



Article An Optimization Technique for Linear Manifold Learning-Based Dimensionality Reduction: Evaluations on Hyperspectral Images

Ümit Öztürk ^{1,*} and Atınç Yılmaz ²

- ¹ Department of Computer Engineering, Institute of Graduate Studies, Beykent University, Taksim, 34433 İstanbul, Turkey
- ² Department of Computer Engineering, Faculty of Engineering-Architecture, Beykent University, Ayazağa, 34396 İstanbul, Turkey; atincyilmaz@beykent.edu.tr
- * Correspondence: 160821006@student.beykent.edu.tr

Abstract: Manifold learning tries to find low-dimensional manifolds on high-dimensional data. It is useful to omit redundant data from input. Linear manifold learning algorithms have applicability for out-of-sample data, in which they are fast and practical especially for classification purposes. Locality preserving projection (LPP) and orthogonal locality preserving projection (OLPP) are two known linear manifold learning algorithms. In this study, scatter information of a distance matrix is used to construct a weight matrix with a supervised approach for the LPP and OLPP algorithms to improve classification accuracy rates. Low-dimensional data are classified with SVM and the results of the proposed method are compared with some other important existing linear manifold learning methods. Class-based enhancements and coefficients proposed for the formulization are reported visually. Furthermore, the change on weight matrices, band information, and correlation matrices with *p*-values are extracted and visualized to understand the effect of the proposed method. Experiments are conducted on hyperspectral imaging (HSI) with two different datasets. According to the experimental results, application of the proposed method with the LPP or OLPP algorithms outperformed traditional LPP, OLPP, neighborhood preserving embedding (NPE) and orthogonal neighborhood preserving embedding (ONPE) algorithms. Furthermore, the analytical findings on visualizations show consistency with obtained classification accuracy enhancements.

Keywords: linear manifold learning; out-of-sample; classification; weight matrix; artificial intelligence

1. Introduction

Dimension reduction (DR) techniques are used to obtain new feature sets in a subspace from high-dimensional original data. This helps to omit the redundant data and noise from the input and provides advantages, especially in terms of computational time and data storage [1,2]. Manifold learning is a special kind of technique for DR, which assumes that there is a low-dimensional manifold on high-dimensional data and its purpose is to find this manifold [3]. Manifold learning is applied to different application domains such as face recognition, medical image segmentation, and object recognition [4]. Furthermore, an increase in the availability of hyperspectral imaging (HSI) data and advanced computing has led to the popularity of manifold learning techniques in this research area [5].

Manifold learning aims to preserve local or global geometric relations and not lose the graph topology of high-dimensional data [5–7]. Isometric feature mapping (Isomap), locally linear embedding (LLE), and Laplacian eigenmaps (LE) are the most widely utilized nonlinear manifold learning methods. While Isomap is a global approach [8], LE [9] and LLE [10] retain the local geometry of data. Local approaches are initiated by obtaining the nearest neighborhood of each data point instead of retaining overall topology. There are more robust domain-specific manifold learning methods in the literature. In this study, an



Citation: Öztürk, Ü.; Yılmaz, A. An Optimization Technique for Linear Manifold Learning-Based Dimensionality Reduction: Evaluations on Hyperspectral Images. *Appl. Sci.* 2021, *11*, 9063. https:// doi.org/10.3390/app11199063

Academic Editors: Akemi Galvez Tomida and Andres Iglesias Prieto

Received: 2 August 2021 Accepted: 24 September 2021 Published: 28 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). effective local manifold representation that makes use of spatial and spectral information is proposed [11].

Traditional nonlinear manifold learning algorithms take all previous samples and new samples as input for the learning process. This is known as an out-of-sample problem and is not practical for classification purposes including face recognition and HSI tasks [3]. Nevertheless, there are some studies on generalizations and extensions of out-of-sample data for nonlinear manifold learning that use polynomial mapping [3], the Nyström approach [12], kernel Hilbert space [13], and sparse grid functions [14]. Additionally, linearized manifold learning algorithms can make embedding generalizations from the learned linear projections for new samples.

There are various linear manifold learning algorithms in the literature. Locality preserving projection (LPP) and neighborhood preserving embedding (NPE) linearize the LE and LLE, respectively [15,16]. Linearization provides applicability for out-of-sample problems, while LLE applies convex optimization to obtain a weight matrix and LE employs a heat kernel function for this purpose. Orthogonal locality preserving projection (OLPP) and orthogonal neighborhood preserving embedding (ONPE) are the same with LPP and NPE, respectively, except for orthogonality restrictions [17,18]. Transformation matrices are restricted to be orthogonal [3]. These algorithms have been analyzed on hyperspectral images in a study by [19]. The aforementioned methods use an unsupervised approach. If the labeled data is scarce, semi-supervised techniques can also be used. One study presented a multitask regression framework to enhance model generalization and the quality of class indicators of unlabeled samples [20].

