

Peer Reviewed Paper openaccess

Adaptive hierarchical clustering for hyperspectral image classification: Umbrella Clustering

S.S.P. Vithana,^a E.M.M.B. Ekanayake,^{b,*} E.M.H.E.B. Ekanayake,^c A.R.M.A.N. Rathnayake,^c G.C. Jayatilaka,^d H.M.V.R. Herath,^e G.M.R.I. Godaliyadda^c and M.P.B. Ekanayake^c

^aSri Lanka Technological Campus, Padukka, Sri Lanka

^bDepartment of Electrical and Electronic Engineering, University of Peradeniya, Peradeniya, Sri Lanka. E-mail: mevanekanayake@gmail.com, https://orcid.org/0000-0002-4768-5073

^cDepartment of Electrical and Electronic Engineering, University of Peradeniya, Peradeniya, Sri Lanka

^dDepartment of Computer Engineering, University of Peradeniya, Peradeniya, Sri Lanka. [®] https://orcid.org/0000-0001-6999-4777

Department of Electrical and Electronic Engineering, University of Peradeniya, Peradeniya, Sri Lanka. (1) https://orcid.org/0000-0002-2094-0716

Hyperspectral Imaging (HSI) utilises the reflectance information of a large number of contiguous spectral bands to solve various problems. However, the relative proximity of spectral signatures among classes can be exploited to generate an adaptive hierarchical structure for HSI classification. This enables a level by level optimisation for clustering at each stage of the hierarchy. The Umbrella Clustering algorithm, introduced in this work, utilises this premise to significantly improve performance compared to non-hierarchical algorithms which attempt to optimise clustering globally. The key feature of the proposed methodology is that, unlike existing hierarchical algorithms which rely on fixed or supervised structures, the proposed method exploits a mechanism in spectral clustering to generate a self-organised hierarchy. The algorithm gradually zooms into the feature space to identify levels of clustering at each stage of the hierarchy. The results further demonstrate that the generated structure tallies with human perception. In addition, an improvement to Linear Discriminant Analysis (LDA) is also introduced to further improve performance. This modification maximises the pairwise class separation in the feature space. The entire algorithm includes this modified LDA step which requires a certain amount of class information in terms of features, at the training phase. The classification algorithm which incorporates all novel concepts was tested on the HSI data set of Pavia University as well the database of Common Sri Lankan Spices and Adulterants in order to assess the versatility of the algorithm.

Keywords: hyperspectral imagery, spectral clustering, hierarchical classification, umbrella clustering, feature extraction, remote sensing, linear discriminant analysis, self-organise, unsupervised

Correspondence

 $E.M.M.B.\ Ekanayake\ (\underline{mevanekanayake@gmail.com})$

Received: 23 October 2018 Revised: 27 May 2019 Accepted: 24 June 2019 Publication: 9 July 2019 doi: 10.1255/jsi.2019.a11 ISSN: 2040-4565

Citation

S.S.P. Vithana, E.M.M.B. Ekanayake, E.M.H.E.B. Ekanayake, A.R.M.A.N. Rathnayake, G.C. Jayatilaka, H.M.V.R. Herath, G.M.R.I. Godaliyadda and M.P.B. Ekanayake, "Adaptive hierarchical clustering for hyperspectral image classification: Umbrella Clustering", J. Spectral Imaging 8, a11 (2019). https://doi.org/10.1255/jsi.2019.a11

© 2019 The Authors

This licence permits you to use, share, copy and redistribute the paper in any medium or any format provided that a full citation to the original paper in this journal is given.

Introduction

Hyperspectral images contain intensity values of a large number of contiguous spectral bands. This facilitates the detection of subtle details in a given image. Hence Hyperspectral Imaging (HSI) is widely used in many applications, such as remote sensing, food quality testing, medical imaging and surveillance.¹⁻⁴ In order to detect the finer details in a given image, a number of classification algorithms for HSI have been developed recently, which include pre-processing techniques, clustering algorithms, probabilistic frameworks etc.

As a large amount of data in hyperspectral images could be redundant as well as misleading in some cases, effective dimensionality reduction techniques and feature extraction mechanisms have also been developed.⁵ Principal Component Analysis (PCA)⁶ is a widely applied linear dimensionality reduction technique in HSI, which reduces the dimension of a data set by projecting data points onto the directions of maximum overall scatter in the original space. Folded-PCA⁷ achieves the same objectives of conventional PCA with lower computational cost and memory requirement, by representing the spectral vectors of the image as a matrix. Linear Discriminant Analysis (LDA)⁸ is a classification technique as well as a mechanism of feature extraction, which maximises the between-class scatter while minimising the within-class scatter of a data set. Local Fisher Discriminant Analysis⁹ proposes a more effective mechanism to determine the within-class scatter matrix and the between-class scatter matrix in LDA, by considering local neighbourhoods in the data space.

Non-linear transformations such as Locally Linear Embedding (LLE)¹⁰⁻¹³ and sparse discriminant manifold embedding¹⁴ are also used to reduce the dimension of hyperspectral image data sets. These techniques first identify low-dimensional manifolds within the high-dimensional data space that can effectively represent the data and then provide the transformation procedure to collapse the high-dimensional data space to the low-dimensional space, while preserving the neighbourhood of each data point.

Feature Space Discriminant Analysis (FSDA)¹⁵ is a dimensionality reduction technique as well as a feature extraction mechanism which consists of a feature extraction process prior to maximising the variance of the data points. There are other feature extraction mechanisms, such as band selection based on Visual Assessment of cluster Tendency (VAT)¹⁶ and Nonparametric Feature

Extraction (NFE).¹⁷ From analysing the above dimensionality reduction and feature extraction techniques, it can be concluded that there is a trade-off between the performance and the computational requirements associated with their implementation.

Pre-processing mechanisms which do not transform the data points into a new space have also been developed recently. Relevance-Based Feature Extraction¹⁸ is one such mechanism of feature extraction, which weights each feature of the data points, such that the maximum separation among classes is achieved in the original space.

Apart from the algorithms developed to perform preprocessing, a number of algorithms have been developed to cluster and classify HSI data. Rank-two Nonnegative Matrix Factorisation (rank-two NMF)¹⁹ is an iterative classification algorithm based on a fixed hierarchical structure. Hierarchical SVM²⁰ is a two-stage classification algorithm based on Support Vector Machine (SVM)²¹ for HSI data. An agglomerative structure of a hierarchical classification algorithm is presented, 22 where a two-stage process is employed. In the first stage, data points are classified together, subjected to spatial constraints as opposed to the second stage where the individually classified components are combined with no spatial constraints considered. Simultaneous Orthogonal Matching Pursuit (SOMP)²³ and Simultaneous Subspace Pursuit (SSP)²³ are two classification algorithms developed, based on the sparse representation of HSI data.

