ & \dots & \\ & & L_{ER_n} \end{bmatrix} [G_1, G_2, \dots, G_n]^T \end{aligned} \quad (22)$$

The objective function of LERDC becomes

$$\arg \min_Z \text{tr} [Z^T L_{ER} Z] \quad (23)$$

As proved in [17], the matrix defined in (22) is a Laplacian operator matrix. However, the objective function for the local exponentially regularized discriminant clustering using regularized discriminant with two constraints  $Z=Y(Y^T Y)^{-1/2}$  and  $Z^T Z=I$  for  $Z$  can be written as

$$\begin{aligned} &\arg \min_Z \text{tr} [Z^T L_{ER} Z] \\ &\text{s.t. } Z = Y(Y^T Y)^{1/2} \text{ and } Z^T Z = I \end{aligned} \quad (24)$$

Since the scaling cluster assignment matrix  $Z$  of the objective function in (24) is constrained by  $Z=Y(Y^T Y)^{-1/2}$ , removing this constraint from [26] and relaxing  $Z$  in the continuous range, the objective function can be written as

$$\arg \min_{Z^T Z} \text{tr} (Z^T L_{ER} Z).$$

To solve such optimization problems, eigenvalue decomposition method is used.

$$L_{ER} \omega_i = \lambda_i \omega_i \quad (25)$$

where  $\lambda_0 < \lambda_1 < \lambda_2 < \dots, < \lambda_{n-1}$  are the eigenvalues and  $\omega_i$ ,  $i=0, 1, \dots, n-1$  are the corresponding eigenvectors. Eliminating the trivial solutions with  $\lambda_0=0$  and  $\omega_0=1_n$ , we have the optimal scaled cluster allocation matrix  $Z^*$  as follows

$$Z^* = [\omega_1, \omega_2, \dots, \omega_{d-1}, \omega_d] \quad (26)$$

We define the mapping to cluster allocation matrix  $Y^*$  from the scaled cluster assignment matrix  $Z^*$  as

$$Y^* = f^{-1}(Z^*) = \text{Diag}(Z^* Z^{*T})^{-1/2} Z^* \quad (27)$$

Then, by solving the following optimization problem

$$\begin{aligned} &\arg \min_{Y \in B^{n \times d}, R \in \mathbb{R}^{d \times d}} \|Y - Y^* R\|^2 \\ &\text{s.t. } Y 1_d = 1_n, \quad R^T R = I \end{aligned} \quad (28)$$

using spectral rotation [19], we obtain the cluster allocation matrix  $Y \in B^{n \times d}$  and the rotation matrix  $R$  simultaneously. We summarize our proposed LERDC model in Algorithm 1.

Algorithm 1

Procedure  $Y = \text{LERDC}(X)$

- 1) For each image  $x_i$ , we construct a set of local images  $\mathbb{S}_k(x_i)$ ,  $k=5$ ,  $i=1, \dots, n$  and obtain the image data matrix  $X_i$ .
- 2) Using [25], we obtain the image complement  $\widetilde{X}_i$  of  $\widetilde{X}_i$ , where  $\widetilde{X}_i = X_i H_k$ ;
- 3) Compute  $L_{ER_i}$ ,  $i=1, 2, \dots, n$  according to (20);
- 4) Compute  $L_{ER}$  using (22);
- 5) To obtain the optimal  $Z^* = [\omega_1, \omega_2, \dots, \omega_{d-1}, \omega_d]$ , we solve the eigenvalue decomposition problem in (25). The solution  $\omega_0=1_n$  corresponding to the eigenvalue  $\lambda_0=0$  is eliminated;
- 6) Compute  $Y^*$  by (27) and solve the optimization problem (28) using spectral rotation [19].

### 3. Comparison with Previous Clustering Models

To compare the optimization problem for the LERDC model presented in this paper, we briefly describe the optimization problem for the local learning-based clustering models proposed in the previous Ncut [19], SEC [18], and LDMGI [17].