In addition, some supervised manifold learning approaches have also been proposed. They use class information of the input and aim to optimize similarities between and within class neighbors, while preserving the geometric structure of the data. They can be applied as linear or nonlinear. Approaches using class information which redefine the distances between the samples in LE and LLE have been shown in some studies [21,22]. Furthermore, linear operator mappings for the LPP have been proposed for better class separability [23–25]. Similarly, a specific method has been developed which was applied during the construction of the adjacency graph step in the NPE for the same purpose [26]. There are also Fisher-based linear manifold learning algorithms that improve class separability in the literature [27–30] and an analysis on supervised nonlinear embeddings has been suggested in a study by [31].

One study explained manifold learning-based dimension reduction methods and the technique to make them supervised [32]. The supervised technique applied to LE shown in [32] was based on maximizing inter-class and minimizing intra-class distances. The method employed two different weight matrices for each task to optimize cost. However, the proposed method employed one weight matrix for optimization. Contrary to the existing method, it created class-specific variables considering scatter information of every class. They employed a heat kernel function with a supervised manner to create the weight matrix. This approach was applied to LPP and OLPP instead of LE and was more suitable for classification applications due to its out-of-sample applicability. The linear manifold learning methods used were LPP, OLPP, NPE, and ONPE. The study also included comparative analytical evaluations of the algorithm giving the best classification accuracy performance to understand the effect of the proposed method. Analytical evaluations were made with the help of visual reports of class specific accuracy enhancements and the coefficients proposed, visualizations of weight and correlation matrices with the *p*values, and band information graphs. This study explores the relations between the analytical visuals and classification results. The application of the proposed method is expected to have better discrimination ability of extracted features as compared with a traditional version and to provide higher classification accuracy performance than the rivals mentioned. The experiments are conducted on HSI data.

The organization of this paper is as follows: In Section 2, the linear manifold learning algorithms which are used in the proposed method and the experiments are briefly described; in Section 3, the proposed method is described clearly; in Section 4, datasets, experiments, and the results are presented; finally, in Section 5, we conclude the paper with some remarks.

2. Linear Projections for Manifold Learning

Some important linear manifold learning methods which are applied to the proposed formulation on and/or used in the experiments are described.

2.1. LPP and NPE

The LPP and the NPE are linear manifold learning methods that are derived from the LE and the LLE, respectively. Given a set for training data is $X = [x_1, x_2, ..., x_m]$ in \mathbb{R}^n , they aim to find a transformation matrix, $A = [a_0, a_1, ..., a_{l-1},]$ that maps X to $Y = [y_1, y_2, ..., y_m]$ in \mathbb{R}^l to obtain lower dimensional space. The methods use the following equation for the linear projection:

$$x_i \to y_i = A^T x_i \tag{1}$$

In the LPP [13], the first step is to construct the adjacency graph. The nodes i and j are connected if i is in k-nearest neighbors of j. The following Euclidean distance formulation is used to determine the distances between the nodes:

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{in} - x_{jn})^2}$$
(2)

The next step of the LPP is to construct the weights. Weights on the connected nodes are calculated by a heat kernel function given by Equation (3). If the nodes *i* and *j* are not connected, then, it is represented by 0 in the *W*. *W* is a symmetric, $m \times m$ weight matrix as follows:

$$W_{ij} = e^{-\frac{||x_i - x_j||^2}{t}}$$
(3)

where $t = 2s^2$. The heat kernel parameter, *s*, is used in the experiments later.

The algorithm continues with the minimization of Equation (4) assuming $y^T = a^T X$, where α is the transformation vector as follows:

$$\sum_{i,j} (y_i - y_j)^2 W_{ij} \tag{4}$$

Finally, the minimization problem leads to a generalized eigenvalue problem as follows:

$$XLX^T a_i = \lambda_i XDX^T a_i \tag{5}$$

where *L* and *D* represent the Laplacian and diagonal matrices, respectively, and are defined as L = D - W and $D_{ii} = \sum_{j} W_{ji}$, respectively. The transformation matrix, *A*, is constructed

as $A = [a_0, a_1, \ldots, a_{l-1},]$, where $a_0, a_1, \ldots, a_{l-1}$ corresponds to l smallest eigenvalue such that $\lambda_0 < \lambda_1 < \ldots < \lambda_{l-1}$. A is substituted into Equation (1) to obtain the low-dimensional space that represents high-dimensional data.