Partitioned clustering techniques have been used to develop a classification algorithm for HSI²⁴ incorporating both spatial and spectral information of the image. In this algorithm, each pixel is classified using SVM on the spectral data, while a probabilistic approach is undertaken to incorporate the spatial information of the image, by considering the class labels of the eight neighbouring pixels of each pixel in the image. Modern optimisation techniques such as the Artificial Bee Colony algorithm have also been incorporated in HSI classification recently.²⁵ A Game Theoretical classification Algorithm (GTA),²⁶ which uses a Conditional Random Field (CRF) to model the spatial features of the image is another HSI classification algorithm which incorporates both spectral and spatial information of the image. Recently, Deep Learning algorithms have been used excessively in HSI classification²⁷⁻³⁰ due to the high levels of accuracy they can achieve.

It can be observed from existing literature that most hierarchical classification algorithms developed so far are supervised, where the number of stages and end members are known prior to implementation, or unsupervised with a fixed structure. The main contribution of this paper is the proposed Umbrella Clustering algorithm which exploits the relative proximity of inter-class scatter to self-organise the algorithm itself in a hierarchical manner. This enables a level by level optimisation of clustering procedures. This is more effective than a global optimisation of clustering algorithms, especially in situations where the inter-class scatter is non-uniformly distributed. This is done by adjusting the scaling parameter of the Gaussian kernel in the standard spectral clustering algorithm³¹ to zoom into the data space at various levels. By zooming into the data set at different levels, the algorithm can identify the major clusters (super-clusters) of data points available at each stage. The scalar sweep or the zooming process actually provides what is called the "modes of clustering" or the most prominent grouping of clusters at each stage and its strength (quantified by the index and strength of the dominant Eigen-gap of the affinity matrix in spectral clustering). This enables the proper identification of the most prominent mode or grouping for a given data set.

The proposed method was tested on the Pavia University data set³² as well as the Database of Common Sri Lankan Spices and Adulterants³³ (Mendeley Data) to gain accurate results for classification. One key advantage of this method is its adaptability to any given dataset. The results of the case studies in this paper show that there is a significant improvement in the overall accuracy (around 30%) from using the Umbrella Clustering algorithm, compared to that of the single-staged global optimisation process carried out on the Pavia University data set with the aid of spectral clustering.

Umbrella Clustering results in similar accuracy levels compared to a number of state-of-the-art classification algorithms. However, the Umbrella Clustering algorithm has the advantage of producing similar levels of accuracy as the other algorithms, with less training data. Furthermore, this algorithm has only one user input variable as opposed to most of the state-of-the-art algorithms which require the user to initialise a number of parameters. The above two advantages are extensively discussed in this paper.

In addition to the concept of Umbrella Clustering, a classification mechanism that maximises the pairwise

class separation is also introduced in this paper. This is a modification of LDA which addresses the issue of overlapping of different classes when maximising the class mean to overall mean separation.³⁴ Apart from maximising the pairwise class separation, this technique minimises the within-class separation, similar to LDA. However, the inclusion of the modified LDA step requires prior information at the training phase, nevertheless it enhances the performance of the entire algorithms.

Datasets

A hyperspectral image dataset from Pavia University³² and Database of Common Sri Lankan Spices and Adulterants³³ (Mendeley Data) were used to test the hyperspectral image classification algorithm developed in this paper.

The HSI data set from Pavia University, acquired by the Reflective Optics System Imaging Spectrometer (ROSIS), consists of 103 spectral bands with 610 × 340 pixels. The data set consists of pixels belonging to nine classes, namely asphalt, meadows, gravel, trees, painted metal sheets, bare soil, bitumen, self-blocking bricks and shadows. Figure 1(a) shows the RGB image of the area while Figure 1(b) shows the ground truth provided. This data set has been used to test a number of previously developed algorithms for HSI. 35,36 The dataset of common Sri Lankan spices captured by a multispectral imaging system consists of nine spectral bands with 315 × 393 pixels. The data set consists of pixels belonging to 16 classes: common Sri Lankan spices (turmeric, pepper, chili and curry powder), adulterants (tartrazine and rice flour) and adulterated versions of the aforementioned spices. Figure 1(c) shows the RGB image of the dataset while Figure 1(d) shows the ground truth. This data set also has been used to test a number of previously developed algorithms for HSI.³⁷

Proposed method

This paper proposes an algorithm to classify HSI data, which consists of two main stages. The first stage is the application of modified LDA in order to increase the interclass separation to intra-class separation ratio of the dataset. This is a supervised procedure which requires class information. However, the second stage, which is

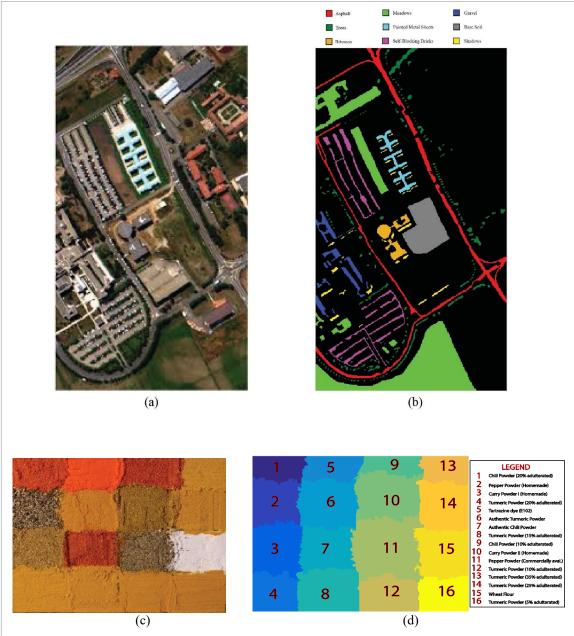


Figure 1. (a) RGB image of the Pavia University Scene. (b) Ground truth of the Pavia University data set. (c) RGB image of the dataset of common Sri Lankan spices. (d) Ground truth of the dataset of common Sri Lankan spices.

the main focus of the paper, organises the hierarchical structure of data by adjusting the scaling parameter of the Gaussian kernel in the standard spectral clustering algorithm in an unsupervised manner.

This novel algorithm zooms into the high-dimensional data space at different levels in a hierarchical manner and identifies clusters of points. These are called super-clusters, as they contain data points belonging to a number of classes. The algorithm zooms into each super-cluster

separately, and identifies the sub-clusters, generating a hierarchical structure for classification. This is discussed further in this paper.

Modified LDA

The mean spectral signatures and the variances of the nine classes in the Pavia University dataset are shown in Figure 2(a) and (b). It can be noticed that the mean spectral signatures of pairs, gravel and bricks, meadows and bare soil, and asphalt and bitumen are quite similar. Hence, the separation of data points among these pairs is not a trivial task in the original 103-dimensional space. As a remedy, the data set can be transformed into a new space, where the between-class similarities are minimised.

Figure 3(a) shows that the spectral signatures of the pixels belonging to the class "trees" are embedded in the band of spectral signatures of the pixels belonging to the class "meadows". This means that the cluster of pixels containing pixels belonging to "trees" is engulfed in the larger and more spacious cluster of pixels belonging to the class "meadows". Hence, the next step of the algorithm which zooms into the data space in a series of levels, would not be able to separate pixels among these two classes, no matter how deep it zooms in.

