# BandClust: An Unsupervised Band Reduction Method for Hyperspectral Remote Sensing

Claude Cariou, Kacem Chehdi, and Steven Le Moan

***Abstract*—We address the problem of unsupervised band reduction in hyperspectral remote sensing imagery. We propose the use of an information theoretic criterion to automatically separate the sensor's spectral range into disjoint subbands without ground truth knowledge. Our approach, named BandClust, preserves the physical sense of the spectral data and automatically provides relevant spectral subbands, i.e., of maximal informational complementarity. Experiments using real hyperspectral images are conducted to compare BandClust with four other unsupervised approaches. The comparison of the selected dimensionality reduction methods is performed via supervised classification using support vector machines and shows the potential of the proposed approach.**

***Index Terms*—Dimensionality reduction, feature extraction, hyperspectral images, information theory, remote sensing, supervised classification.**

## I. Introduction

HYPERSPECTRAL imaging consists of acquiring a scene with numerous and contiguous spectral bands (from several tens to a few hundreds of bands) centered on uniformly distributed wavelengths ranging in the visible to near-infrared (VNIR) spectrum. This technology is now fully operational in many applicative fields, including remote sensing, astronomy, medicine, food safety, and forensics [1].

This letter addresses the problem of band reduction in hyperspectral remote sensing imagery. Hyperspectral imaging offers high richness of information which is often necessary to achieve good classification performance at the pixel level. However, several problems occur when using hyperspectral images. One problem comes from the high dimensionality of hyperspectral data which is often an obstacle for data classification and interpretation, since there are generally not enough samples to fill the hyperspace spanned by the large set of variables and thus to infer some structure within the data. This problem of space emptiness is related to the well-known *curse of dimensionality* [2]. Another problem is due to the large size of hyperspectral images which makes them difficult to store and handle. Most images can amount up to several gigabytes for a single scene acquired from an airborne hyperspectral platform. Hyperspectral images generally show a high amount of correlation between adjacent spectral bands. Therefore, removing this redundancy would reduce the amount of data that are relevant to further classification and interpretation stages. Moreover, a band reduction technique which could help to automatically determine the optimal band set for a given application from available hyperspectral data would be useful in a band programming procedure prior to acquisition. Band reduction is thus desirable for many practical reasons.

Several approaches exist for dimensionality reduction in hyperspectral data and can be split into two major groups. The first group relates to *transformation-based approaches*. Such methods aim at projecting the original data set onto adequate subspaces, chosen for their relevance to explain the data. This group includes not only linear methods such as principal component (PC) analysis (PCA), projection pursuit [3], and independent component analysis [4] but also nonlinear methods such as kernel PCA [5]. A detailed experimental analysis of several transformation approaches (although not using hyperspectral data) shows that nonlinear transformation methods for dimensionality reduction do not outperform linear approaches in average [6]. These methods, when applied to hyperspectral data, suffer from a lack of "explainability" from a physical viewpoint since, generally, the original spectral information is not preserved. The second group includes *feature-selection-based approaches*. Most of them require knowledge of the ground truth. Some methods use the Jeffries–Matusita distance [7] between multidimensional class-conditional distributions. Many others use either entropy or mutual information [8]–[10] to derive appropriate band selection criteria. Due to the high correlation between neighboring bands, several selection-based methods proceed by sequential forward selection [11] or by grouping similar bands in a bottom-up way. Only few of these approaches are partially unsupervised [12] or totally unsupervised [10], [13]. In [13], a linear prediction or orthogonal subspace projection (LP/OSP) approach is proposed which tries, at each step, to select the spectral band which is the most dissimilar in the $l_2$ sense to its prediction from the previously selected bands. In [10], the authors propose the use of a hierarchical band-clustering strategy based on Ward's linkage [14] which can involve different correlation measures, such as the mutual information for *Ward's linkage strategy using mutual information* (WaLuMI) and the spectral information divergence [15] for *Ward's linkage strategy using divergence* (WaLuDi). More precisely, WaLuMI and WaLuDi sequentially merge the two most similar bands until a specified number of bands are reached and then choose one representative band for each band cluster formed.