NPE provides linearization to the LLE algorithm as stated formerly [16]. The first step of the algorithm is to construct the adjacency graph, as in the LPP. This provides neighborhood information. Unlike the LPP algorithm, the weight matrix, W, is obtained by using the convex optimization instead of the heat kernel function. Assuming $y^T = a^T X$, the minimization of Equation (6) is solved to obtain W as follows:

$$\sum_{i} \left(y_i - \sum_{j} W_{ij} y_j \right)^2 \tag{6}$$

Finally, the minimization problem leads to generalized eigenvalue problem as follows:

$$XMX^{T}a_{i} = \lambda_{i}XX^{T}a_{i} \tag{7}$$

where *M* equals $(I - W)^T (I - W)$ and *I* is identity matrix. As in the LPP, the transformation matrix, *A*, is constructed as $A = [a_0, a_1, \dots, a_{l-1},]$, where a_0, a_1, \dots, a_{l-1} corresponds to the *l* smallest eigenvalues such that $\lambda_0 < \lambda_1 < \dots < \lambda_{l-1}$. Linear projection is implemented by substituting *A* into Equation (1) to get low-dimensional representation.

2.2. OLPP and ONPE

The LPP and the NPE are redesigned with the orthogonality restriction of the transformation matrices. The minimization equations defined for the LPP and the NPE, i.e., Equations (4) and (6), are solved by taking $AA^T = I$ constraint into consideration to get the OLPP and the ONPE respectively.

The construction of the adjacency graph and finding of weight matrix phases, W, are the same as defined in the LPP and the NPE. The restriction, $AA^T = I$, leads to an easier eigenvalue solution as compared with that of the LPP and NPE. The *l* smallest eigenvalues are calculated from XLX^T and XMX^T in the OLPP and the ONPE respectively. Equation (1) is used for the projection. The reader can refer to [17,18] for more details about these algorithms.

3. The Proposed Method

The LE attempts to preserve the local structure of data. First, it constructs an adjacency graph, and then finds the weight matrix. The heat kernel function defined in Equation (3) can be applied to the graph at this phase. These steps are common for both the LPP and the OLPP.

A supervised version of the LE is overviewed in a study by [32]. It finds within-class and between-class neighbors upon a specific similarity threshold value. The algorithm continues by constructing two different weight matrices for within and between classes, W_w and W_b . At these steps, the heat kernel function is constructed with a preselected constant heat kernel parameter, *t*. The next step is the optimization of two objective functions, i.e., maximizing $\sum_{i,j} (y_i - y_j)^2 W_{b,ij}$ and minimizing $\sum_{i,j} (y_i - y_j)^2 W_{w,ij}$. Finally, these are combined to give a generalized eigenvalue solution.

In the proposed method, the weight matrix, *W*, is constructed with the help of a changing parameter in the denominator of Equation (3). The denominator is formulated based on class-specific scattering information of training data with a supervised manner. The proposed method optimizes an objective function depending on this *W*. It means that the proposed method requires only one weight matrix instead of two for optimization. The method is applied on the LPP and the OLPP for classification purpose.

Let μ_c be the coefficient that keep the specific values calculated for every class with c in 1, 2, ..., n and n is the total class number, and D is the $m \times m$ distance matrix where m is the number of training samples. Then, μ_c is found by applying Equations (8) and (9), sequentially, based on the information of D as follows:

$$S_{i} = \sqrt{\frac{1}{k_{i} - 1} \sum_{j=1}^{j=k_{i}} \left(D(i, I_{ij}) - avg_{i} \right)^{2}}$$
(8)

$$u_{c} = \frac{1}{t_{c}} \sum_{j=1}^{j=t_{c}} \left(s_{I_{grouped,c}} \right)_{j}$$
(9)

Equation (8) calculates the standard deviation of every *i*-th sample of *D* considering only intra-class distance information. Here, the *i*-th sample means the *i*-th row and *i* changes from 1 to *m*. The output of the equation is shown by S_i which corresponds to the *i*-th row. In the equation, k_i keeps the total number of intra-class samples for every row. I_{ij} keeps the intra-class sample indices for every *i*-th row; the maximum number of *j* depends

on the *i*-th sample intra-class number, k_i . The average of distances of intra-class samples of *i*-th rows is shown with avg_i .

After S_i is calculated for every row of D, the values that belong to the same classes on S are grouped by holding the indices of S elements. This information is kept on $s_{I_{grouped,c}}$ and allowed us to calculate the average of each grouped class as given by Equation (9). The indices of S elements belonging to the same classes are represented by $I_{grouped,c}$, where c is for class id. The total sample number of every class is shown by t_c . The maximum number of j depends on the sample number belonging to the specific classes, t_c .

