Ahmad Khoder
University of the Basque Country
San Sebastian, Spain
E-mail: ahmadkhoder10@hotmail.com

Fadi Dornaika
University of the Basque Country
IKERBASQUE Research Foundation
San Sebastian, Spain
E-mail: fadi.dornaika@ehu.es

Feature Extraction and Selection via Robust Discriminant Analysis and Class Sparsity

1. Introduction

The main goal of discriminant embedding is to extract features that can be compact and informative representations of the original set of features.
We introduce a hybrid scheme for linear feature extraction for supervised multiclass classification. We introduce a unifying criterion that is able to retain the advantages of robust sparse LDA and Interclass sparsity. Thus, the estimated transformation includes two types of discrimination which are the inter-class sparsity and robust Linear Discriminant Analysis with feature selection. In order to optimize the proposed objective function, we deploy an iterative alternating minimization scheme for estimating the linear transformation and the orthogonal matrix. The introduced scheme is generic in the sense that it can be used for combining and tuning many other linear embedding methods.

2. Proposed Method (two Variants)

2.1. Objective function

- **First Term: LDA Criterion.**
- **Second Term: Inter-class Sparsity.**
- **Third Term: PCA Variant (Reconstruction Term)**

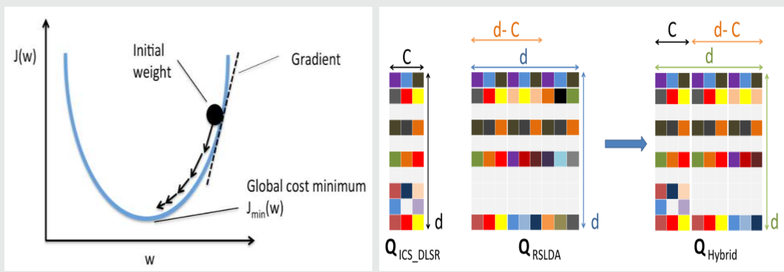
$$f(\mathbf{Q},\mathbf{P}) = Tr\left(\mathbf{Q}^T\mathbf{S}\mathbf{Q}\right) + \lambda_1 \sum_{i=1}^{C} \|\mathbf{Q}^T\mathbf{X}_i\|_{2,1} + \lambda_2 \|\mathbf{X} - \mathbf{P}\mathbf{Q}^T\mathbf{X}\|_2^2 \quad s.t. \quad \mathbf{P}^T\mathbf{P} = \mathbf{I}$$

- Encapsulates two different types of discrimination. Namely: robust LDA and inter-class sparsity

-A fine tuning tool for linear models. Our proposed method is generic in the sence it can be used to fine tune the solution obtained by different linear methods.

2.2. Optimization + initialization Scheme

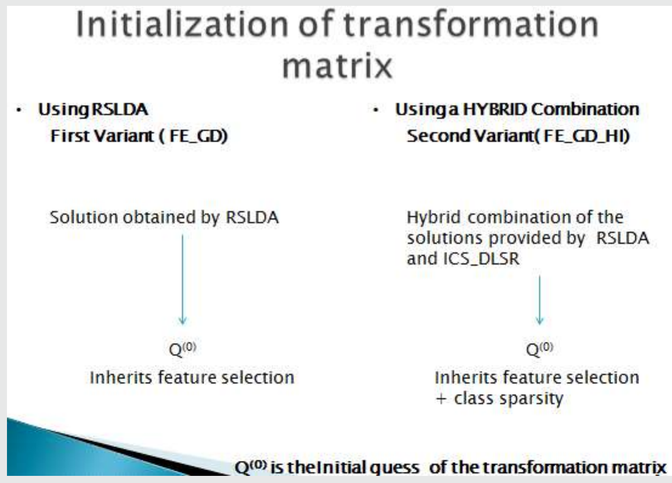
- **According to our optimization problem, there is no analytical solution for the problem. Thus, we adopted an alternating scheme for solving the two matrices.**
- **The linear transformation is updated using a gradient descent approach..**



- Hybrid initialization scheme.
- Derived transformation space, inherits feature ranking and class sparsity.

2.3. Two proposed variants

- Two variants are proposed, each one inherits the advantages of different methods.



3. Experimental Results

We have conducted our experiments over the following public datasets in addition to a large-scale dataset: USPS, digits dataset, Honda, COIL20 object dataset, Extended Yale B, FEI dataset, and the large scale MNIST dataset consisting of 60,000 images. The two proposed variants have been compared with the following methods: K-nearest neighbors (KNN), Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), Local Discriminant Embedding (LDE), PCE (unsupervised method), ICS DLSR and Robust sparse LDA (RSLDA). All experiments and all compared methods used the same conditions in order to guarantee a fair comparison. For each compared embedding method, the whole dataset is randomly split into a training part and a test part. For each dataset and for each embedding method, an embedding is computed first using the training part of the data. The training and test data are then projected using the estimated embedding. Classification of the test data is then performed using the Nearest Neighbor classifier (NN).

Table II: Mean classification accuracies (%) of different methods on the tested datasets.