Based on this observation, a modification was proposed to LDA. As LDA maximises the between-class scatter by increasing the variance between each of the class means and the overall mean, there is a possibility of two classes overlapping in the transformed space while the separation between the class means and the overall mean is at its maximum level. Basically, the LDA metric can be maximised in a global sense while pairwise separations are reduced. This is because a scatter increase from the overall mean does not necessarily imply optimal pairwise separation.

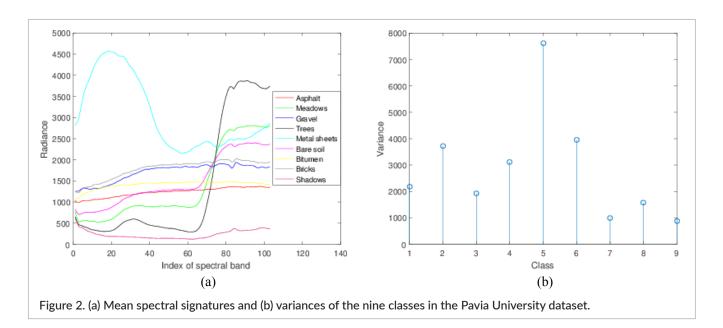
This is illustrated in Figure 4(a) and (b). Figure 4(a) shows the clusters of data points in the original space and the directions of which the separation will be maximised through LDA. Figure 4(b) shows a possible outcome of LDA, leading to an overlap of two clusters in the transformed space.

In order to address the aforementioned shortcoming of LDA, this paper suggests maximising the pairwise class separation instead of the class-to-overall mean separation in LDA, which is illustrated in Figure 4(c). This ensures minimal overlapping among clusters of data points while maximising the pairwise class separation and minimising the within-class scatter of data points, similar to LDA. Within-class scatter of a given class was represented by the variance of each class (using centroid distances) and the between-class scatter was calculated by obtaining the sum of all outer products of the vectors representing the separation between each of the class mean pairs. The equations calculating the within and between-class scatter matrices are given by:

$$S_{\mathbf{b}} = \frac{1}{n_{C_2}} \sum_{j=1}^{n} \sum_{i=1}^{n} (\overline{\mathbf{x}}_i - \overline{\mathbf{x}}_j) (\overline{\mathbf{x}}_i - \overline{\mathbf{x}}_j)^{\mathsf{T}}; i \le j$$
 (1)

$$S_{\mathbf{w}} = \frac{1}{\sum_{j=1}^{n} L_j} \sum_{j=1}^{n} \sum_{j=1}^{L_j} (x_i - \overline{\mathbf{x}}_j)(x_i - \overline{\mathbf{x}}_j)^T$$
 (2)

where S_b is the between-class scatter matrix, S_w is the within-class scatter matrix, n is the number of classes, n_{C_2} is the number of pairwise combinations obtainable from n classes, L_j is the number of pixels in class j, \mathbf{x} is the feature vector of a pixel and $\overline{\mathbf{x}}$ is the class mean. By calculating the Eigen vectors corresponding to the



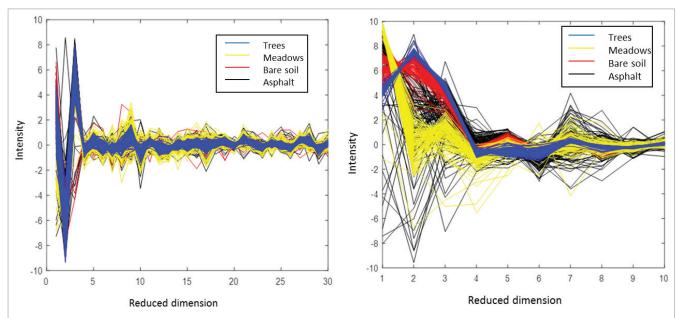


Figure 3. Spectral signatures of a training sample representing four classes of the data set when (a) LDA and (b) modified LDA are applied.

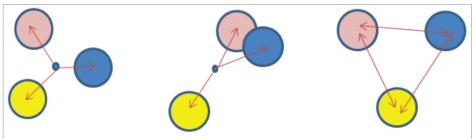


Figure 4. (a) Maximising vectors in classical LDA, (b) overlapping of different classes and (c) maximising vectors in the proposed modification to LDA.

dominant Eigen values of $S_w^{-1}S_b$, the transformation matrix can be obtained. The spectral signatures of the same training sample after applying modified LDA are shown in Figure 3(b). They are not overlapped, as opposed to the previous case in Figure 3(a). This means that the clusters of data points in the transformed space are separated more effectively from one another, which is a favourable condition for convenient classification.

The between- to within-class scatter ratio of the nine classes in the Pavia University data set, with and without the proposed modified LDA technique, is illustrated in Figure 5, which shows that the ratio has increased upon the application of modified LDA. Although the ratio has decreased in classes 7 and 9, the effect of the reduction is insignificant due to the high absolute value of the ratio. The overall increase in the between- to within-

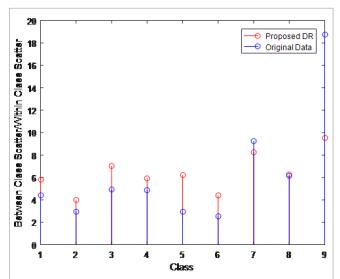


Figure 5. Between- to within-class scatter ratio with and without the proposed modified LDA technique.

class scatter implies a higher degree of separation among classes in the new space, compared to that of the original space.

In order to compare the within- and between-class scatters of each class in Figure 5, the class variance and the mean squared distance (Euclidean distance) between each point in the class with all other points in the data set which do not belong to the same class, were used respectively.

Adaptive hierarchical clustering: Umbrella Clustering

An unsupervised clustering mechanism is required to generate a self-organising hierarchical structure for classification. As spectral clustering can be adaptively used to zoom in and out of the space, using the trainable parameter in the Gaussian kernel employed in its implementation, spectral clustering becomes a good candidate for the generation of the hierarchical structure of classification. Hence, this algorithm is based on spectral clustering, which identifies the structure of the data set and clusters them according to their degree of affinity. The steps and respective equations of spectral clustering are as follows.

1) Generating the distance matrix

$$\mathbf{D}_{ij} = \left\| \mathbf{x}_i - \mathbf{x}_j \right\| \quad \forall i, j \in \{1, N\}$$
 (3)

where D_{ij} = measure of distance between pixels i and j (Euclidean distance), \mathbf{x} = feature vector of a pixel and N = number of pixels.

2) Generating the affinity matrix using a Gaussian kernel

$$\mathbf{A}_{ii} = e^{-\frac{\mathbf{D}_{ij}}{2\sigma^2}} \tag{4}$$

where Aij = Measure of affinity between pixels i and j σ = scale factor.

3) Generating the graph Laplacian

$$\mathbf{W} = \begin{cases} W_{kj} = \sum_{i=1}^{N} \mathbf{A}_{ij} & k = j \\ W_{kj} = O & k \neq j \end{cases}$$
 (5)

$$L = W^{-\frac{1}{2}}AW^{-\frac{1}{2}} \tag{6}$$

where L = graph Laplacian, W = normalising matrix.