At the end, *n* different μ are obtained that are represented as μ_c . Due to the symmetrical structure of *D*, the explanations made for notations are also valid for the columnwise perspective.

Up to this point, the procedure to find μ_c was explained. This parameter is exclusive for each class of training samples and used in the formulation to find *W*. The main steps of the application of the proposed method to LPP and OLPP are described as follows:

- The adjacency graph is constructed.
- Equation (10) is applied to samples among k-nearest neighbors.

$$W_{ij} = \begin{cases} e^{-\frac{||x_i - x_j||^2}{t}} \frac{1}{u_c p}, & \text{if } c(x_i) = c(x_j) \\ e^{-\frac{||x_i - x_j||^2}{t}}, & \text{otherwise} \end{cases}$$
(10)

- New weights for the samples among k-nearest neighbors are calculated with Equation (10). The class exclusive parameter, μ_c, is used when two samples belong to the same class that c(x_i) = c(x_j) holds. The parameter p is a constant determined manually to scale W. If the classes of the samples are not the same, the heat kernel function is applied directly.
- The samples that are not connected, i.e., not among k-nearest neighbors, are marked with 0 on W or this must be guaranteed in Step 1.
- The intra-class samples that satisfies $W_{ij} = 0$ among the k-nearest neighbors are determined and their weights are updated to new values determined by the first part of Equation (10) and, correspondingly, the maximum values of the samples on the row of *W* among k-nearest neighbors and not in the same class are marked as 0, i.e., this value is replaced with 0.
- The generalized eigenvector solution and the projection are implemented as the LPP or the OLPP does.

The main steps of the above algorithm are implemented considering Algorithm 1 given below.

The last part of the algorithm is the eigenvalue solution and is the same with the LPP and the OLPP. Time and space complexity is $O(m^2)$, but extra memory is required to keep 11 different variables to initialize the algorithm.

Algorithm 1 Proposed method

Given $D_{m \times m}$, the distance matrix; $K_{m \times m}$, the neighborhood information matrix (1, neighbor and 2, not); $L_{m \times m}$, the non-intra-class information matrix (1, non-intra and 2, not); $W_{m \times m}$, $J_{m \times m}$, $N_{c \times m}$, $S_{1 \times m}$, $Y_{1 \times m}$, and $\mu_{c \times 1}$, u, an integer; and p, a float used in Equation (10), are zero matrices and variables that are needed during the implementation:

c: Total number of classes and used as class enumeration. *m*: Total number of training number. *u*, *p* are independent variables created due to need of algorithm. For i = 1 to m do For j = 1 to m do If (Y(1,i) == Y(1,j))J(i,j) = D(i,j)end end $S(1,i) \leftarrow$ calculate standard deviation for every row for $J_{m \times m}$ (exclude zeros) For i = 1 to c do For j = 1 to m do If (Y(1,j) == c)N(c,i) = S(1,j)end end $\mu(c,1) \leftarrow calculate average of every row of N_{c \times m}$ (exclude zeros) For i = 1 to m do For j = 1 to m do If (Y(1,i) == Y(1,j) && K(i,j) == 1) $W(i,j) = exp(-D(i,j)^2/2.s^2 . \mu(Y(1,j),1).p)$ *Elseif* (Y(1,i) != Y(1,j) & & K(i,j) == 1) $W(i,j) = exp(-D(i,j)^2/2.s^2)$ end end For i = 1 to m do For j = 1 to m do If (Y(1,i) == Y(1,j) && K(i,j) == 0) $u = find_max_non_intra_indice (W(i,:) from L(i,:) = 1)$ W(i,u) = 0L(i,u) = 0 $W(i,j) = exp(-D(i,j)^2/2.s^2 .\mu(Y(1,j),1).p)$ end end

4. Experiments

4.1. Datasets

The experiments are carried out on HSI datasets, but first, synthetic data are employed. The synthetic data used are "3d_clusters" and can be created via [33]. The dataset includes 500 samples and has six different classes. Random noise is added to the samples, but most experiments are performed on hyperspectral data. Two different hyperspectral datasets are used in the experiments, i.e., the Indian Pines and Kennedy Space Center (KSC) datasets.

The Indian Pines and KSC datasets are both acquired by NASA AVIRIS sensor. Indian Pines is a 16-class dataset. After removing 20 noisy bands, 200 bands remain for the analysis. It primarily consists of regular-shaped agricultural fields with 10,249 samples in the ground truth. The KSC dataset includes 13 classes of land cover. After removing the water absorption bands, 176 bands remain. It includes 5211 samples in the ground truth. The Indian Pines and KSC data samples are not equally distributed among classes. The reader can refer to [34] to access these datasets to obtain more details about them.