Dataset \ Method	Training Samples	KNN	SVM	LDA	LDE	PCE	ICS_DLSR	RSLDA	FE_GD	FE_GD_HI
USPS	30	87.01	88.21	84.91	83.54	72.01	88.46	89.45	89.50	**90.29**
	40	88.56	90.40	86.19	85.3	72.30	90.16	91.11	**91.81**	91.46
	55	90.51	92.09	88.64	87.16	73.32	91.25	92.65	**93.07**	92.87
	65	91.76	93.16	89.29	88.58	74.11	91.53	92.89	**93.71**	93.49
Ext. Yale B	10	69.80	73.85	82.32	79.92	86.39	86.56	86.79	87.10	**88.42**
	15	75.20	80.02	86.76	83.77	89.23	89.53	89.93	90.04	**91.21**
	20	80.24	85.79	90.7	88.44	92.19	93.14	93.59	93.75	**93.81**
	25	82.24	89.03	92.17	90.43	93.35	94.50	94.92	95.02	**95.09**
Honda	10	64.12	71.32	65.95	65.74	61.86	70.79	69.90	70.16	**72.14**
	20	77.69	83.60	79.39	79.25	75.33	82.95	83.03	83.60	**84.64**
	30	84.78	89.09	85.84	86.24	82.55	88.20	89.04	89.41	**90.12**
	50	91.36	94.15	92.28	92.34	90.03	93.53	94.13	94.53	**95.10**
FEI	5	88.98	91.18	92.60	90.67	86.04	92.16	93.19	93.81	**94.58**
	6	90.35	92.93	94.18	92.15	88.73	93.65	94.25	94.75	**95.08**
	7	92.60	94.31	95.60	94.26	91.09	95.20	95.66	96.20	**96.29**
	8	94.27	95.23	96.03	95.57	93.20	96.17	96.43	**96.97**	96.40
COIL20	20	94.58	97.65	96.19	95.00	94.87	**98.04**	96.73	96.89	97.66
	25	95.79	98.22	97.07	96.12	95.99	98.22	97.74	97.89	**98.59**
	30	96.65	98.70	97.81	97.01	97.49	98.75	98.26	98.52	**99.08**
	35	97.14	98.81	98.15	97.42	98.11	99.12	98.68	98.80	**99.39**
MNIST	1000	91.75	97.58	85.74	93.22	93.77	98.02	97.95	98.21	**98.33**

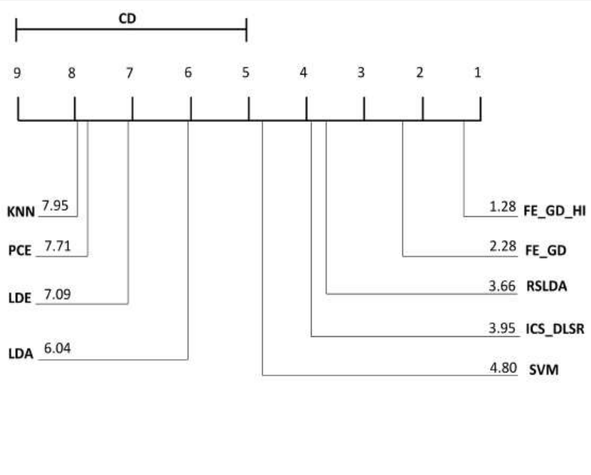
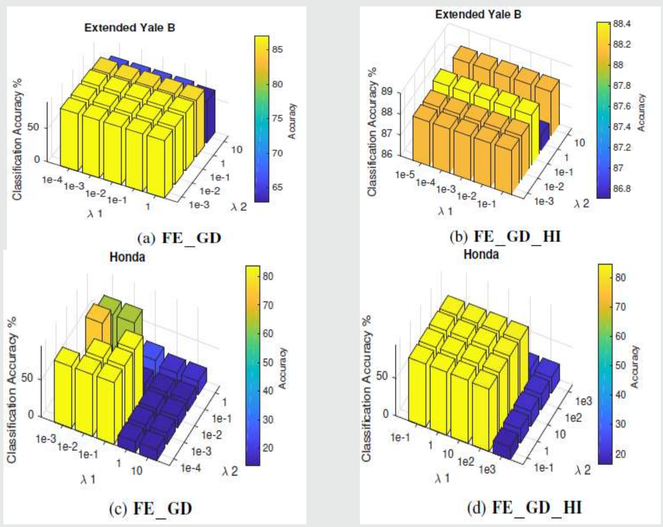


Figure 2: Statistical Analysis - CD diagram.



Statistical Analysis showing that our proposed variants outperformed competitors. Figure on the right shows the effect of the parameter sensitivity over the classification performance.

4. Conclusions

We introduced a novel criterion in order to obtain a discriminant linear transformation. The obtained linear transformation encapsulates two different types of discrimination which are the inter-class sparsity in addition to robust LDA. We deployed an iterative alternating minimization scheme to estimate the linear transformation and the orthogonal matrix associated with the robust LDA. The linear transformation is efficiently updated via the steepest descent gradient technique.
We proposed two initialization variants for the linear transformation. The first scheme sets the initial solution to the linear transformation obtained by robust sparse LDA method (RSLDA). The second variant initializes the solution to a hybrid combination of the two transformations obtained by RSLDA and ICS DLSR methods. The two variants of the proposed method have demonstrated superiority over competing methods and led to a more discriminative transformation matrix. The proposed framework is generic in the sense it allows the combination and tuning of other linear discriminant embedding methods. By combining these two types of constraints, and proposing the hybrid initialization scheme, the proposed method achieved higher classification rates than many competing methods.