4) Finding the Eigenvectors corresponding to the dominant Eigenvalues of the Laplacian matrix

$$LV = \lambda V \tag{7}$$

$$\mathbf{X}_{new} = [\mathbf{V}_1 \quad \mathbf{V}_2 \dots \mathbf{V}_t] \tag{8}$$

where V = Eigen vectors of L, λ = Eigen values of L, X_{new} = feature vectors in the new space, t = required number of dimensions.

The degree of affinity, of which spectral clustering is based on, is a subjective measurement which depends on the point of observation. The parameter σ can be identified as a parameter which controls the level of zooming in and out of the data set when performing spectral clustering.⁴

When σ decreases, the degree of affinity between a given pair of points decreases.⁴ This is similar to zooming into the data space where the distance between two points seems to appear much larger, lowering the degree of affinity. Similarly, an increase in σ can be explained as the process of zooming out of the data space. This concept is exploited in Umbrella Clustering, which is a hierarchical clustering algorithm that organises its hierarchical structure autonomously.

The number of clusters present in a data set, at a given value of σ can be identified by the structure of the affinity matrix. Let us call the number of prominent Eigenvalues of the affinity matrix at a given value of σ , the "mode" of clustering.

The prominent mode of clustering or the most prominent number of clusters present in a dataset at a given level of zooming in, can be effectively calculated by

$$G(i) = \lambda_i - \lambda_{i+1}$$
 $i = 1, 2, ..., N-1$ (9)

$$PM = \arg\max_{i} G \tag{10}$$

where λ s are the Eigenvalues of the normalised affinity matrix, \mathbf{G} is the vector containing the differences between the adjacent Eigenvalues when sorted in the descending order, N is the number of pixels and PM is the prominent mode of the given data set. As the differences between adjacent Eigenvalues called Eigen gaps [the value $\mathbf{G}(i)$] spike, this signals the possibility of clustering the data set by the corresponding number of classes i. This is called a mode in the cluster formation. The magnitude of the spike at i [i.e. $\mathbf{G}(i)$], symbolises the clarity or the precision of separation of i number of clusters for the given data set.

Multiple modes can exist for a given dataset. Each mode has a corresponding index i and strength G(i). Hence, a measure of precision (P) for the most prominent number of clusters calculated by Eqn (10) is given by

$$P = \max_{i} G \tag{11}$$

Hence the mode with the maximum strength as per Eqn (11) generates the most prominent mode for the given data set. This mode's index i yields the optimal number of clusters the data set can be arranged in, in the current iteration (zoom level). At this point the separation between i number of classes supersedes the separation between any j; $j \neq i$ classes for the current iteration (zoom level) of the data set.

The proposed algorithm first calculates the modes of clustering present in a data set for a range of σ values (different zoom levels into the data set) and identifies the mode with the highest precision (most prominent mode), along with the respective value of σ . Next, it performs spectral clustering on the data set using Eqns (3–8), with the σ value and the number of Eigenvectors to be extracted specified based on the index of the most prominent mode. The algorithm then labels the resulting clusters using the k-means algorithm. In this way, the Umbrella Clustering algorithm decides the number of branches (super-clusters) in the hierarchical structure autonomously, along with an optimum value for the tuning parameter σ .

In the next step, the algorithm considers only one of the clusters separated in the previous stage, for which it calculates the modes of clustering and selects the most prominent mode with the corresponding value of σ , and performs spectral clustering and labels the clusters. In this manner, the algorithm zooms into different parts of the data set at different levels and clusters the data points at each level of zooming. This is repeated until only two modes are available for sub-clustering a given cluster, namely, the mode one and mode N, where N is the total number of data points present in the current cluster. This symbolises the end node of the hierarchical structure, where the current cluster cannot be sub-divided into any more sub-clusters.

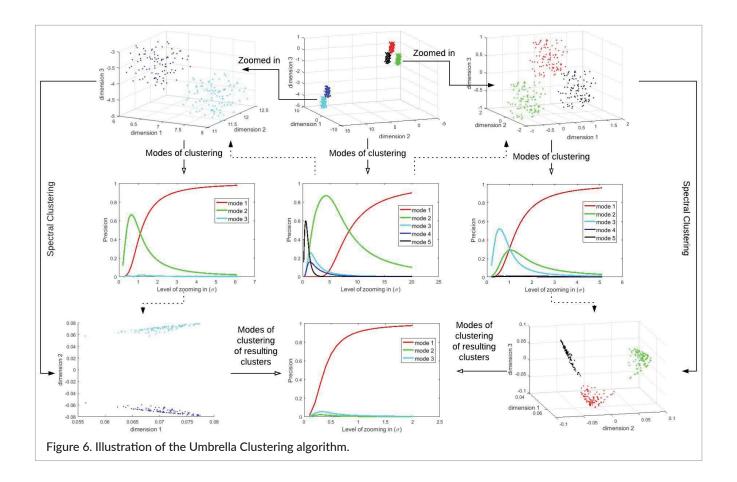
This is an unsupervised hierarchical clustering algorithm which organises its hierarchical structure adaptively and autonomously. The algorithm is illustrated using a three-dimensional dataset (selected for clarity of illustrative demonstration) shown in Figure 6. The three-dimensional dataset consists of five classes that are shown in

five different colours. In this scenario, two classes are located closer together, while the other three classes are lumped together separately.

The modes of clustering and their respective strengths or precisions for a wide range of σ values, considering the entire data set, are shown below the 3-D data set in Figure 6. The range of σ values is chosen according to the overall scatter of the dataset. This could be a trial and error approach for the first few stages of the algorithm and the range would remain the same for the rest of the stages. If the number of data points is N, there will exist a σ value which is large enough, where the number of clusters is N, and there will exist a σ value small enough, where the number of clusters is one. In any reliable clustering algorithm, if the number of classes equals the number of data points or if the number classes equal one, it does not convey any reliable information. The overall scatter of the data points determines these bounds. In other words, if the scatter of a particular dataset is small, it is not necessary to zoom out by a large amount in order to visualise it as a single cluster. In contrast, if the scatter of a particular dataset is large, it is essential to zoom out by a reasonable amount in order to visualise it as a single cluster.

Mode 1, with a maximum precision closer to one, is present in the data set for higher values of σ . The higher the value of σ , the higher the degree of affinity which explains the case where the data set is viewed from afar, where the distances between pixels (data points) seem very small, resulting in a single lumped cluster. When the value of σ is reduced, the level of zooming into the data set is increased. As expected, more clusters are present when zoomed in, which can be explained by the higher values of precision in modes 2 and 5. Mode 5 can be physically explained as there are five clusters in total while mode 2 is explained by the two lumped clusters available in the data space.

Out of the three modes of clustering identified, mode 1 can be eliminated as it will always be present at higher values of σ with a high degree of precision. Out of the other modes available, mode 2 has the highest precision. Therefore, spectral clustering is applied to the data set with σ =5, considering the Eigenvectors corresponding to the first two largest Eigenvalues of the Laplacian matrix of the data set. Then the clusters of pixels formed in the new space are identified using the k-means algorithm. The resulting super-clusters, when zoomed in, in the



original space, are shown on either side of the overall data set.

