

FEATURE FUSION OF HYPERSPECTRAL AND LIDAR DATA USING EXTINCTION PROFILES AND TOTAL VARIATION

Pedram Ghamisi^{1,2}, Behnood Rasti³, and Xiao X. Zhu^{1,2}

¹German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF)

²Technische Universität München (TUM), Signal Processing in Earth Observation

³Faculty of Electrical and Computer Engineering, University of Iceland

ABSTRACT

To improve the classification of hyperspectral images, this paper proposes an approach for multi-sensor data fusion of LiDAR and hyperspectral data using extinction profiles and Orthogonal Total Variation Component Analysis (OTVCA). Results on the benchmark Houston data indicate the superior performance of the proposed approach compared to other approaches used in the experiments based on classification accuracies.

Index Terms— Feature Fusion; Orthogonal Total Variation Component Analysis; Extinction Profiles; Random Forest; Support Vector Machines.

1. INTRODUCTION

At this point, an enormous amount of data captured by diverse sensors are available. Such data vary from spectral information taken by passive sensors [e.g., multispectral and hyperspectral images (HSI)], to height and shape information acquired by Light Detection And Ranging (LiDAR) sensors. This availability makes it possible to integrate rich multisensor information to further improve object detection ability and classification performance.

Recently, the fusion of hyperspectral images (HSI) and Light Detection And Ranging (LiDAR) has received lots of attention from researchers due to its benefits in terms of classification accuracy [1, 2].

Unfortunately, the automatic fusion of LiDAR-derived features and HSI is not trivial [1]. In addition, the simple stacking of extracted features obtained by different sensors might lead to *curse of dimensionality* [3]. To solve this shortcoming, different feature reduction approaches can be used [4]. This fact encourages researchers to develop an effective and efficient multi-sensor data fusion approach to perform both dimensionality reduction and feature fusion at once.

This research has been partly supported by Alexander von Humboldt Fellowship for postdoctoral researchers, Helmholtz Young Investigators Group “SiPEO” (VH-NG-1018, www.sipeco.bgu.tum.de) and FAPESP grants 2013/23514-0, 2015/12127-0.

High resolution remote sensing images contain considerable amount of spatial information, which could be useful for classification [4]. Recently, extinction profiles (EPs) [5], have shown superior performance in terms of spatial information extraction. In addition, the EPs are automatic in nature, which are independent from the kind of the attribute being used (e.g., area, volume, etc).

In this paper, a feature fusion approach for HSI and LiDAR is proposed, which is based on two main stages, spatial feature extraction and multi-sensor data fusion. In the first stage, EPs are used to extract spatial and elevation information from HSI and LiDAR, respectively. In the second step, the extracted features from HSI and LiDAR are fused using the orthogonal total variation component analysis (OTVCA) [6]. OTVCA estimates the fused features in a lower dimensional space while promotes piece-wise smoothness and maintains the spatial structures [6].

The rest of the paper is organized as follows: Section 2.1 describes the proposed OTVCA-based fusion technique. The experiments are described in Section 3. Finally, Section 4 concludes the paper.

2. METHODOLOGY

In this section, the proposed fusion techniques is described in detail in two parts, feature extraction using EPs (subsection 2.1) and fusion using OTCVA (subsection 2.2). Here, the number of bands and pixels in each band are denoted by p and n , respectively. Matrices are denoted by bold and capital letters, column vectors by bold letters, the element placed in the i th row and j th column of matrix \mathbf{X} by x_{ij} and the i th column by $\mathbf{x}_{(i)}$. Identity matrix of size $p \times p$ is denoted by \mathbf{I}_p . $\hat{\mathbf{X}}$ stands for the estimate of the variable \mathbf{X} , and \mathbf{X}^m denotes the estimate of the variable \mathbf{X} at m th iteration.