### 3.1. Normalized Cut

In NCut [16],

$$\tilde{L} = I - D^{-1/2}GD^{-1/2} \quad (29)$$

is a Laplace matrix, where  $D$  is a diagonal matrix with the diagonal elements  $D_{ii} = \sum_{j=1}^n G_{ij}$  for  $i=1, \dots, n$  and  $G$  is a graph affinity matrix computed with the Gaussian function

$$G_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) & x_i \text{ and } x_j \text{ are } k\text{-nearest neighbours.} \\ 0 & \text{otherwise} \end{cases}$$

The  $k$ -nearest neighbors and the bandwidth  $\sigma$  are the clustering parameters.

### 3.2. SEC (Spectral Embedded Clustering)

In [18], the SEC model becomes the following optimization problem.

$$\arg \min_A \text{tr} [A^T (\sum_{i=1}^n L_i + \mu L_g) A]$$

where

$$L_g = H_n - (X^T (XX^T + \gamma I))^{-1} X \quad (30)$$

and

$$L_i = H_k - (X_i^T (X_i X_i^T + \gamma I))^{-1} X_i \quad (31)$$

The  $k$ -nearest neighbors and the regularization parameters  $\mu$  and  $\gamma$  are the clustering parameters.

**Table 1.** Clustering parameters in clustering models.

| Clustering Models | Clustering parameters | Total clustering parameters |
|-------------------|-----------------------|-----------------------------|
| NCut [19]         | $k, \sigma$           | 2                           |
| SEC [18]          | $k, \gamma, \mu$      | 3                           |
| LDMGI [17]        | $k, \lambda$          | 2                           |
| proposed LERDC    | $k, \lambda$          | 2                           |

### 3.3. LDMGI (Local Discriminant Model and Global Integration)

The LDMGI clustering model proposed in [17] is the following optimization problem

$$\arg \min_A \text{tr} \left[ A^T \left( \sum_{i=1}^n S_i L_i S_i^T \right) A \right],$$

where  $S_i$  is the selection matrix such as (10) and  $L_i$  is the same as (4). The clustering parameters are  $k$ -nearest neighbors and the regularization parameter  $\lambda$ .

As shown in Table 1, compared with previous clustering methods [17-19], we also find that the number of nearest neighbor images  $k$  is a clustering parameter in the proposed LERDC model.

In this study, we show in (2) that the clustering parameter  $\lambda$  involved to deal with the SSS problem of LDA is a restriction. Also, when we carefully compare (30) and (31) with (4), we see that the objective of the regularization parameter  $\gamma$  is equal to  $\lambda$ .

In the SEC model, parameter  $\gamma$  was included because of the singularity problem in the scattering matrices for high-dimensional image datasets.

## 4. Experiment

We compare the performance of the proposed LERDC model with K-means, NCut [19], SEC [18], and LDMGI [17] clustering models using twelve benchmark image datasets.

A comparative study about the performances of all clustering models is performed by clustering accuracy (ACC) and normalized mutual information (NMI) performance measures.

The initialization [19] is affected to the results for all clustering algorithms.

In order to reduce statistical variation, the random initialization was repeated 20 times during the simulation.

The average of the results of repeated runs for each clustering model is estimated by  $\text{ACC} \pm \text{std}$  and  $\text{NMI} \pm \text{std}$ , where std is the standard deviation from the mean.

The clustering performance is the best when ACC and NMI are the best in 20 runs.

#### 4.1. Image Dataset

To evaluate the performance of the new image clustering model, we used 10 870 samples from 10 benchmark image datasets.

We used 20 objects from the Caltech101 image database [27] containing an image of objects belonging to 101 classes with 40 ~ 800 images per class.

The Caltech faces dataset [28] involves 450 face images for 27 people under different lighting, expressions and backgrounds, where 380 face images for 19 people were used.

The COIL-20 database has 1 200 images for 20 objects, each consisting of 60 images captured from different views [29].

Corel database [31] divided the images from [30] into 80 concept groups, where 20 concept groups were used, such as aircraft, car, doll, flag, boat, steam engine and train, according

to the category of objects.

The JAFFE database used 150 images of different facial expressions of 10 female models [32].

The MPEG7 shape image database [33, 34] used 540 images aligned by geometric transformation using [25].

The ORL image database [35] contains 400 images of different time, facial expression and face detail for 40 distinct subjects.

The Pointing 04 face database contains images of 15 people covering a wide range of poses with various skin colors, with or without glasses [36].