Manuscript received July 12, 2010; revised October 8, 2010; accepted November 3, 2010. Date of publication December 16, 2010; date of current version April 22, 2011.

C. Cariou and K. Chehdi are with the Laboratoire de Traitement des Signaux et Images Multicomposantes et Multimodales, École Nationale Supérieure des Sciences Appliquées et de Technologie, Université de Rennes 1, 22305 Lannion, France (e-mail: cariou@enssat.fr; chehdi@enssat.fr).

S. Le Moan is with the Laboratoire Electronique, Informatique et Image, UMR CNRS 5158, 89010 Auxerre, France.

Digital Object Identifier 10.1109/LGRS.2010.2091673

In this letter, we propose an unsupervised approach to band reduction for hyperspectral images named BandClust, which stands for *band clustering*. It consists in *splitting* the initial range of spectral bands into disjoint clusters or subbands. BandClust has several benefits: First, it does not require any ground truth knowledge; second, it automatically provides an estimate of the optimal number of bands for further classification purposes; third, it does preserve the physical meaning of the hyperspectral data, which is useful in a band specification procedure for a given application; finally, it is easy to implement.

In Section II, we describe the proposed method in detail. In Section III, we present some experimental results obtained on three hyperspectral data sets. More precisely, we compare the performances of BandClust with four other unsupervised approaches: PCA, LP/OSP [13], WaLuMI, and WaLuDi [10]. The comparison is performed in terms of overall accuracies (OAs) and kappa indices ($\kappa$) obtained by each band reduction technique in supervised support vector machine (SVM) classification. We conclude and give some perspectives of this work in Section IV.

## II. Proposed Method

Hyperspectral data can be thought as a collection $\boldsymbol{X} = \{\boldsymbol{x}_i = [x_i(1), \ldots, x_i(B)]; i = 1, \ldots, N\}$ of realizations of $B$-dimensional random (rowwise) vectors. $B$ is the original number of spectral bands (or features), and $N$ is the number of collected spectral signatures. Hyperspectral images (also known as data cubes) can be easily rearranged in this manner. Let $[b_{\min}, b_{\max}]$ be some interval of the original band index range $[1, B]$. The basis of BandClust is to split $[b_{\min}, b_{\max}]$ into two disjoint contiguous subbands.

Let $\boldsymbol{s}_1(b) = [s_{1,1}(b), \ldots, s_{1,N}(b)]^T$ ($\boldsymbol{s}_2(b) = [s_{2,1}(b), \ldots, s_{2,N}(b)]^T$) be the (column) vector obtained by averaging each row vector $\boldsymbol{x}_i$, with $i = 1, \ldots, N$, over the range of spectral bands $[b_{\min}, b]$ ($[b, b_{\max}]$). The $i$th components of $\boldsymbol{s}_1(b)$ and $\boldsymbol{s}_2(b)$ are given by

$$\begin{cases} s_{1,i}(b) = \frac{1}{(b - b_{\min}+1)} \sum_{k=b_{\min}}^{b} x_i(k) \\ s_{2,i}(b) = \frac{1}{(b_{\max} - b+1)} \sum_{k=b}^{b_{\max}} x_i(k). \end{cases} \quad (1)$$

The vectors $\boldsymbol{s}_1(b)$ and $\boldsymbol{s}_2(b)$ can then be seen as collections of realizations of two random variables (RVs) $S_1(b)$ and $S_2(b)$.

The principle of BandClust is to find (provided that it exists) the band index $b_{\text{opt}} \in ]b_{\min}, b_{\max}[$ which minimizes the following mutual information criterion:

$$\mathcal{J}_{b_{\min}, b_{\max}}(b) = I\left(S_1(b); S_2(b)\right). \quad (2)$$

Let us recall that the mutual information between two real-valued RVs $S_1$ and $S_2$ is given by

$$I(S_1; S_2) = \int_{\mathbb{R}^2} f_{S_1,S_2}(s_1, s_2) \log \frac{f_{S_1,S_2}(s_1, s_2)}{f_{S_1}(s_1) \cdot f_{S_2}(s_2)} ds_1 ds_2 \quad (3)$$

where $f_{S_1,S_2}$ is the joint probability density function (pdf) of $S_1$ and $S_2$ and $f_{S_1}$ and $f_{S_2}$ are the marginal pdfs. From a practical viewpoint, the pdfs required for BandClust are replaced by histogram estimates after quantization of the data into 32 bins between the minimum and maximum values of each subband.