4.2. Experimental Settings

The experiments are conducted on hyperspectral images; but, first, the effect of the proposed method is shown on synthetic data. LPP and Proposed + LPP embeddings are used, and the parameters are chosen heuristically to reduce three dimensions to two dimensions. Then, the classification accuracy performances that belong to LPP, OLPP, NPE, ONPE, the proposed method + LPP, and the proposed method + OLPP are searched on hyperspectral images. The reduced dimensionality (RD) is fixed to 20. The optimum neighborhood k and heat kernel parameter s are searched based on the classification performance of each algorithm that achieves its maximum over a range. The boundaries of the search loops and interval values for every iteration are determined heuristically. All experiments are repeated five times with random samples and classification rates are averaged. The overall accuracy values are shown with tables in the next subsection. The algorithm flowchart for the classification experiments on a randomly selected train and test data is illustrated in Figure 1.



Figure 1. The flowchart used for classification. (Red rectangle contains the process on a selected training/test dataset for feature learning. This part is repeated five times with randomly selected training/test datasets and classification results are kept, and then averaged to find optimal *k* and *s*. Cross validation is applied to train embeddings to determine hyperparameters. Then, the test data is evaluated. The test data are completely unseen data and used just for prediction).

As shown in Figure 1, the first step is to split the data into training and test datasets. The training dataset is employed for dimensionality reduction learning. A transformation matrix is created and with the help of this matrix, training and test embeddings are constructed. The process shown in the red rectangle continues over a range of s and k. The accuracy values are recorded during this period. Training and test dataset sample numbers are given with Table 1. The classes belonging to the HSI datasets do not have the same number of samples.

Table 1. Train/test sample numbers of the HSI datasets used.

Dataset	Indian P.	KSC
Total	10,249	5211
Training	1031	528
Test	9218	4683

Support vector machine (SVM) is chosen as the classifier in this study. In the SVM, a radial basis function (RBF) kernel is used with the penalty parameter, $C \in [2^{-4}, 2^8]$ and kernel width parameter, $\gamma \in [2^{-4}, 2^4]$. A grid search is applied on the closed intervals indicated. Cross validation is applied to search parameters.

Another main part of the study is to conduct experiments for analytical evaluations. The manifold learning algorithm that yields the highest classification accuracy performance is taken into consideration for the analysis to show the effects of the proposed method, which is the experimental strategy. The best resulting algorithm may become LPP or OLPP with the proposed versions, depending on the dataset.

The class-based accuracy enhancements and coefficients, μ_c , that are proposed with this study are visually reported. The other analytical elements can be expressed as the evaluation of band information, weight, and correlation matrices with their significance. The analytical experiments are carried out with randomly chosen data by considering the sample rates shown in Table 1. The parameters remained the same as previously determined in classification phase.

Classes 3 and 12 for Indian Pines and Classes 2, 7, 8, and 9 for KSC are declared as difficult to discriminate [35]. In the band information experiments, these classes are evaluated based on the first two bands. Furthermore, Bands 11 and 12 are chosen empirically to search whether they provide discrimination for any class.

4.3. Experimental Results

The visualization of the 2D embeddings for the synthetic dataset are given in Figure 2. According to this figure, the proposed method provides better class separation as compared with LPP. Furthermore, Table 2 shows the classification results for hyperspectral images. The most successful linear manifold learning method is shown in bold type.

The proposed method is implemented for both LPP and OLPP. According to Table 2, KSC has higher general classification accuracy rates than Indian Pines when all manifold learning algorithms used are considered. Proposed + LPP and Proposed + OLPP are the best linear manifold learning algorithms for Indian Pines and KSC, respectively. Furthermore, the most successful linear manifold learning algorithm among the existing ones are OLPP for Indian Pines and NPE for KSC datasets. The enhancement rates are 1.6% and 1.4% for Indian Pines and KSC datasets, respectively.



Figure 2. Visualizations of subspaces obtained upon "3d_clusters" toy dataset: (**a**) Original data; (**b**) LPP; (**c**) Proposed + LPP, (*x y* and *z* represent the features; LPP: Locality Preserving Projection).