The Scenes image database includes forest, mountain and city centers, which are composed of eight different categories for both urban and natural [37].

The Yalefaces database contains 150 images of 15 subjects with different facial expressions and illumination conditions and there are 10 images per subject [38].

In Table 2, detailed descriptions of sample size, the number of classes, original image size and feature dimension for each image dataset are given.

**Table 2.** Description of image datasets and image feature dimension.

| Dataset       | Sample size | Image size | Feature dimension | Class number |
|---------------|-------------|------------|-------------------|--------------|
| Caltech101    | 1 000       | 260×300    | 1 200             | 20           |
| Caltech faces | 380         | 896×592    | 2 072             | 19           |
| COIL-20       | 1 200       | 256×256    | 1 024             | 20           |
| COREL         | 2 400       | 120×80     | 1 600             | 20           |
| JAFFE         | 150         | 256×256    | 676               | 10           |
| MPEG7         | 540         | 300×250    | 200               | 30           |
| ORL           | 400         | 112×92     | 644               | 40           |
| Pointing04    | 2 250       | 384×288    | 1 120             | 15           |
| Scenes        | 2 400       | 256×256    | 676               | 8            |
| Yelefaces     | 150         | 320×243    | 1 024             | 15           |

#### 4.2. Performance Evaluation of LERDC

In this paper, we used clustering accuracy (ACC) and NMI to evaluate the performance of the classification method.

##### 4.2.1. Clustering Accuracy (ACC)

For the  $i$ th image, let  $q_i$  be the result of clustering by clustering algorithm and  $p_i$  be the ground truth label. The ACC is defined as follows

$$ACC = \frac{\sum_{i=1}^n \delta(p_i, \text{map}(q_i))}{n} \quad (32)$$

where  $n$  is the total number of images,  $\text{map}(q_i)$  is the optimal mapping function that permutes clustering labels to match the ground truth labels, and  $\delta(p, q) = 1$  if  $p = q$ .

Using the Kuhn-Munkres algorithm [39], we obtain the optimal mapping. The larger ACC indicates better clustering performance.

However, when the number of classes in the image dataset is large, it is tedious to compute the ACC using the Kuhn-

Munkres algorithm [39].

Zhu et al. improved the Kuhn-Munkres algorithm by choosing a reasonable size of permutation in class labels to maximize ACC [40].

#### 4.2.2. Normalized Mutual Information (NMI)

Another widely used measure to evaluate clustering results is NMI. For two arbitrary variables  $S$  and  $T$ , NMI is defined as follows

$$NMI(S, T) = \frac{I(S, T)}{\sqrt{H(S)H(T)}} \quad (33)$$

where  $I(S, T)$  represents the mutual information between  $S$  and  $T$ .

$H(S)$  and  $H(T)$  are the entropies of  $S$  and  $T$ , respectively. It is obvious that  $NMI(S, T)$  is 1 if  $S$  is identical to  $T$ , and 0 if  $S$  is not similar to  $T$ .

Let  $t_h$  be the total samples in the  $h$ th ground truth class ( $1 \leq h \leq d$ ) and  $t_m$  be the number of samples in cluster  $D_m$  ( $1 \leq m \leq d$ ) obtained by using the clustering algorithms.

NMI is defined as

$$NMI = \frac{\sum_{h=1}^d \sum_{m=1}^d t_{h,m} \log(n \cdot t_{h,m} / t_h t_m)}{\sqrt{\left(\sum_{h=1}^d t_h \log(t_h / n)\right) \left(\sum_{m=1}^d t_m \log(t_m / n)\right)}} \quad (34)$$

where  $t_{h,m}$  is the number of samples belonging to both the  $h$ th ground truth class and the cluster  $D_m$ . When the NMI is large, the clustering results are good.

#### 4.3. Discussion About Clustering Parameters

For comparison with all clustering models, we have  $k = 5$  as in [17, 18].

The optimal value of  $\sigma$  and  $\lambda$  were determined from the set  $\{10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6, 10^8\}$  as in [17]. Also, with  $\mu = 1$ , the optimal value of  $\gamma$  was determined from the set  $\{10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6, 10^8\}$ .

In the clustering models based on local learning, the parameter  $k$  affects the clustering performance [17].