The criterion in (2) ensures that the two subbands created by splitting the interval $[b_{\min}, b_{\max}]$ share the minimum of mutual information and thus are the most statistically independent from each other. It is important to notice that, under this setting and assuming a uniform spectral sampling interval, the quantities $s_{1,i}(b)$ and $s_{2,i}(b)$ represent the actual average radiance (or reflectance) within the obtained spectral subbands for a given pixel or spectral measurement.

The search for the optimal band index $b_{\text{opt}}$ given an initial range of indices $[b_{\min}, b_{\max}]$ can be set up in different ways. We have chosen to use a greedy procedure which consists of explicitly computing $\mathcal{J}_{b_{\min}, b_{\max}}(b)$ and selecting $b_{\text{opt}}$ as the band index minimizing this criterion, provided that a minimum exists. In order to smooth local variations of the criterion and better detect its global minimum, the values of $\mathcal{J}_{b_{\min}, b_{\max}}(b)$ are filtered using a Gaussian filter $f_\sigma$ with standard deviation $\sigma$.

The BandClust method is described in Algorithm 1. The splitting of the initial band index range $[1, B]$ into finer subbands is performed by successive applications of the same procedure over each previously created subband. More precisely, once a subband $[b_k, b_{k+1}]$ has been processed, the neighboring subband to its right is processed next ($k \leftarrow k + 1$) until $b_{k+1} = B$. The algorithm then performs another loop starting from the leftmost subband. BandClust therefore allows splitting, in a balanced manner, the whole spectral range of band indices $[1, B]$ into as many subbands as necessary to satisfy the splitting criterion. Therefore, BandClust is essentially a recursive binary band-splitting algorithm. The algorithm stops whenever the subbands previously created cannot further be split, i.e., no minimum of criterion (2) can be found between any pair of consecutive band limits. The final number of subbands is then considered as the optimal one.

**Algorithm 1** BandClust algorithm
**Initialization:**
  Given $\boldsymbol{X} = \{\boldsymbol{x}_i = [x_i(1), \ldots, x_i(B)]; i = 1, \ldots, N\}$ a matrix of (rowwise) spectral signatures;
  $b_1 = 1$, $b_2 = B$;
  Define $\mathcal{L} = \{b_1, b_2\}$; (set of band limits)
  $m = 2$; (# of band limits in $\mathcal{L}$)
**Iterations:**
  **repeat**
    $k = 1$; $\mathcal{L}_{\text{temp}} = \varnothing$;
    **while** $k < m$ **do**
      $b_{\min} = b_k$, $b_{\max} = b_{k+1}$;
      **if** $(\exists b_{\text{opt}} : b_{\text{opt}} = \arg\min\{[\mathcal{J}_{b_{\min}, b_{\max}} * f_\sigma](b)\})$ **then**
        $\mathcal{L}_{\text{temp}} \leftarrow \mathcal{L}_{\text{temp}} \cup b_{\text{opt}}$;
      **end if**
      $k \leftarrow k + 1$;
    **end while**
    $\mathcal{L} \leftarrow \mathcal{L} \cup \mathcal{L}_{\text{temp}}$;
    $m = \text{card}(\mathcal{L})$;
    Sort $\mathcal{L}$ by ascending band index order: $\mathcal{L} = \{b_1 = 1, \ldots, b_m = B\}$;
  **until** $\mathcal{L}$ is unchanged

$M = m - 1$;
**for** $i = 1 : N$ **do**
  **for** $j = 1 : M$ **do**
    $s_i(j) = (1/(b_{j+1} - b_j + 1)) \sum_{k=b_j}^{b_{j+1}} x_i(k)$;
  **end for**
**end for**
**return** $M$ (final # of bands)
**return** $\mathcal{L}$ (set of band limits)
**return** $\{s_i(j); 1 \leq i \leq N; 1 \leq j \leq M\}$ (reduced data set)

It is noteworthy that, under this setting, BandClust cannot output a specified number of bands to retain. On the one hand, this could be considered as a limitation of our approach for some applications where a fixed number of bands are required. On the other hand, this makes BandClust fully unsupervised since it automatically provides an optimal number of band clusters at the end of the algorithm, with each cluster being represented by its band average.