Dataset	Linear Manifold Learning Method	OA (%) \pm std	Parameters
Indian Pines	LPP	69.2 ± 1.25	k = 50, s = 1600
	NPE	73.2 ± 0.25	k = 70
	OLPP	76.6 ± 0.72	k = 90, s = 800
	ONPE	74.7 ± 0.67	k = 40
	Proposed + LPP	78.2 ± 0.77	$k = 110, s = 400, p = 10^{-4}$
	Proposed + OLPP	75.2 ± 1.25	$k = 90, s = 1000, p = 10^{-4}$
KSC	LPP	82.9 ± 2.05	k = 2, s = 19,500
	NPE	90.0 ± 0.82	<i>k</i> = 32
	OLPP	87.4 ± 2.10	k = 18, s = 36,000
	ONPE	87.0 ± 1.00	k = 60
	Proposed + LPP	85.0 ± 0.87	$k = 50, s = 19,500, p = 10^{-6}$
	Proposed + OLPP	91.4 ± 0.83	$k = 20, s = 20,000, p = 10^{-6}$

Table 2. Classification performance on linear manifold learning methods for the HSI datasets. (OA, overall classification accuracy in percentage; std, standard deviation; *k*, neighborhood; *s*, heat kernel parameter; the algorithm with the best performance is bolded for every dataset).

Class-based μ_c values and enhancement rates for every dataset are reported and shown in Figure 3. According to this figure, the number of classes that show decrement in accuracy is only two out of 29 classes. Furthermore, in Figure 3a, a correlation pattern is noted. Another remarkable point is the classes that seem to match with zero μ_c , as shown in Figure 3b. In fact, these are not zero, but very low values as compared with other classes. This occurs because the classes having very similar band information that causes low scatter values.



Figure 3. Enhancement vs. μ_c graphs upon best resulting algorithms reported in Table 2: (a) Indian P.; (b) KSC (enhancement, class-based classification accuracy difference in percent between the proposed method + existing manifold learning algorithm and existing manifold learning algorithm; coeff, μ_c , class-based coefficients proposed with this study). (Note that every circle represents a class. The figure reports class-based accuracy enhancements and μ_c coefficients calculated for every class. The coefficient μ_c is related to the scattering level of intra-class samples and specific for every class. Class-based success of the proposed method can be evaluated for every dataset and compared with the coefficients proposed).

The other part of the experimental results consists of some visualizations for analytical evaluations. First, the weight matrices are extracted. The obtained weight matrices are shown with Figure 4 for hyperspectral datasets. These matrices show the effect of the changes that the proposed method causes. According to these figures, the values of coefficients around diagonals, i.e., the similarities for intra-class values, are increased. Additionally, off-diagonal values for non-intra-class values are decreased.

Figure 5a–d shows the graphs that belong to the first two bands of low-dimensional data of hyperspectral images. According to Figure 5a–d which is obtained upon "difficult classes", the class discrimination information is increased predominantly by applying the proposed method for the Indian Pines and KSC datasets. Furthermore, the first two bands for KSC show a significant discriminative property for the classes marked with green and red circles. It helps to enhance classification accuracies for KSC. Additionally, the feature information of some classes belonging to the Indian Pines and KSC datasets is improved on randomly chosen Bands 11 and 12. This is shown in Figure 5e–h.

The last evaluation metric is getting the visualizations of correlation matrices and *p*-values that give the relation between the samples for every dataset and their significance. If an off-diagonal value is smaller than 0.05, then, the correlation is considered to be significant. A smaller *p*-value is identified as a more significant correlation. The correlation matrices and *p*-values are given in Figures 6a–d and 6e–h, respectively. In Figure 6a–d, it can be seen that the correlation values of off-diagonal values are decreased, while there is no significant change around diagonals. Furthermore, *p*-values are increased for off-diagonal values, as observed in Figure 6e–h.



Figure 4. Weight matrices obtained with HSI datasets: (a) LPP for Indian Pines; (b) Proposed + LPP for Indian Pines; (c) OLPP for KSC; (d) Proposed + OLPP for KSC. (LPP: Locality Preserving Projection, OLPP: Orthogonal Locality Preserving Projection).



Figure 5. Information of Bands 1 and 2 for "difficult classes": (a) LPP for Indian Pines; (b) Proposed + LPP for Indian Pines; (c) OLPP for KSC; (d) Proposed + OLPP for KSC. Information of Bands 11 and12 for empirically chosen classes: (e) LPP for Indian Pines; (f) Proposed + LPP for Indian Pines; (g) OLPP for KSC; (h) Proposed + OLPP for KSC. (LPP: Locality Preserving Projection, OLPP: Orthogonal Locality Preserving Projection).



Figure 6. Correlation matrices obtained upon face and HSI datasets: (a) LPP for Indian Pines; (b) Proposed + LPP for Indian Pines; (c) OLPP for KSC; (d) Proposed + OLPP for KSC. Significance matrices (*p*-values) for the correlations: (e) LPP for Indian Pines; (f) Proposed + LPP for Indian Pines; (g) OLPP for KSC; (h) Proposed + OLPP for KSC. (LPP: Locality Preserving Projection, OLPP: Orthogonal Locality Preserving Projection).