For most image data sets, we observed that the optimal value of parameter  $k$  in the proposed LERDC model from the set  $\{2, 3, 4, 5, 6, 7, 8, 10\}$  is 5.

In [41], the neighborhood size  $k$  was defined from the set  $\{5, 10, 20, 50, 100\}$ .

**Table 3.** Performance comparison of proposed LERDC with previous clustering models (best ACC and mean ACC $\pm$ std).

| Dataset       | best ACC |           |          |            |       | mean ACC $\pm$ std |                |                |                |                |
|---------------|----------|-----------|----------|------------|-------|--------------------|----------------|----------------|----------------|----------------|
|               | K-meas   | NCut [19] | SEC [18] | LDMGI [17] | LERDC | K-meas             | NCut [19]      | SEC [18]       | LDMGI [17]     | LERDC          |
| JAFFE         | 83.2     | 95.8      | 96.0     | 96.8       | 97.5  | 75.8 $\pm$ 6.2     | 94.8 $\pm$ 0.8 | 95.0 $\pm$ 0.6 | 95.8 $\pm$ 0.6 | 96.8 $\pm$ 0.4 |
| COIL-20       | 66.9     | 83.1      | 81.2     | 84.3       | 86.4  | 62.0 $\pm$ 3.5     | 82.2 $\pm$ 0.7 | 78.8 $\pm$ 1.2 | 83.5 $\pm$ 0.4 | 85.4 $\pm$ 0.5 |
| ORL           | 73.3     | 78.6      | 82.9     | 82.3       | 83.4  | 67.1 $\pm$ 4.0     | 77.5 $\pm$ 1.0 | 82.0 $\pm$ 0.6 | 81.6 $\pm$ 0.3 | 82.7 $\pm$ 0.3 |
| Pointing04    | 52.8     | 64.7      | 65.0     | 67.0       | 69.2  | 49.3 $\pm$ 3.2     | 63.8 $\pm$ 0.4 | 63.2 $\pm$ 0.8 | 65.3 $\pm$ 1.2 | 68.3 $\pm$ 0.7 |
| Yalefaces     | 60.7     | 70.5      | 73.7     | 66.7       | 68.4  | 56.4 $\pm$ 4.0     | 69.7 $\pm$ 0.5 | 73.2 $\pm$ 0.3 | 65.5 $\pm$ 0.8 | 67.2 $\pm$ 0.4 |
| MPEG7         | 72.0     | 68.7      | 73.6     | 67.9       | 70.6  | 66.7 $\pm$ 3.4     | 67.2 $\pm$ 1.0 | 72.8 $\pm$ 0.5 | 65.8 $\pm$ 1.5 | 69.8 $\pm$ 0.5 |
| COREL         | 37.4     | 36.0      | 35.3     | 38.6       | 38.3  | 34.8 $\pm$ 2.1     | 34.8 $\pm$ 0.7 | 34.6 $\pm$ 0.4 | 37.4 $\pm$ 1.0 | 37.8 $\pm$ 0.3 |
| Caltech faces | 26.8     | 28.0      | 26.9     | 27.9       | 28.8  | 25.0 $\pm$ 1.2     | 26.8 $\pm$ 0.5 | 26.4 $\pm$ 0.3 | 27.2 $\pm$ 0.3 | 28.0 $\pm$ 0.4 |
| Scenes        | 25.7     | 24.9      | 27.2     | 27.1       | 26.3  | 23.8 $\pm$ 2.4     | 24.2 $\pm$ 0.4 | 25.9 $\pm$ 0.4 | 26.5 $\pm$ 0.3 | 25.8 $\pm$ 0.4 |
| Caltech101    | 18.3     | 17.5      | 18.3     | 19.2       | 19.5  | 16.9 $\pm$ 1.3     | 16.5 $\pm$ 0.7 | 17.8 $\pm$ 0.2 | 18.6 $\pm$ 0.5 | 19.0 $\pm$ 0.3 |
| Overall mean  | 51.7     | 56.8      | 58.0     | 57.8       | 58.8  | 47.8 $\pm$ 3.1     | 55.8 $\pm$ 0.6 | 57.0 $\pm$ 0.5 | 56.7 $\pm$ 0.7 | 58.1 $\pm$ 0.4 |