## III. Experimental Results

In this section, we provide the experimental results using three hyperspectral images. More precisely, we compare BandClust with four other unsupervised approaches: PCA, LP/OSP [13], WaLuMI, and WaLuDi [10]. These methods are applied on the same data for each hyperspectral image. Due to the strong limitation of the LP/OSP approach with respect to the size of the required data, only a part of the original data set is used depending on the size of the data set considered; indeed, the dimension of the orthogonal subspace projection matrix required by this method is $N$ by $N$, with $N$ being the number of available data vectors [13]. Therefore, we have limited $N$ to less than 3000 vectors. For each data set, the retained band indices, PCs for PCA, or subband limits $\mathcal{L}$ for BandClust, obtained from the said part of the original data, are used to perform the reduction on the original number of pixels in each data set.

The assessment of each band reduction approach is performed via supervised classification using SVMs [16]. One-against-one multiclass SVMs are trained, owing to the availability of ground truth, and tested on the reduced data sets. For this, we use the *LIBSVM* software [17]. The Gaussian radial basis function (RBF) is the kernel chosen in the SVM implementation [18]. In all the experiments, a fivefold cross-validation procedure is performed to optimize the parameters of the SVM, i.e., the aperture $\gamma$ of the Gaussian RBF and the penalty parameter $C$. The OA and kappa index of agreement are computed for each band reduction method.

### A. Data Sets

The first hyperspectral data used in this experiment are the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) 92AV3C *Indian Pine* data, originally consisting of 220 spectral bands in the [0.4–2.5]-$\mu$m VNIR range [19]. Each band is a 145 $\times$ 145 pixel image, but only nearly half of the pixels (10 366) have a ground truth label into one of 16 classes. These almost include different vegetation classes and a few of man-made materials. This data set has been extensively used in the literature for hyperspectral dimensional reduction and classification purposes [7], [20]. In our experiments, we removed, as specified in [20], the most noisy bands from this data set and kept 185 spectral bands as the initial data for every method. More precisely, we removed bands [1–3], [103–112], [148–165], and [217–220]. For band reduction, we used only 2627 pixels scattered throughout the labeled pixels to determine the bands to be selected/averaged. For the assessment by SVM classification, we randomly selected from the original labeled data set roughly one-half of the labeled pixels (5169) for training and the other half (5197) for computing the OA and kappa index. For the AVIRIS data set, we have taken $\sigma = 0.5$ for the standard deviation of the Gaussian filter used for criterion filtering in BandClust. This choice was made after several tests $(0 < \sigma \leq 2)$ and was found to give a good compromise between the number of retained bands and their relevance for supervised classification.

The second data set is the *University of Pavia* hyperspectral image acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor on behalf of the Deutsches Zentrum für Luft- und Raumfahrt (Germany). This datum consists of 103 spectral bands in the range [430–770] nm, after noisy band removal. The spatial dimension of this image is 610 $\times$ 340 pixels, and the ground truth consists of 43 923 pixels labeled among nine classes and split into 3921 pixels for training and 40 002 for testing. Only 2821 pixels are used to compute the relevant bands to be selected or averaged, which, in turn, are used to perform the band reduction on the whole data set.