As the next stage of the hierarchical structure, the first super-cluster containing the red, green and black classes is selected and the procedure is repeated. The modes of clustering and corresponding values of precision are calculated based on their original three-dimensional data. Mode one is present as usual for high values of σ . Mode 2 is also present with a lower degree of precision, as the red and black classes are placed relatively closer together compared to the green class, resulting in two sub-clusters. However, mode three has the highest value of precision, which is due to the presence of the three different classes in the first umbrella cluster. Hence, spectral clustering is applied to the first umbrella cluster for the separation of three clusters with $\sigma = 0.6$. Each of the three clusters generated by the above stage is considered for the repetition of the same procedure as above and the resulting graph depicting the modes of clustering demonstrates that the clusters cannot be separated any further as the only mode available for a wide range of σ is one.

Next, the second super-cluster (light blue and dark blue) is analysed for its modes of clustering. Since mode 2 is the only mode of clustering present, except for mode 1 at high values of σ , spectral clustering is applied to the second umbrella cluster with σ = 0.7, in order to separate it into two sub-clusters.

The same procedure is applied to the two resulting clusters, to find out the modes of clustering, which indicates that mode 1 is the only available mode. Hence the umbrella clustering algorithm for this data set ends at this point. Figure 7 shows the resulting adaptive hierar-

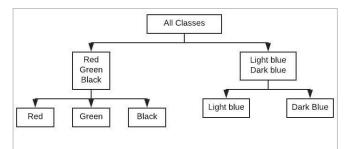


Figure 7. Adaptive hierarchical structure for classification, based on the example considered.

chical structure for classification, which classifies the five classes in the data set with 100% accuracy.

Case studies

Case study 1

Combining the proposed modified LDA and the unsupervised adaptive hierarchical clustering (Umbrella Clustering) algorithm, a classification process was carried out on the hyperspectral image data set of Pavia University. The training phase of case study 1 was carried out using a training sample of 900 pixels (100 from each class) in order to generate the hierarchical structure of clustering along with the corresponding σ values, and to find the transformations that perform the modified LDA at each stage of the hierarchy.

As depicted in Figure 8, mode 5 with the highest precision of 0.35 was chosen to perform spectral clustering on the training set, in the first stage of the hierarchy, with a corresponding σ value of 150. Note that mode 1 can be neglected since it is a universal mode. Once spectral clustering was applied and the data space was transformed into a new space with enhanced separation among the five clusters identified, the k-means algorithm was initialised with input parameter set to five for the number of clusters as per the prominent mode. Similarly, the process was repeated in all sub-clusters of the training sample until the most prominent mode was one for all values of σ . The overall flow of the hierarchical

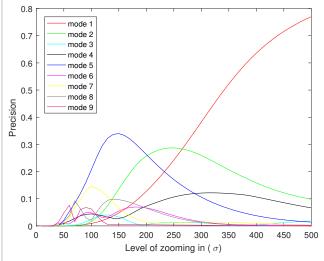


Figure 8. Modes of clustering of the Pavia University data set, at the first stage.

structure of the classification, upon the application of the developed algorithm on the training sample, is shown in Figure 9.

Case study 2

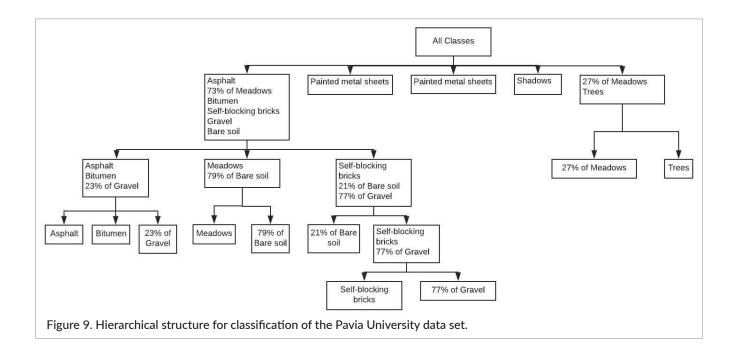
Using the umbrella novel methodology, a classification process was carried out on the Image Database of Common Sri Lankan Spices and Adulterants. The training phase of case study 2 was carried out using a training sample of 1600 pixels (100 from each class) in order to generate the hierarchical structure of clustering along with the corresponding σ values, and to find the transformations that perform the modified LDA at each stage of the hierarchy.

As depicted in Figure 10, mode 2 with the highest precision of 0.32 was chosen to perform spectral clustering on the training set, in the first stage of the hierarchy, with a corresponding σ value of 28. Note that mode 1 can be neglected since it is a universal mode. Once spectral clustering was applied and the data space was transformed into a new space with enhanced separation among the two clusters identified, the k-means algorithm was initialised with input parameter set to two for the number of clusters as per the prominent mode. Similarly, the process was repeated in all sub-clusters of the training sample until the most prominent mode was one for all values of σ . The overall flow of the hierarchical structure of the classification, upon the application of the developed algorithm on the training sample, is shown in Figure 11.

Results and discussion

The results of the case studies reveal the levels of accuracy, upon the application of the main algorithm in case study 1 followed by case study 2. While both the case studies were carried out in high-dimensional feature spaces of HSI data, case study 1 is based on a remote sensing application while case study 2 is aligned towards an agricultural or food engineering perspective.

The results of the hierarchical classification algorithm applied on the Pavia University data set, with the truth labels shown in columns and predicted labels in rows are presented in Table 1. It can be noticed that smaller fractions of meadows have been misclassified as trees and bare soil which can be justified by the two types of meadows (green and brown) present in the actual image



[Figure 1(a)]. On the other hand, portions of trees and bare soil have been misclassified as meadows, for the same reason. A few minor misclassifications in asphalt, gravel, bitumen and self-blocking bricks are also evident in the confusion matrix, due to their similarity in composition and colour.

The application of the Umbrella Clustering algorithm on the Pavia University dataset has resulted in five main super-clusters in the first stage, which are then clustered further into a number of sub-clusters, in a process of four

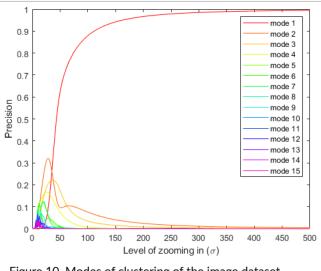


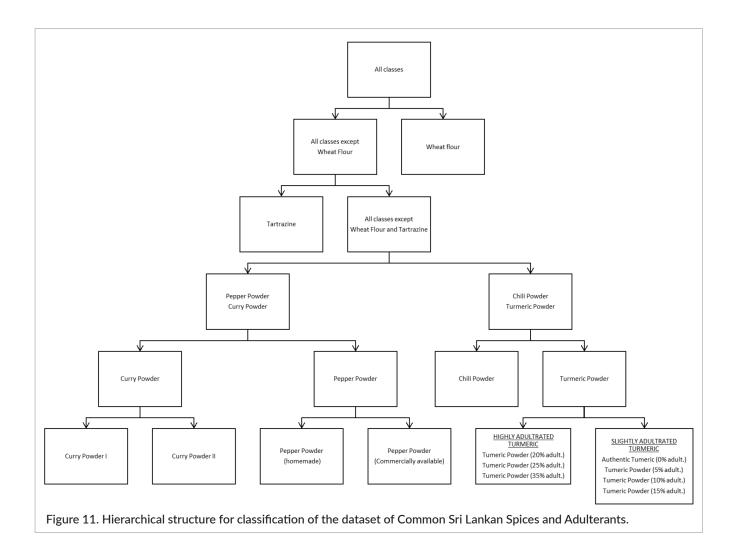
Figure 10. Modes of clustering of the image dataset of Common Sri Lankan Spices and Adulterants, at first stage.

stages. These numbers of clusters at different levels of zooming into the dataset were determined autonomously, based on the Eigen structure of the affinity matrix of the data set. The number of sub-clusters identified at each stage and the nature of the resulting clusters correspond to their physical properties.