**Table 4.** Performance comparison of proposed LERDC with previous clustering models (best NMI and mean NMI  $\pm$  std).

| Dataset       | best NMI |           |          |            |       | mean NMI $\pm$ std |                |                |                |                |
|---------------|----------|-----------|----------|------------|-------|--------------------|----------------|----------------|----------------|----------------|
|               | K-means  | NCut [19] | SEC [18] | LDMGI [17] | LERDC | K-means            | NCut [19]      | SEC [18]       | LDMGI [17]     | LERDC          |
| JAFFE         | 90.8     | 95.4      | 95.7     | 96.5       | 97.8  | 85.3 $\pm$ 2.7     | 94.3 $\pm$ 0.5 | 94.9 $\pm$ 0.3 | 95.6 $\pm$ 0.8 | 96.7 $\pm$ 0.6 |
| COIL-20       | 80.2     | 90.3      | 87.6     | 95.2       | 95.8  | 76.8 $\pm$ 1.5     | 89.2 $\pm$ 0.8 | 87.0 $\pm$ 0.4 | 94.8 $\pm$ 0.2 | 95.3 $\pm$ 0.2 |
| ORL           | 87.3     | 89.6      | 93.1     | 93.7       | 94.3  | 84.5 $\pm$ 0.9     | 87.7 $\pm$ 0.5 | 92.5 $\pm$ 0.3 | 92.8 $\pm$ 0.4 | 93.5 $\pm$ 0.3 |
| Pointing04    | 53.5     | 78.3      | 74.5     | 80.4       | 81.2  | 51.8 $\pm$ 1.5     | 77.5 $\pm$ 0.4 | 73.5 $\pm$ 0.3 | 78.7 $\pm$ 0.6 | 79.8 $\pm$ 0.5 |
| Yalefaces     | 67.2     | 72.7      | 73.4     | 72.8       | 72.7  | 65.2 $\pm$ 1.8     | 71.3 $\pm$ 0.7 | 72.8 $\pm$ 0.3 | 71.8 $\pm$ 0.8 | 71.5 $\pm$ 0.3 |
| MPEG7         | 86.1     | 86.7      | 87.8     | 86.5       | 87.6  | 83.7 $\pm$ 1.3     | 85.8 $\pm$ 0.5 | 86.3 $\pm$ 0.7 | 85.3 $\pm$ 0.5 | 86.5 $\pm$ 0.5 |
| COREL         | 41.2     | 39.8      | 39.4     | 41.8       | 42.5  | 39.7 $\pm$ 0.6     | 38.6 $\pm$ 0.5 | 38.4 $\pm$ 0.5 | 40.4 $\pm$ 0.8 | 41.3 $\pm$ 0.6 |
| Caltech faces | 31.7     | 35.6      | 36.8     | 35.1       | 36.4  | 30.3 $\pm$ 1.2     | 34.3 $\pm$ 1.0 | 35.5 $\pm$ 0.8 | 34.5 $\pm$ 0.4 | 35.2 $\pm$ 0.7 |
| Scenes        | 8.9      | 10.6      | 11.2     | 13.6       | 13.2  | 8.5 $\pm$ 0.2      | 10.1 $\pm$ 0.3 | 10.5 $\pm$ 0.5 | 12.9 $\pm$ 0.3 | 12.4 $\pm$ 0.6 |
| Caltech101    | 21.3     | 19.8      | 20.3     | 20.1       | 20.4  | 18.5 $\pm$ 1.0     | 18.9 $\pm$ 0.6 | 19.1 $\pm$ 0.6 | 19.7 $\pm$ 0.2 | 19.5 $\pm$ 0.4 |
| Overall mean  | 56.8     | 61.9      | 62.0     | 63.6       | 64.2  | 54.4 $\pm$ 1.3     | 60.8 $\pm$ 0.5 | 61.1 $\pm$ 0.4 | 62.6 $\pm$ 0.5 | 63.2 $\pm$ 0.4 |

#### 4.4. Performance Comparison

Tables 3 and 4 shows the image clustering performances of K-means, Ncut [19], SEC [18], LDMGI [17] and proposed LERDC models.

The LERDC model achieved a comparable clustering performance as that of the near competitors Ncut [19], SEC [18] and LDMGI [17] by means of ACC and Ncut [19], LDMGI [18] in terms of NMI.