The third data set was acquired by our laboratory on behalf of the U.K. Institute of Grassland and Environmental Research in June 2005 with a Compact Airborne Spectrographic Imager (CASI) sensor over a field plot $(52°16'30.00'' \text{ N}-4°11'23.40'' \text{ W})$ in Llanon, Wales, U.K., during the Interreg IIIB Atlantic Area European project PIMHAI. It consists of 48 spectral bands in the [405–947]-nm range. Some 747 pixels measure the spectral radiance and are assumed to belong to one of four classes, corresponding to two different barley species (*Optic* and *Golden Promise*) which are infected or not by mildew after spraying pesticide. A ground truth image is available for these data. Fig. 1 shows sample bands of the hyperspectral CASI image, as well as the ground truth image. The image is not geometrically corrected to avoid mixing radiances of neighboring pixels. From this set, two disjoint subsets with approximately halves of the data are created for supervised classification, i.e., 356 pixels for training and 391 pixels for testing. The whole data set is used for band reduction.

No filtering of the criterion in (2) was performed for the ROSIS and CASI data sets.

### B. Results

Figs. 2–4 show the OAs of the supervised SVM classification for the three hyperspectral image data sets. For each method and each data set, the OA is plotted versus the number of retained bands. We have chosen to limit the number of retained bands to 20 for each method. For each data set, the OA obtained using the whole set of original bands is also plotted for comparison. Since BandClust ends with an optimal number of bands, only the last point should be considered on the graph plots for this method. Nevertheless, to better compare our method with the

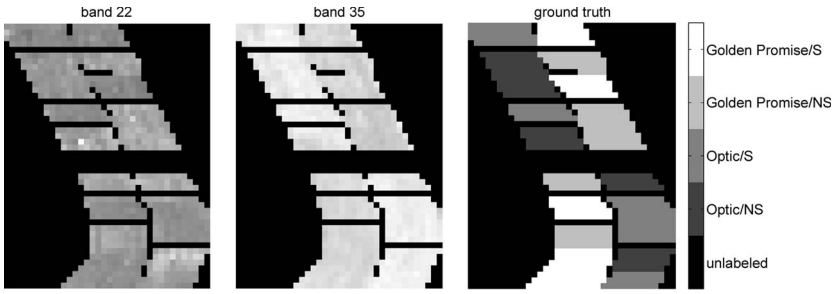


Fig. 1. Sample bands from the CASI hyperspectral image used in the experiments and associated ground truth data (black pixels are not labeled; Optic and Golden Promise are the two barley species; NS: not sprayed with pesticides; S: sprayed with pesticides). Band 22 is located at (643.7 $\pm$ 6.2) nm; band 35 is located at (792.6 $\pm$ 6.3) nm.

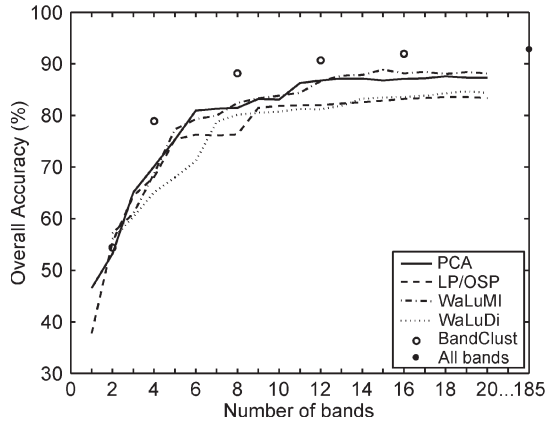


Fig. 2. Comparison of OAs versus selected number of bands for the AVIRIS *Indian Pine* data set.

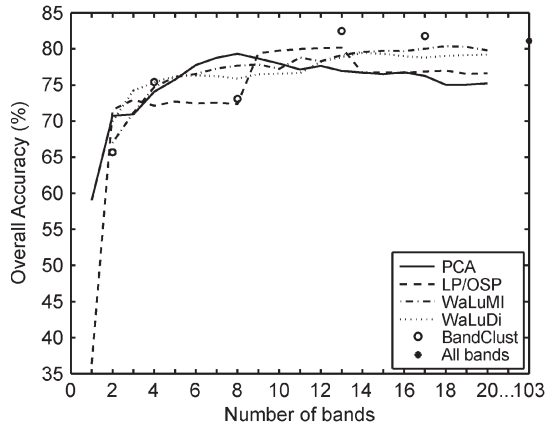


Fig. 3. Comparison of OAs versus selected number of bands for the ROSIS *University of Pavia* data set.

other ones, we also computed the OAs for smaller numbers of retained bands, i.e., at the end of each `while` loop in Algorithm 1. This corresponds to the scanning of previously created subbands for new criterion minima along the whole band index range.