5. Discussion

First, the proposed method is tested on a toy dataset. The proposed method gives better class separation for six randomly created noisy classes. Furthermore, the effectiveness of the proposed method is shown on two different hyperspectral datasets by applying classification. When the proposed method is evaluated among all linear manifold learning algorithms used, 1.6% improvement is achieved for the Indian Pines dataset and 1.4% for the KSC dataset. Improvements are observed in many classes of both datasets.

Weight matrices are extracted to observe the changes made by the proposed method on the algorithm for the HSI data. The value increment around diagonals and the value decrement on off-diagonals for weight matrices can be an indicator to get better classification results. However, obtaining a weight matrix is a mid-step and occurs before the construction of the embeddings. Therefore, it does not guarantee that the classification success rates always increase.

The band information and correlation matrices are extracted using the embeddings for both datasets. In the band analysis, it is shown that the information for Bands 1 and 2 is improved for the "difficult classes". Indeed, some bands can be discriminative only for some classes. In this context, the features belonging to some classes that have no information on Bands 11 and 12 are greatly improved with the proposed method. In addition, the correlation matrices for the bands can give information about distinguishability of samples. A remarkable point is that a general decrease in the correlation values for non-intra-class samples is observed. This can, likewise, have a positive effect on classification success, because it shows that the effect of irrelevant class samples is decreased. Furthermore, some correlations of off-diagonal, i.e., correlations belonging to non-intra-class samples lose their significance (greater *p*). This can be evaluated as a meaningless relation for non-intra-class samples and supports the supervised approach.

6. Conclusions

This study proposes an alternative algorithm for linear manifold learning based on dimensionality reduction improving feature learning by a supervised approach. It demonstrates that implementation of the proposed method with the LPP or OLPP algorithms outperforms not only LPP and OLPP but also NPE and ONPE algorithms for both hyperspectral datasets. Furthermore, the discriminative effect of the proposed method is shown with weight matrices. In addition, according to the statistical evaluations, the classification results are not a chance. Improvement of the information contributing to classification is shown on randomly chosen bands. Additionally, general correlations of non-intra-class samples and the significance of them are decreased. This supports the quality of classification. All the statistical findings are consistent with the classification improvements of both datasets.

The proposed method uses only Euclidean distance. In the future, state-of-the-art classifiers could be investigated to understand whether the classification success could be enhanced further or not. Experiments could be expanded to search optimum reduced dimensionalities to increase efficiency.