2.1. Extinction Profiles (EPs)

Ghamisi *et al.* [5] proposed EPs using a set of connected filters, extinction filters. Let $Max(\mathbf{X}) = \{M_1, M_2, \dots, M_N\}$ represent the regional maxima of the gray scale image \mathbf{X} . For

each M_i , we can estimate an extinction value ϵ_i with respect to the increasing attribute being analyzed. For the input gray scale image \mathbf{X} , the extinction filter preserves the n maxima with the highest extinction values [5].

EPs are constructed by applying several extinction filters, *i.e.*, a sequence of thinning and thickening transformations, with progressively higher threshold values to extract spatial and contextual information from the input data. Thinning and thickening are obtained from max-tree and min-tree, respectively [5]. The EP for the input gray scale image, \mathbf{X} , is obtained by (1):

$$\text{EP}(\mathbf{X}) = \underbrace{\{\phi^{P_{\lambda_L}}(\mathbf{X}), \phi^{P_{\lambda_{L-1}}}(\mathbf{X}), \dots, \phi^{P_{\lambda_1}}(\mathbf{X}), \mathbf{X}\}}_{\text{thickening profile}}, \underbrace{\{\gamma^{P_{\lambda_1}}(\mathbf{X}), \dots, \gamma^{P_{\lambda_{L-1}}}(\mathbf{X}), \gamma^{P_{\lambda_L}}(\mathbf{X})\}}_{\text{thinning profile}}, \quad (1)$$

where $P_\lambda : \{P_{\lambda_i}\} (i = 1, \dots, L)$ is a set of L ordered predicates (*i.e.*, $P_{\lambda_i} \subseteq P_{\lambda_k}, i \leq k$). For EPs, the number of extrema is considered to be the predicates. ϕ and γ are thickening and thinning transformations, respectively.

In order to generalize the concept of EPs from a gray scale image to HSI, one can extract a few informative features from the whole dimensionality using an approach such as independent component analysis (ICA). Then, the extracted features are considered to be the based images to construct EPs [7], known as extended extinction profiles (EEPs), which can mathematically be given by: $\text{EEP}(\mathbf{Q}) = \{\text{EP}(\mathbf{Q}_1), \text{EP}(\mathbf{Q}_2), \dots, \text{EP}(\mathbf{Q}_m)\}$.

Extended multi-EP (EMEP) concatenates different EEPs (*e.g.*, area, height, volume, diagonal of bounding box, and standard deviation on different extracted features) into a single stacked vector, which can given as follows: $\text{EMEP} = \{\text{EEP}_{a_1}, \text{EEP}_{a_2}, \dots, \text{EEP}_{a_w}\}$, where $a_k, k = \{1, \dots, w\}$ denotes different types of attributes [5, 7].

2.2. Feature Fusion Using Orthogonal Total Variation Component Analysis

Let \mathbf{HSI} be the input HSI, which contains detailed spectral information. \mathbf{EP}_{HSI} represents the spatial features produced by EPs on the first three independent components. $\mathbf{EP}_{\text{LiDAR}}$ is the elevation features produced by EPs on the LiDAR-derived digital surface model (DSM).

To fuse \mathbf{HSI} , \mathbf{EP}_{HSI} , and $\mathbf{EP}_{\text{LiDAR}}$, the number of features should be normalized to put the same weight on each type of the features and reduce the computational cost and noise throughout the feature space [8]. To do so, Kernel PCA [9] was used. The normalized dimension of \mathbf{HSI} , \mathbf{EP}_{HSI} , and $\mathbf{EP}_{\text{LiDAR}}$ is automatically set to the smallest dimension of the above-mentioned features.

Extracted features from HSI and LiDAR are highly redundant. In order to reduce the features redundancy, we propose

a low-rank model to fuse the normalized spectral, spatial, and elevation features. The low-rank model is suggested as

$$\mathbf{F} = \mathbf{A}\mathbf{V}^T + \mathbf{N}, \quad (2)$$

where $\mathbf{F} = [\mathbf{f}_{(i)}]$ is an $n \times p$ matrix containing the vectorized features at band i in its i th column, \mathbf{V} is an unknown subspace (low-rank) basis ($p \times r$), $\mathbf{A} = [\mathbf{a}_{(i)}]$ contains the i th vectorized unknown fused feature in its i th column, and $\mathbf{N} = [\mathbf{n}_{(i)}]$ is an $n \times p$ matrix containing the vectorized noise and error at band i in its i th column. Note that r is the number of fused features ($1 \leq r \leq p$), and $\mathbf{F} = [\mathbf{F}_{\text{HSI}}, \mathbf{H}, \mathbf{F}_{\text{LiDAR}}]$ contains hyperspectral bands and features extracted from both HSI and LiDAR rasterized data.