For the AVIRIS data set, the averaged subbands given by BandClust provide the best overall classification accuracies among all the unsupervised methods for most of identical numbers of selected bands. BandClust stopped with an optimal number of 16 averaged bands and a corresponding 91.96% OA. Comparatively, for 16 retained bands, the OAs of the other methods are (in decreasing order) 88.17% for WaLuMI, 87.09% for PCA, 83.59% for WaLuDi, and 83.24% for LP/OSP.

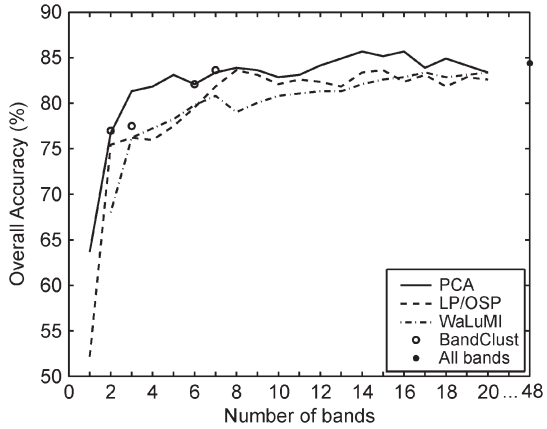


Fig. 4. Comparison of OAs versus selected number of bands for the CASI *Llanon* data set.

This is in accordance with the results in [10] which show a slightly better behavior of WaLuMI over WaLuDi in terms of classification accuracy.

For the ROSIS *University of Pavia* image, BandClust gives an optimal number of 17 averaged bands with an 81.80% OA and once again outperforms the other methods for this specific number of bands. However, the highest OA (82.48%) of BandClust is obtained for only 13 retained bands. Noticeable is the behavior of the OA for PCA, which attains its maximum value for the first eight PCs and then slowly decreases as the number of PCs increases. WaLuMI also gives slightly better OAs than WaLuDi in the range from 15 to 20 selected bands. The behavior of the curve for the LP/OSP method looks very "chaotic" with highly pronounced changes in the OA for adjacent numbers of retained bands.

The results for the CASI *Llanon* data set are more contrasted. The maximum OA of 85.68% is attained by PCA for 14 PCs. However, our method provides the highest OA (83.63%) among all the methods for the same number of seven bands which is found to be the optimal one. Notice that WaLuDi, for some unknown reason, fails to output any set of selected bands with the CASI data.

Table I gives the OAs, the kappa indices, and the number of retained bands corresponding to the best OA results of each of the PCA, LP/OSP, WaLuMI, and WaLuDi methods on each of the three data sets, as well as the results obtained using all the bands of the original data sets. For BandClust, we have given the OA and kappa indices corresponding to the optimal number of averaged bands found. One can observe that, for the

TABLE I
BEST OAS (IN PERCENTAGE), KAPPA INDICES ($\kappa$), AND CORRESPONDING NUMBER OF RETAINED BANDS FOR EACH METHOD ON EACH OF THE THREE DATA SETS

| Dataset | AVIRIS | | | ROSIS | | | CASI | | |
|---|---|---|---|---|---|---|---|---|---|
| | OA | $\kappa$ | # bands | OA | $\kappa$ | # bands | OA | $\kappa$ | # bands |
| PCA | 87.61 | 0.8582 | 18 | 79.34 | 0.7320 | 8 | **85.68** | **0.8087** | **14** |
| LP/OSP | 83.59 | 0.8127 | 19 | 80.19 | 0.7467 | 13 | 83.63 | 0.7813 | 8 |
| WaLuMI | 88.88 | 0.8731 | 15 | 80.41 | 0.7496 | 18 | 83.38 | 0.7778 | 17 |
| WaLuDi | 84.66 | 0.8247 | 19 | 79.47 | 0.7379 | 14 | – | – | – |
| BANDCLUST | **91.96** | **0.9083** | **16** | **81.80** | **0.7661** | **17** | 83.63 | 0.7816 | 7 |
| *All bands* | *92.86* | *0.9186* | *185* | *81.13* | *0.7600* | *103* | *84.40* | *0.7915* | *48* |

AVIRIS and ROSIS data sets, BandClust gives better results than the other unsupervised band reduction methods compared. However, this is not the case for the CASI data set, where PCA provides the best OA among the band reduction methods, followed by LP/OSP and BandClust. Also, it is remarkable that, for the ROSIS data set, BandClust gives a better classification result (OA and kappa index) than using the whole band set.