Overall performance of Ncut [19], SEC [18], LDMGI [17] and proposed LERDC is 55.8%, 57%, 56.7% and 58.1% in terms of mean ACC, 60.8%, 61.1%, 62.6% and 63.2% by mean NMI.

The overall cluster performance of the proposed LERDC model is comparable to that of the previous LDMGI model.

However, on image datasets such as Scenes, Caltech101, and Caltech faces, it is shown that the performance of all clustering models is not optimal.

In this study, all clustering models used image features based on pixel intensity values.

The image features were obtained using linear interpolation to adjust the size of the original images as shown in [17].

To improve the clustering performance on these challenging image datasets, we used an optimization image descriptor [42].

The clustering result for Pointing 04 in our work is different from that reported in [17] because the 1120-dimensional feature vector is not normalized.

We also used [25] to align images in the MPEG7 shape image database by geometric transformation from control pair

points.

In the proposed LERDC model, we also studied the effect of rescaling after normalization of the exponential regularized matrix.

We rescaled the matrix obtained from

$$\exp\left(-\left(\tilde{\tilde{X}}_i^T \tilde{\tilde{X}}_i + \lambda I\right) \tilde{\tilde{X}}_i^T \tilde{\tilde{X}}_i / \left\| \left(\tilde{\tilde{X}}_i^T \tilde{\tilde{X}}_i + \lambda I\right) \tilde{\tilde{X}}_i^T \tilde{\tilde{X}}_i \right\| \right)$$

in (20) by multiplying it with Frobenius norm of

$$\left(\tilde{\tilde{X}}_i^T \tilde{\tilde{X}}_i + \lambda I\right) \tilde{\tilde{X}}_i^T \tilde{\tilde{X}}_i.$$

The experimental results show that after normalization, the effect of rescaling is not significant, and comparable clustering results were observed without rescaling the exponential regularized matrix after normalization.

We now discuss how and when the proposed LERDC model performs better as compared with the near competitor LDMGI clustering model [17].

Theoretically, these two clustering models utilized the discriminant performance of LDA at local level.

The LDMGI is based on RDA [10] whereas the proposed LERDC is based on EDA [20].

In [20], it was shown that the discriminant information of LDA is not lost even in EDA.

Thus, in the proposed LERDC model, discriminant performance of LDA is not decreased.

Furthermore, even if we project the scatter matrices into an

exponential region, there is no significant change in the discriminant information of LDA.

That is why the LERDC can compare clustering performance with the LDMGI model.

As shown in Table 1, the proposed LERDC is a local learning-based clustering model with parameters such as the LDMGI clustering model [17], when compared with previous local learning-based clustering methods [17-19].

#### 4.5. Performance Variation of Previous Clustering Models

In this study, the optimal value of the clustering parameters  $\lambda$ ,  $\gamma$ ,  $\sigma$  were selected from the set  $\{10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6, 10^8\}$  as in [17].

Therefore, a total of nine observations for each image database are made in the previous clustering models (Ncut [19], SEC [18] and LDMGI [17]).

Tuning of clustering parameters for optimal clustering performance in all previous clustering models [17-19] is required for each image database.

However, in contrast, such tuning is not required in the proposed LERDC clustering model.

## 5. Conclusion

We propose LERDC as a new local learning-based image clustering model.

In the LERDC model, the local scatter matrices in RDA are projected into the exponential domain.

Therefore, the small sample size problem of LDA is treated at the local level with the local scatter matrix added regularization parameter.

In the proposed LERDC model,  $k$ -nearest neighbors and  $\lambda$  is the clustering parameter as compared with existing local learning based clustering approaches.

Experimental results on ten reference image databases show that the proposed LERDC model is able to compare clustering performance with the RDA-based nearby competitor LDMGI model [17].

The clustering performance is comparable because the discriminant information of LDA is not lost even if the scatter matrix is projected onto the exponential domain.

We conclude that the proposed LERDC model is able to compare clustering performance with the same parameters as existing local learning based clustering methods.

In the future, we will investigate how to determine the optimal parameters and the appropriate discriminative image feature extraction methods to improve the clustering performance, together with using global information in addition to local information to effectively learn nonlinear manifolds in image data.