Author Contributions: Investigation, methodology, software, resources, data acquisition, writing original draft preparation, visualization, Ü.Ö.; conceptualization, validation, formal analysis, review and editing, supervision, A.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (accessed on 1 August 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Pal, M.; Foody, G.M. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Remote Sens.* 2010, 48, 2297–2307. [CrossRef]
- Zhuang, L.; Gao, L.; Zhang, B.; Fu, X.; Bioucas-Dias, J.M. Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations. *IEEE Trans. Geosci. Remote Sens.* 2020, 1–17. [CrossRef]
- Qiao, H.; Zhang, P.; Wang, D.; Zhang, B. An explicit nonlinear mapping for manifold learning. *IEEE Trans. Cybern.* 2012, 43, 51–63. [CrossRef] [PubMed]
- 4. Pless, R.; Souvenir, R. A survey of manifold learning for images. IPSJ Trans. Comput. Vis. Appl. 2009, 1, 83–94. [CrossRef]
- Lunga, D.; Prasad, S.; Crawford, M.M.; Ersoy, O. Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning. *IEEE Signal Process. Mag.* 2013, 31, 55–66. [CrossRef]
- Zhang, J.; Huang, H.; Wang, J. Manifold learning for visualizing and analyzing high-dimensional data. *IEEE Ann. Hist. Comput.* 2010, 25, 54–61. [CrossRef]
- 7. Van Der Maaten, L.; Postma, E.; Van den Herik, J. Dimensionality reduction: A comparative. J Mach Learn Res 2009, 10, 66–71.
- Tenenbaum, J.B.; De Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000, 290, 2319–2323. [CrossRef]
- 9. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* 2003, 15, 1373–1396. [CrossRef]
- 10. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. Science 2000, 290, 2323–2326. [CrossRef]
- 11. Hong, D.; Yokoya, N.; Zhu, X.X. Learning a robust local manifold representation for hyperspectral dimensionality reduction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2960–2975. [CrossRef]
- 12. Bengio, Y.; Paiement, J.F.; Vincent, P.; Delalleau, O.; Roux, N.; Ouimet, M. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–13 December 2003; pp. 177–184. [CrossRef]
- Chen, G.H.; Wachinger, C.; Golland, P. Sparse projections of medical images onto manifolds. In Proceedings of the International Conference on Information Processing in Medical Imaging, Asilomar, CA, USA, 28 June–3 July 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 292–303. [CrossRef]
- 14. Peherstorfer, B.; Pflüger, D.; Bungartz, H.J. A sparse-grid-based out-of-sample extension for dimensionality reduction and clustering with laplacian eigenmaps. In Proceedings of the Australasian Joint Conference on Artificial Intelligence, Perth, WA, Australia, 5–8 December 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 112–121. [CrossRef]
- 15. He, X.; Niyogi, P. Locality preserving projections. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 13–18 December 2004; Volume 16, pp. 153–160.
- 16. He, X.; Cai, D.; Yan, S.; Zhang, H.J. Neighborhood preserving embedding. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005; pp. 1208–1213. [CrossRef]
- 17. Cai, D.; He, X.; Han, J.; Zhang, H.J. Orthogonal laplacianfaces for face recognition. *IEEE Trans. Image Process.* **2006**, *15*, 3608–3614. [CrossRef] [PubMed]
- 18. Liu, X.; Yin, J.; Feng, Z.; Dong, J.; Wang, L. Orthogonal neighborhood preserving embedding for face recognition. In Proceedings of the 2007 IEEE International Conference on Image Processing, San Antonio, TX, USA, 16–19 September 2007; p. I-133. [CrossRef]
- Yılmaz, A.; Öztürk, Ü. An Analysis of Dimension Reduction Methods Applicable for Out of Sample Problem in Hyperspectral Images. In Proceedings of the 2019 9th International Conference on Recent Advances in Space Technologies (RAST), İstanbul, Turkey, 11–14 June 2019; pp. 381–386. [CrossRef]
- Hong, D.; Yokoya, N.; Chanussot, J.; Xu, J.; Zhu, X.X. Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction. *ISPRS J. Photogramm. Remote Sens.* 2019, 158, 35–49. [CrossRef]
- 21. Raducanu, B.; Dornaika, F. A supervised non-linear dimensionality reduction approach for manifold learning. *Pattern Recognit*. **2012**, *45*, 2432–2444. [CrossRef]
- 22. Chen, J.; Liu, Y. Locally linear embedding: A survey. Artif. Intell. Rev. 2011, 36, 29–48. [CrossRef]
- Yang, W.; Sun, C.; Zhang, L. A multi-manifold discriminant analysis method for image feature extraction. *Pattern Recognit.* 2011, 44, 1649–1657. [CrossRef]
- 24. Zhang, Z.; Zhao, M.; Chow, T.W. Marginal semi-supervised sub-manifold projections with informative constraints for dimensionality reduction and recognition. *Neural Netw.* 2012, *36*, 97–111. [CrossRef]
- Hua, Q.; Bai, L.; Wang, X.; Liu, Y. Local similarity and diversity preserving discriminant projection for face and handwriting digits recognition. *Neurocomputing* 2012, *86*, 150–157. [CrossRef]
- Wen, J.; Yan, W.; Lin, W. Supervised linear manifold learning feature extraction for hyperspectral image classification. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 3710–3713. [CrossRef]
- 27. Sugiyama, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.* **2007**, *8*, 1027–1061.
- Li, B.; Liu, J.; Zhao, Z.Q.; Zhang, W.S. Locally linear representation fisher criterion. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013; pp. 1–7. [CrossRef]

- 29. Cui, Y.; Fan, L. A novel supervised dimensionality reduction algorithm: Graph-based Fisher analysis. *Pattern Recognit.* **2012**, *45*, 1471–1481. [CrossRef]
- 30. Wang, R.; Chen, X. Manifold discriminant analysis. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 429–436. [CrossRef]
- 31. Vural, E.; Guillemot, C. A Study of the Classification of Low-Dimensional Data with Supervised Manifold Learning. J. Mach. Learn. Res. 2017, 18, 5741–5795.
- 32. Chao, G.; Luo, Y.; Ding, W. Recent advances in supervised dimension reduction: A survey. *Mach. Learn. Knowl. Extr.* 2019, 1, 341–358. [CrossRef]
- Matlab Toolbox for Dimensionality Reduction. Available online: https://lvdmaaten.github.io/drtoolbox/ (accessed on 20 February 2021).
- 34. Hyperspectral Remote Sensing Scenes. Available online: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_ Remote_Sensing_Scenes (accessed on 20 February 2021).
- 35. Crawford, M.M.; Ma, L.; Kim, W. Exploring nonlinear manifold learning for classification of hyperspectral data. In *Optical Remote Sensing*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 207–234.