In the first stage, shadows and painted metal sheets, which are the two classes that stand out from the rest, have been clustered separately. Shadows inherently reflect less compared to all other classes and painted metal sheets have a different chemical composition. It should also be noted that painted metal sheets have broken into two clusters due to the large variance present in the spectral signatures of the pixels containing them.

Trees and the green part of meadows which have similarities in both their colour and composition are combined together creating a separate umbrella cluster. Asphalt, gravel, bitumen, self-blocking bricks, bare soil and the brown part of meadows have combined together resulting in another umbrella cluster. This cluster contains black/brown material which consists of soil, minerals and other organic substances.

While the "green" cluster has broken down to its individual components in the second stage when zoomed in, the black/brown cluster has separated further into three sub-clusters, the first being the black sub-cluster, the second being the brown sub-cluster and the third being the cluster in between.



The first sub-cluster contains asphalt, part of the gravel and bitumen which are made of rocks. The second contains part bare soil and part meadows which have a mixed composition of soil and withered plants. The third class contains part gravel, self-blocking bricks and part bare soil. This is the sub-cluster containing a composition of soil and rocks in it. The final stage classifies all the pixels under their respective classes, identifying no further sub-clusters.

The results of the proposed algorithm have been compared with some of the well-known classification algorithms, both supervised and unsupervised, related to hyperspectral image classification in Table 2. The highest classification accuracies of all classes are marked in bold. These algorithms include Principal Component Analysis (PCA),⁶ LDA,⁸ Locally Linear Embedding (LLE),¹⁰⁻¹³ Feature Space Discriminant Analysis (FSDA),¹⁵ Nonparametric Feature Extraction (NFE),¹⁷ Simultaneous Orthogonal Matching Pursuit (SOMP),²³ Simultaneous

Subspace Pursuit (SSP)²³ and Spectral Clustering. FSDA is an algorithm which tries to produce features that are as different from each other as possible while increasing the separability between classes. The reduced dimensional data is then classified using SVM in this algorithm. NFE is a modification to LDA, weighted according to the degree of belonging of a data point to a particular class (weight calculated by the Euclidean distance to the class mean from the data point). The transformed data is classified using the SVM classifier in this algorithm. SOMP tries to decompose the pixel spectrum to a sparse linear combination of class signatures in a lower dimensional space while incorporating spatial information. Sparsity level and weighting factor are two of the user inputs used in SOMP.

Apart from the main contribution of generating the hierarchical structure of classification autonomously, the Umbrella Clustering algorithm has the advantage of producing accurate results using a smaller amount of training data, compared to what is used in state-of-the-

Table 1. Confusion matrix of classification—Case Study 1.

	Asphalt	Meadows	Gravel	Trees	Metal sheets	Bare soil	Bitumen	Bricks	Shadows	Total
Asphalt	4710	1	46	0	1	9	42	8	0	4817
Meadows	15	15939	3	442	0	705	0	26	0	17130
Gravel	462	4	1943	0	0	0	9	8	0	2426
Trees	0	1362	0	2598	0	5	0	0	0	3965
Metal sheets	38	0	0	4	1325	76	0	0	0	1443
Bare soil	31	710	0	5	2	3626	0	4	0	4378
Bitumen	822	5	17	3	3	521	1138	33	0	2542
Bricks	548	627	30	0	14	87	141	3663	0	5110
Shadows	5	1	0	12	0	0	0	0	947	965
Total pixels	6631	18649	2039	3064	1345	5029	1330	3742	947	42776

Table 2. Comparison of accuracy levels (class wise and overall) between different classification methods and the proposed method—Case Study 1.

	Accuracy (%) (to the nearest integer)								
								Spectral	Umbrella
Class	PCA	LDA	LLE	FSDA	NFE	SOMP	SSP	clustering	Clustering
Asphalt	82	65	57	60	73	59	70	41	71
Meadows	63	73	77	66	73	78	72	46	81
Gravel	79	49	50	67	73	84	74	53	93
Trees	93	79	92	91	89	97	95	90	85
Metal sheets	95	98	58	100	100	91	100	36	99
Bare soil	61	56	65	50	73	77	87	42	56
Bitumen	58	50	90	88	90	99	90	92	86
Bricks	40	55	62	76	76	89	90	96	99
Shadows	100	99	100	100	100	92	90	100	100
Overall	68	68	71	78	82	79	82	55	80

art algorithms. This is a major advantage in many practical situations. A comparison of accuracies among a number of state-of-the-art algorithms with a different number of training samples is illustrated in Figure 12. It shows that Umbrella Clustering performs well with a limited number of training samples compared to the other algorithms considered. It should be noted that, for all the methods illustrated in Figure 12, external data had been utilised in the validation process in order to generate results.

Machine learning algorithms have two types of parameters as fixed parameters and free variables. Fixed parameters are coded into the program. The free variables consist of both the user input variables and trainable parameters. The user input variables are input manually (thus they are subjective) and the trainable parameters are learned from the training data and used on the test data. The proposed algorithm has only one user input—the range of sigma values to sweep through. In contrast,

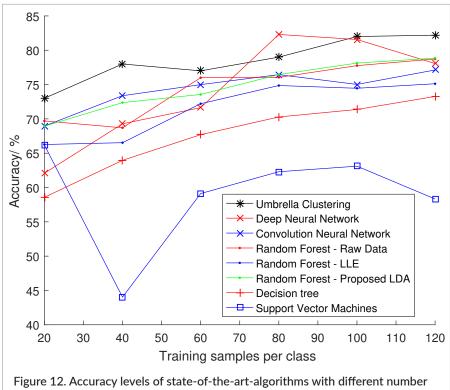
the neural-network-based algorithms have multiple such parameters (these solutions require updating the layer sizes, number of kernels etc. according to the dataset which requires user input/manual calibration). Having a lower number of user inputs makes the algorithm more versatile and the results less subjective.

The number of trainable parameters has a correlation with the complexity of the model and the amount of information that could be extracted from the dataset. For neural networks, this is the count of neuron weights²⁸ (around 80,000) and for linear classifiers, this is the number of the elements of the transformation matrices generated. Both those types of algorithms have fixed numbers of trainable parameters. The proposed solution has two types of trainable parameters—the elements of the LDA transformation matrices, sigma values for zooming into the dataset, cluster means of training data's labelled classes and the overall hierarchical structure. The number of trainable parameters is not fixed for the algorithm since the hierarchical structure is generated according to the training dataset. A variable number of trainable parameters makes the algorithm versatile over different scenarios (datasets and intended classifications) since the complexity of the model and the amount

of information extracted is adapting according to the scenario.