## Abbreviations

|           |                                                       |
|-----------|-------------------------------------------------------|
| ACC       | Accuracy                                              |
| DisKmeans | Discriminant K-means                                  |
| EDA       | Exponential Discriminant Analysis                     |
| ELDE      | Exponential Local Discriminant Embedding              |
| LERDC     | Local Exponential Regularized Discriminant Clustering |
| LERDM     | Local Exponential Regularized Discriminant Model      |
| LDA       | Linear Discriminant Analysis                          |
| LDMGI     | Local Discriminant Models and Global Integration      |
| Ncut      | Normalized Cutting                                    |
| NMI       | Normalized Mutual Information                         |
| PCA       | Principal Components Analysis                         |
| RDA       | Regularized Discriminant Analysis                     |
| RLDA      | Regularized Linear Discriminant Analysis              |
| SEC       | Spectral Embedded Clustering                          |
| SSS       | Small-sample-size                                     |
| std       | Standard Deviation                                    |

## Author Contributions

Kwang Jun Pak is the sole author. The author read and approved the final manuscript.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- [1] Adil, M., Abid, M., Khan, A, Q., et al. Exponential discriminant analysis for fault diagnosis [J]. Neurocomputing. 2016, 171, 1344-1353.
- [2] Zhu, J., et al. Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data [J]. IEEE Transactions on Industrial Informatics. 2017, 13(4), 1877-1885.
- [3] Yu, W., Zhao, C. Sparse exponential discriminant analysis and its application to fault diagnosis [J]. IEEE Transactions on Industrial Informatics. 2018, 65(7), 5931-5940.
- [4] Shuiwang, Ji., Jieping, Ye. Generalized Linear Discriminant Analysis: A Unified Framework and Efficient Model Selection. IEEE Transactions on Neural Network. Vol, 19., No, 10., October, 2008, 1768-1782.
- [5] la, Torre, F, D., Kanade, T. Discriminative cluster analysis. Int. Conf, on Machine Learning (ICML). 2006, 241-248.
- [6] Ding, C., Li, T. Adaptive dimension reduction using discriminant analysis and k-means clustering. Int. Conf, on Machine Learning (ICML). 2007, 521-528.