In (2), it is assumed that the fused features, \mathbf{F} , and the basis matrix, \mathbf{V} , are unknown. To preserve the spatial structure of the features and promote piece-wise smoothness on the fused features, here, we use OTVCA [6], which is based on solving the following TV penalized least squares problem

$$(\hat{\mathbf{A}}, \hat{\mathbf{V}}) = \arg \min_{\mathbf{A}, \mathbf{V}} J(\mathbf{A}, \mathbf{V}) = \arg \min_{\mathbf{A}, \mathbf{V}} \frac{1}{2} \|\mathbf{Y} - \mathbf{A}\mathbf{V}^T\|_F^2 + \lambda \sum_{i=1}^r \left\| \sqrt{(\mathbf{D}_h \mathbf{a}_{(i)})^2 + (\mathbf{D}_v \mathbf{a}_{(i)})^2} \right\|_1 \quad \text{s.t.} \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}_r, \quad (3)$$

where \mathbf{D}_h and \mathbf{D}_v are the matrix operators for calculating the first order vertical and horizontal differences, respectively, for a vectorized image. For an image of size $n_1 \times n_2$ we have $\mathbf{D}_h = \mathbf{R} \otimes \mathbf{I}_{n_1}$ and $\mathbf{D}_v = \mathbf{I}_{n_2} \otimes \mathbf{R}$, where \mathbf{R} is the first order difference matrix.

A cyclic descent (CD) algorithm given in [6] is used to solve (3) called OTVCA-CD. This algorithm solves the non-convex problem (3) w.r.t. one matrix at a time while the other matrix is assumed to be fixed. Therefore, OTVCA-CD iterates between the following two steps until a convergence criterion achieves:

2.2.1. A-step

When matrix \mathbf{V} is fixed, it can be shown that the minimization problem (3) can be considered as r separable TV denoising problems, which can be solved by using split Bregman iterations as $\mathbf{A}^{m+1} = \text{SplitBregman}(\mathbf{F}\mathbf{V}^m, \lambda)$.

2.2.2. V-step

When Matrix \mathbf{A} is fixed, it can be shown that the minimization problem (3) turns to an orthogonal (low-rank) Procrustes problem where the solution is given by a low-rank Procrustes rotation given by $\mathbf{V}^{m+1} = \mathbf{P}\mathbf{Q}^T$, where $\mathbf{A}^{m+1T} \mathbf{Y} = \mathbf{P}\Sigma\mathbf{Q}^T$.

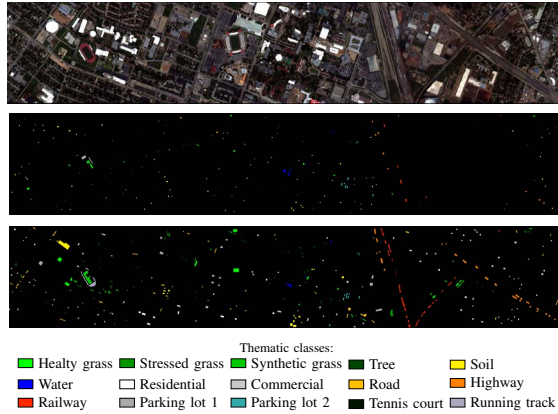


Fig. 1. Houston - From top to bottom: A color composite representation of the HSI; Training samples; Test samples; and legend of different classes.