No method is globally superior to any other method and different methods have different strengths and weaknesses. Deep Learning will not require supervisory information but will consume a significant amount of training resources. However, the proposed method, as opposed to a purely LDA method, would not require that much supervisory information, yet a certain amount of supervisory information is required due to the spectral clustering aspect of it. Hence the comparative analysis does not provide a statement on the superiority of a method as opposed to another, but concludes on the different utilisation of training resources, which is significant for deep learning, whereas deep learning would continue to be superior in terms of the requirement of supervisory information.

The graph in Figure 12 demonstrates a universally accepted behaviour of algorithms, which is that for example, the deep learning methods will outperform all the other methods when given a reasonable amount of training information. However, if the situation lacks training resources, this has shown the deterioration of the said algorithms. However, algorithms such as the



of training samples.

proposed algorithms, which require a minimal amount of supervisory information while also having a minimum requirement of training information, will be superior under such circumstances.

In order to assess the versatility of the Umbrella Clustering algorithm, it was applied to the dataset of common Sri Lankan spices in case study 2. The illustration of the hierarchical structure for the classification of this data set, in Figure 11, shows that the structure is in line with human perception. The two super-clusters in the first stage succeed in separating "Wheat Flour" as it directly shows a dissimilarity with the other classes. In the next stage, similarly "Tartrazine" is separated due to the same reason. In the following stage, the sub-cluster containing the spices (Pepper, Chili, Turmeric and Curry powder) are taken into consideration. Applying the selforganising hierarchical method, this sub-cluster is divided into two more clusters—one containing pepper and curry powder and the other containing chili and turmeric. This again is in line with human perception with respect to colour and texture. The algorithm proceeds further to result in the hierarchical classification shown in Figure 11.

Conclusion

Hyperspectral image processing is an emerging technique used for feature detection and classification. This stands out from other image processing techniques due to its ability to identify fine features of an image which are not detectable by other image processing techniques.

The classification algorithm introduced in this paper is based on spectral clustering, which is applied to hyperspectral image data in a hierarchical manner. Before implementing spectral clustering, a modified LDA method is also introduced in order to increase between-class separability and decrease within-class scatter at each stage of the hierarchy. The algorithm organises the hierarchical structure of clustering autonomously considering the positional patterns of data points. This is more effective than single-staged spectral clustering, as shown in the comparative analysis in the results section.

The hierarchical classification makes sense in a physical aspect as well, resulting in large clusters combining data points belonging to different classes with similar physical properties. Results of each stage can be interpreted based on their physical characteristics. The splitting of a few of the original classes into different umbrella clusters

conveys large physical disparities within those classes which were not available in the ground truth of the image. It also enables the identification of such high variance classes with large intra-class distance.

Umbrella Clustering has the ability to produce high levels of accuracy using a smaller number of training samples, compared to the state-of-the-art algorithms as shown in Figure 12. It also has the advantage of having only one user input and one trainable parameter which limits the degrees of freedom of the algorithm, resulting in improved efficiency.

The algorithm proposed in this paper has been able to reach a class-wise accuracy of more than 70% for all the classes in the Pavia University dataset, with four classes being classified with an accuracy level greater than 93%.

In order to assess the versatility of the proposed Umbrella Clustering algorithm, it was applied to the dataset of Common Sri Lankan Spices and Adulterants, which consists of 16 classes that are different. The resulting adaptive hierarchical structure was in line with human perception, and the results of classification were comparably high, compared to recently published work.

Based on the results of the case studies, it could be concluded that the proposed Umbrella Clustering algorithm develops its hierarchical clustering structure autonomously, which coincides with the human perception and performs well in terms of accuracy. The proposed algorithm also has the ability to adapt to different types of datasets and generate the hierarchical structure of clustering autonomously, confirming its versatility in implementation.

Acknowledgement

We are sincerely grateful to the Computational Intelligence Group of the University of the Basque Country for providing the hyperspectral image databases of the Pavia University scene, which the proposed approach was experimented upon.

References

1. Z. Liu, B. Tang, X. He, Q. Qiu and F. Liu, "Class specific random forest with cross correlation constraints for spectral–spatial hyperspectral image classifica-

- tion", IEEE Geosci. Remote Sens. Lett. **14**, 257–261 (2017). https://doi.org/10.1109/LGRS.2016.2637561
- G. Lu and B. Fei, "Medical hyperspectral imaging: a review", J. Biomed. Optics 19, 010901 (2014). https://doi.org/10.1117/1.JBO.19.1.010901
- **3.** C. Zhang, C. Guo, F. Liu, W. Kong, Y. He and B. Lou, "Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine", *J. Food Eng.* **179**, 11–18 (2016). https://doi.org/10.1016/j.jfoodeng.2016.01.002
- 4. E.M.M.B. Ekanayake, E.M.H.E.B. Ekanayake, A.R.M.A.N. Rathnayake, S.S.P. Vithana, H.M.V.R. Herath, G.M.R.I. Godaliyadda and M.P.B. Ekanayake, "A semi-supervised algorithm to map major vegetation zones using satellite hyperspectral data", 2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). Amsterdam, Netherlands, pp. 1–5 (2018). https://doi.org/10.1109/ WHISPERS.2018.8747025
- **5.** D.M.D.W. Ronghua Yan and J. Peng, "Dimensionality reduction method based on a tensor model", *J. Appl. Remote Sens.* **11**, 025011 (2017). https://doi.org/10.1117/1.JRS.11.025011
- **6.** T. Raiko, A. Ilin and J. Karhunen, *Principal Component Analysis for Large Scale Problems with Lots of Missing Values*. Warsaw, Poland (2007).
- J. Zabaiza, M. Ren, Y. Yang, Y. Zhang, J. Wang, S. Marshall and J. Han, "Novel Folded PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing", ISPRS J. Photogramm. Remote Sens.
 112–122 (2014). https://doi.org/10.1016/j.isprsjprs.2014.04.006
- **8.** M. Welling, *Fisher Linear Discriminant Analysis*. Department of Computer Science, University of Toronto 3(1) (2005).
- **9.** M. Sugiyama, "Local Fisher Discriminant Analysis for supervised dimensionality reduction", in *ICML* '06 Proceedings of the 23rd International Conference on Machine Learning, pp. 905–912. Pittsburgh, PA, USA (2006). https://doi.org/10.1145/1143844.1143958
- **10.** L. Zhang and C. Zhao, "Sparsity divergence index based on locally linear embedding for hyperspectral anomaly detection", *J. Appl. Remote Sens.* **10**, 025026 (2016). https://doi.org/10.1117/1.JRS.10.025026
- **11.** Y. Fang, H. Li, Y. Ma, K. Liang, Y. Hu, S. Zhang and H. Wang, "Dimensionality reduction of hyperspec-