## IV. CONCLUSION AND PERSPECTIVES

In this letter, we have proposed a top-down approach for hyperspectral dimensionality reduction. This approach is unsupervised, i.e., it does not require either ground truth data or final number of retained bands. BandClust works by iteratively splitting a band interval into two finer bands, based upon a criterion of minimization of the mutual information between averaged subbands. This method is very easily implementable and was found to provide encouraging results on real hyperspectral data. Compared to PCA or recently published methods (LP/OSP [13], WaLuMI, and WaLuDi [10]), BandClust performs well as evidenced by the OAs obtained after supervised classification using multiclass SVM. The experiments also confirm the superiority of the information theoretic paradigm to cope with general independence, compared to an $l_2$ distance-based methodology: Indeed, BandClust, WaLuMI, and WaLuDi have shown better performances, in average, than LP/OSP in terms of overall classification accuracy.

Several extensions of this work can be foreseen, such as the refinement of band limits found in a further pass, in the same spirit as the (supervised) procedure described in [7]. Indeed, BandClust could be improved by optimizing the limits found at a previous stage of the band formation procedure. Another possible extension which we have started to investigate is the use of a mutual information criterion involving several RVs; this could allow one to manage and optimize multiple band limits at the same time through an appropriate criterion.

## ACKNOWLEDGMENT

The authors would like to thank D. Landgrebe at Purdue University, West Lafayette, IN, P. Gamba at the University of Pavia, Pavia, Italy, and A. Gay at the Institute of Grassland and Environmental Research, Aberystwyth, U.K., for providing the hyperspectral images used in this work and also A. Martínez-Usó at the Jaume I University, Castellón de la Plana, Spain, for making the WaLuMI and WaLuDi codes available.

## REFERENCES

[1] H. Grahn and P. Geladi, Eds., *Techniques and Applications of Hyperspectral Image Analysis*. Chichester, U.K.: Wiley, 2007.
[2] R. Bellman, *Adaptive Control Processes*. Princeton, NJ: Princeton Univ. Press, 1961.
[3] J. Friedman and J. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Comput.*, vol. C-23, no. 9, pp. 881–890, Sep. 1974.
[4] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
[5] B. Scholkopf, A. Smola, and K. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
[6] L. van der Maaten, E. Postma, and H. van den Herik, "Dimensionality reduction: A comparative review," Tilburg Univ., Tilburg, The Netherlands, Tech. Rep. TiCC-TR 2009-005, 2009.
[7] S. Serpico and G. Moser, "Extraction of spectral channels from hyperspectral images for classification purposes," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 484–495, Feb. 2007.
[8] B. Guo, R. Damper, S. Gunn, and J. Nelson, "A fast separability-based feature-selection method for high-dimensional remotely sensed image classification," *Pattern Recognit.*, vol. 41, no. 5, pp. 1653–1662, May 2008.
[9] S. Prasad and L. Bruce, "Decision fusion with confidence-based weight assignment for hyperspectral target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1448–1456, May 2008.
[10] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
[11] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.
[12] B. Guo, S. Gunn, R. Damper, and J. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, Oct. 2006.
[13] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.
[14] J. H. Ward, "Hierarchical grouping to optimize an objective function," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 236–244, Mar. 1963.
[15] C.-I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.
[16] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[17] C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines, 2001, [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm
[18] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.
[19] [Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/
[20] B. Mojaradi, H. Abrishami-Moghaddam, M. Zoej, and R. Duin, "Dimensionality reduction of hyperspectral data via spectral feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2091–2105, Jul. 2009.