3. EXPERIMENTAL RESULTS

3.1. Data Description

Houston Data: Both HSI and DSM are of the same spatial resolution (2.5 m). The size of the data is 349×1905 with the spatial resolution of 2.5m. The HSI consists of 144 spectral bands ranging $0.38-1.05\mu m$. Fig. 1 shows a color composite representation of the HSI and the corresponding training and test samples. Table 1 gives detailed information about the number of training and test samples of different classes.¹

3.2. Algorithm Setup

For the EPs, one only needs to define the number of desired levels (s) since the whole process is automatic. In this context, in order to generate the EP for area, volume, and diagonal of the bounding box, the values of n used to generate the profile are automatically given by $\lfloor 3^j \rfloor$, where $j = 0, 1, \dots, s-1$. The size of the EPs is $2s + 1$, since the original image should also be included in the profile. Here, s is set to seven, as suggested in [7].

OTVCA is initialized as suggested in [6]. The tuning parameter λ indicates the level of smoothness. In the experiments, λ is set to one percent of the intensity range of the features extracted.

In terms of the SVM, a radial basis function (RBF) kernel is used. The optimal hyperplane parameters are defined using 5-fold cross validation. For the RF, the number of trees is set to 300.

For the sake of simplicity, the following names are used in the experimental part: **LiDAR** and **HSI** show the classification accuracies of the LiDAR-derived DSM and HSI, respec-

¹The enhanced data is provided by Prof. Naoto Yokoya from Technical University of Munich (TUM).

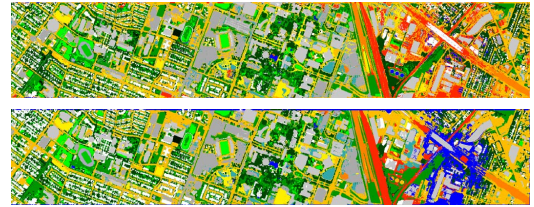


Fig. 2. Classification maps of the proposed method using (from top to bottom): SVM on the Houston data, RF on the Houston data, SVM on the Trento data, and RF on the Houston data

tively. EP_{LiDAR} , and EP_{HSI} show the classification accuracies of EPs applied to LiDAR, and HSI. $EP_{LiDAR+HSI}$ refers to the classification accuracies of EPs applied to the stack of LiDAR and HSI.

3.3. Classification Experiments

The classification results are shown in Table 1 and compared using class accuracies, OA, AA, and Kappa Coefficient (κ).

With reference to Table 1, the spatial information extracted by the EP significantly improves classification accuracies compared to the situations where the SVM and RF have directly been applied to the input datasets. For example, on the LiDAR (71 features) significantly improves the OA by almost 38% for the SVM classifier. For HSI, due to the rich spectral information, the consideration of the EPs can slightly improve the OA by almost 1% and 3%, using SVM and RF, respectively. $EP_{LiDAR+HSI}$ outperform the individual use of each data, which confirms that HSI and LiDAR provide complement information to differentiate different classes of interest.

The proposed approach provides the highest classification accuracy among all the approaches considered in this paper. The OTVCA-based fusion method clearly captures the redundant information existing in the HSI and the profiles and leads to the accuracy of over 90%. A similar trend can be seen in the case of using RF, where the OA obtained from the 50 fused features using OTVCA reaches over 92%, which is 5% more than the integrated profiles having 284 features.

The classification maps obtained by applying RF and SVM on the fused features using OTVCA are shown in Fig. 2. As can be seen, the proposed fusion technique provides classification maps having homogeneous regions while preserving the structures, which is greatly of interests specifically in the case of urban data sets.