- tral images based on robust spatial information using locally linear embedding", *IEEE Geosci. Remote Sens. Lett.* **11**, 1712–1716 (2014). https://doi.org/10.1109/LGRS.2014.2306689
- **12.** L. Sha, D. Schonfeld and J. Wang, "Locally linear embedded sparse coding for image representation", *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, pp. 2527–2531 (2017). https://doi.org/10.1109/ICASSP.2017.7952612
- **13.** G. Chen, "Dimensionality reduction of hyperspectral imagery using improved locally linear embedding", *J. Appl. Remote Sens.* **1,** 013509 (2007). https://doi.org/10.1117/1.2723663
- **14.** H. Huang, F. Luo, J. Liu and Y. Yang, "Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding", *ISPRS J. Photogramm. Remote Sens.* **106**, 42–54 (2015). https://doi.org/10.1016/j.isprsjprs.2015.04.015
- **15.** M. Imani and H. Ghassemian, "Feature space discriminant analysis for hyperspectral data feature reduction", *ISPRS J. Photogramm. Remote Sens.* **102**, 1–13 (2015). https://doi.org/10.1016/j.isprsjprs.2014.12.024
- 16. A.M. Fahim, A.M. Salem, A. Torkey and M.A. Ramadan, "An efficient enhanced k-means clustering algorithm", J. Zhegiang University-SCIENCE A 7, 1626–1633 (2006). https://doi.org/10.1631/jzus.2006.A1626
- **17.** A. Kianisarkaleh and H. Ghassemian, "Nonparametric feature extraction for classification of hyperspectral images with limited training samples", *ISPRS J. Photogramm. Remote Sens.* **119**, 64–78 (2016). https://doi.org/10.1016/j.isprsjprs.2016.05.009
- **18.** E. Mernyi, W.H. Farrand, J.V. Taranik and T.B. Minor, "Classification of hyperspectral imagery with neural networks: comparison to conventional tools", *EURASIP J. Adv. Signal Proc.* 2014 (2014). https://doi.org/10.1186/1687-6180-2014-71
- 19. N. Gillis, D. Kuang and H. Park, "Hierarchical clustering of hyperspectral images using rank two nonnegative matrix factorization", *IEEE Trans. Geosci. Remote Sens.* 53, 2066–2078 (2015). https://doi.org/10.1109/TGRS.2014.2352857
- **20.** L. Hosseini and R.S. Kandovan, "Hyperspectral image classification based on hierarchical SVM algorithm for improving overall accuracy", Adv. Remote

- Sens. **6(1)**, 66–75 (2017). https://doi.org/10.4236/ ars.2017.61005
- **21.** J.C. Burges, "A tutorial on support vector machines for pattern recognition", *Data Min. Knowl. Discover.* **2,** 121–167 (1998). https://doi.org/10.1023/A:1009715923555
- **22.** S. Lee and M. Crawford, "Hierarchical clustering approach for unsupervised image classification of hyperspectral data", *IEEE International Symposium Geoscience and Remote Sensing (IGARSS)*. IEEE, Vol. 2, pp. 941–944 (2004). https://doi.org/10.1109/IGARSS.2004.1368563
- **23.** Y. Chen, N.M. Nasrabadi and T.D. Tran, "Hyperspectral image classification using dictionary-based sparse representation", *IEEE Trans. Geosci. Remote Sens.* **49,** 3973–3985 (2011). https://doi.org/10.1109/TGRS.2011.2129595
- **24.** Y. Tarabalka, J. Benediktsson and J. Chanussot, "Spectral–Spatial classification of hyperspectral imagery based on partitional clustering techniques", *IEEE Trans. Geosci. Remote Sens.* **47**, 2973–2987 (2009). https://doi.org/10.1109/TGRS.2009.2016214
- **25.** X. Sun, Y. Lina, L. Gao, B. Zhang, S. Li and J. Li, "Hyperspectral image clustering method based on artificial bee colony algorithm and Markov random fields", **9**, 095047 (2015). https://doi.org/10.1117/1.JRS.9.095047
- 26. J. Zhao, Y. Zhong, T. Jia, X. Wang, Y. Xu, H. Shu and L. Zhang, "Spectral-spatial classification of hyperspectral imagery with cooperative game", *ISPRS J. Photogramm. Remote Sens.* 35, 31–42 (2018). https://doi.org/10.1016/j.isprsjprs.2017.10.006
- **27.** W. Zeng, D. Zhang, Y. Fang, J. Wu and J. Huang, "Comparison of partial least square regression, support vector machine, and deep learning techniques for estimating soil salinity from hyperspectral data", *J. Appl. Remote Sens.* **12(2),** 022204 (2018). https://doi.org/10.1117/1.JRS.12.022204
- 28. M. Zhang, W. Li and Q. Du "Collaborative classification of hyperspectral and visible images with convolutional neural network", *J. Appl. Remote Sens.* 11(4), 042607 (2017). https://doi.org/10.1117/1.JRS.11.042607

- **29.** G. Abdi and F. Samadzadegan, "Spectral-spatial feature learning for hyperspectral imagery classification using deep stacked sparse autoencoder", *J. Appl. Remote Sens.* **11(4),** 042604 (2017). https://doi.org/10.1117/1.JRS.11.042604
- 30. G. Abdi, F. Samadzadegan and P. Reinartz, "Deep learning decision fusion for the classification of urban remote sensing data", J. Appl. Remote Sens. 12(1), 016038 (2018). https://doi.org/10.1117/1.JRS.12.016038
- **31.** A.Y. Ng, M.I. Jordan and Y. Weiss, "On spectral clustering: Analysis and an algorithm", in *Advances in Neural Information Processing Systems*. MIT Press, pp. 849–856 (2001).
- **32.** Hyperspectral Remote Sensing Scenes—Grupo de Inteligencia Computacional (GIC) (2014).
- **33.** C. Bandara, K. Prabhath, M. Ekanayake, V.R. Herath and R.I. Godaliyadda, A Multispectral Image Database of Common Sri Lankan Spices and Adulterants.

 Mendeley Data, v3 (2019). http://doi.org/10.17632/chffh6pdnw.4
- **34.** J. Kustra, A. Jalba and A. Telea, "Robust segmentation of multiple intersecting manifolds from unoriented noisy point clouds", *Comput. Graphics Forum* **33**, 73–87 (2014). https://doi.org/10.1111/cgf.12255
- 35. X. Cao, B. Ji, Y. Ji, L. Wang and L. Jiao,
 "Hyperspectral image classification based on filtering: a comparative study", *J. Appl. Remote Sens.*11(3), 035007 (2017). https://doi.org/10.1117/1.JRS.11.035007
- **36.** P. Wang, S. Xu, Y. Li. J. Wang and S. Liu, "Hyperspectral image classification based on joint sparsity model with low-dimensional spectral-spatial features", *J. Appl. Remote Sens.* **11(1)**, 015010 (2017). https://doi.org/10.1117/1.JRS.11.015010
- 37. W.G.C. Bandara, G.W.K. Prabhath,
 D.W.S.C.B. Dissanayake, V.R. Herath,
 G.M.R.I. Godaliyadda, M.P. Ekanayake, S. Vithana,
 D. Demini and T. Madhujith, "A multispectral imaging system to assess meat quality", in 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (R10 HTC'18). Colombo, Sri Lanka (2018). https://doi.org/10.1109/R10-HTC.2018.8629858