- [7] Ye, J., Zhao, Z., Wu, M. Discriminative k-means for clustering. Neural Information Processing System (NIPS). 2008, 1649–1656.
- [8] Fukunaga, K. Feature extraction and linear mapping for classification., in, Introduction to statistical pattern recognition. Ed, Rheinboldt, W. (Academic., San, Diego., CA. 1990, 2nd, edn).
- [9] Liu, K., et al. A generalized optimal set of discriminant vectors. Pattern Recognit. 1992, 25, 731–739.
- [10] Friedman, J, H. Regularized discriminant analysis. J, Am, Stat, Assoc. 1989, 84(405), 165–175.
- [11] Tebbens, J, D., Schlesinger, P. Improving implementation of linear discriminant analysis for the high dimension/small sample size problem. Comput, Stat, Data, Anal. 2007, 52, 423–437.
- [12] Howland, P., Wang, J., Park, H. Solving the small sample size problem in face recognition using generalized discriminant analysis. Pattern Recognit. 2006, 39, 277–287.
- [13] Wu, M., Scholkopf, B. A local learning approach for clustering. Neural Information Processing System (NIPS). 2006, 1529–1536.
- [14] Saul, L, K., Roweis, S, T. Think globally fit locally: unsupervised learning of low dimensional manifolds. J, Mach, Learn, Res. 2003, 4, 119–155.
- [15] Sugiyama, M. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. J, Mach, Learn, Res. 2007, 8, 1027–1061.
- [16] Shi, J., Malik, J. Normalized cuts and image segmentation. IEEE Trans, Pattern Anal, Mach, Intell. 2000, 22(8), 888–905.
- [17] Yang, Y., Xu, D., Nie, F., Yan, S., Zhuang, Y. Image clustering using local discriminant models and global integration. IEEE Trans, Image Process. 2010, 19(10), 2761–2773.
- [18] Nie, F., Zeng, Z., Tsang, I, W., Xu, D., Zhang, C. Spectral embedded clustering: a framework for in-sample and out-of-sample spectral clustering. IEEE Trans, Neural Netw. 2011, 22(11), 1796–1808.
- [19] Yu, S, X., Shi, J. Multiclass spectral clustering. International Conference on Computer Vision (ICCV), 2003, 313–339.
- [20] Zhang, T., Fang, B., Tang, Y, Y., Shang, Z., Xu, B. Generalized discriminant analysis: a matrix exponential approach. IEEE Trans, Syst, Man Cybern, B. 2010, 40(1), 186–197.
- [21] Dornaika, F., Bosaghzadeh, A. Exponential local discriminant embedding and its application to face recognition. IEEE Trans, Cybern. 2013, 43(3), 921–934.
- [22] Nasir, Ahmed. Image clustering using exponential discriminant analysis. IET Comput, Vis. 2015, Vol, 9, Iss, 1, 1–12.
- [23] Higham, N, J. Functions of matrices: theory and computation. Society for Industrial and Applied Mathematics. Philadelphia, PA, USA, 2008.
- [24] Williamson, R. Introduction to differential equations. Prentice-Hall, Englewood Cliffs, NJ. 1986.
- [25] MATLAB, version 7.10.0 (R2011a). The MathWorks Inc, Natick, MA. 2011.
- [26] Duda, R., Hart, P., Stroke, D. Pattern classification. Wiley. 2007, 2nd, edn.
- [27] Li, F., Andreetto, M., Ranzato, M, A. Caltech101. Computational Vision Group, California Institute of Technology. 2003. Available at, <http://www.vision.caltech.edu/ImageDatasets/Caltech101>
- [28] Weber, M. Frontal face dataset. Computational Vision Group, California Institute of Technology. 2012. Available at, <http://www.vision.caltech.edu/html-files/archive.html>
- [29] Nene, S, A., Nayar, S, K., Murase, H. Columbia Object Image Library (Coil-20). 1996.
- [30] Xu, D., Yan, S., Tao, D., Lin, S., Zhang, H., Huang, T. Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis. IEEE Trans, Circuits Syst, Video Technol. 2008, 18(1), 36–47.
- [31] Tao, D. The COREL Database for content based image retrieval. 2009. Available at, <https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval>
- [32] Lyons, M, J., Budynek, J., Akamatsu, S. Automatic classification of single facial images. IEEE Trans, Pattern Anal, Mach, Intell. 1999, 21(12), 1357–1362.
- [33] Bicego, M., Murino, V., Figueiredo, M, A. Similarity-based classification of sequences using hidden Markov models. Pattern Recognit. 2004, 37(12), 2281–2291.
- [34] Thakoor, N., Gao, J., Jung, S. Hidden Markov model-based weighted likelihood discriminant for 2-d shape classification. IEEE Trans, Image Process. 2007, 16(11), 2707–2719.
- [35] Samaria, F., Harter, A. Parameterisation of a stochastic model for human face identification. Proc, 2nd IEEE Workshop on Applications of Computer Vision. 1994, 138–142.
- [36] Gourier, N., Hall, D., Crowley, J, L. Estimating face orientation from robust detection of salient facial features. Int, Workshop on Visual Observation of Deictic Gestures (ICPR). 2004, 17–25.
- [37] Oliva, A., Torralba, A. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int, J, Comput, Vis. 2001. 42(3), 145–175.
- [38] Belhumeur, P, N., Hespanha, J, P., Kriegman, D, J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans, Pattern Anal, Mach, Intell. 1997, 19(7), 711–720.
- [39] Papadimitriou, C, H., Steiglitz, K. Combinatorial optimization: algorithms and complexity. Dover, New York. 1998.
- [40] Zhu, H., Zhou, M., Alkins, R. Group role assignment via a Kuhn-Munkres algorithm-based solution. IEEE Trans, Syst, Man Cybern, A. 2012, 42(3), 739–750.

- [41] Wang, F., Zhang, C. Clustering with local and global regularization. *IEEE Trans, Knowl, Data Eng.* 2009, 21(12), 1665–1678.
- [42] Ahmed, N., Jalil, A. Multimode image clustering using optimal image descriptor. *IEICE Trans, Inf, Syst.* 2014, E97-D (4), 743–751.