4. CONCLUSIONS

In this paper, a technique has been proposed for the fusion of HSI and LiDAR. The proposed approach is based on EPs

Table 1. Houston: Classification accuracies obtained by different approaches using RF and SVM. The metrics AA and OA are reported in percentage. Kappa coefficient is of no units. The best result is shown in bold.

| Class name | Train./Test | LiDAR (1) | | HSI (144) | | EP _{LiDAR} (71) | | EP _{HSI} (213) | | EP _{LiDAR+HSI} (284) | | TVCA fusion (50) | |
|-----------------|-------------|-----------|--------|---------------|--------|--------------------------|---------------|-------------------------|---------------|-------------------------------|---------------|------------------|---------------|
| | | SVM | RF | SVM | RF | SVM | RF | SVM | RF | SVM | RF | SVM | RF |
| Grass Healthy | 198/1053 | 11.68 | 13.49 | 83.48 | 83.38 | 57.36 | 74.26 | 79.39 | 77.49 | 79.39 | 78.06 | 79.77 | 80.63 |
| Grass Stressed | 190/1064 | 0.00 | 16.26 | 96.43 | 98.40 | 40.79 | 61.75 | 78.85 | 78.48 | 80.36 | 84.96 | 97.84 | 99.62 |
| Grass Synthetis | 192/505 | 87.13 | 56.63 | 99.80 | 98.02 | 98.61 | 97.23 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| Tree | 188/1056 | 51.80 | 44.03 | 98.77 | 97.54 | 92.33 | 58.14 | 87.78 | 82.77 | 95.83 | 95.45 | 96.02 | 96.02 |
| Soil | 186/1056 | 12.12 | 58.05 | 98.11 | 96.40 | 83.43 | 82.10 | 99.81 | 97.73 | 99.81 | 98.77 | 98.67 | 99.43 |
| Water | 182/143 | 78.32 | 58.04 | 95.10 | 97.20 | 78.32 | 83.22 | 95.80 | 95.80 | 95.80 | 95.80 | 95.80 | 95.80 |
| Residential | 196/1072 | 56.90 | 39.09 | 89.09 | 82.09 | 55.22 | 77.33 | 85.17 | 73.23 | 80.41 | 73.41 | 88.90 | 86.01 |
| Commercial | 191/1053 | 13.11 | 29.53 | 45.87 | 40.65 | 29.06 | 68.28 | 65.15 | 59.92 | 90.41 | 85.28 | 87.65 | 93.54 |
| Road | 193/1059 | 14.92 | 13.60 | 82.53 | 69.78 | 67.33 | 59.40 | 89.90 | 83.00 | 89.80 | 93.96 | 87.35 | 97.07 |
| Highway | 191/1036 | 8.30 | 11.29 | 83.20 | 57.63 | 61.39 | 66.89 | 51.54 | 64.09 | 56.66 | 67.08 | 60.33 | 68.53 |
| Railway | 181/1054 | 72.68 | 40.42 | 83.87 | 76.09 | 99.72 | 99.91 | 87.76 | 84.72 | 90.70 | 90.89 | 99.34 | 98.86 |
| Parking Lot 1 | 192/1041 | 0.00 | 9.99 | 70.99 | 49.38 | 63.11 | 64.75 | 84.34 | 78.10 | 89.91 | 88.57 | 97.69 | 100.00 |
| Parking Lot 2 | 184/285 | 12.28 | 15.09 | 70.53 | 61.40 | 49.12 | 58.60 | 84.56 | 77.89 | 84.56 | 76.14 | 80.35 | 74.74 |
| Tennis Court | 181/247 | 97.57 | 80.16 | 100.00 | 99.60 | 100.00 | 100.00 | 100.00 | 99.60 | 100.00 | 100.00 | 100.00 | 100.00 |
| Running Track | 187/473 | 27.91 | 75.90 | 97.46 | 97.67 | 74.21 | 87.74 | 97.25 | 99.37 | 98.10 | 99.79 | 100.00 | 100.00 |
| OA | | 28.82 | 31.83 | 84.69 | 77.47 | 67.2 | 73.42 | 85.82 | 80.36 | 86.87 | 86.98 | 90.33 | 92.45 |
| AA | | 36.31 | 37.43 | 86.34 | 80.34 | 70.00 | 75.97 | 83.08 | 83.47 | 88.78 | 88.54 | 91.31 | 92.68 |
| κ | | 0.2422 | 0.2677 | 0.8340 | 0.7563 | 0.6440 | 0.7120 | 0.8168 | 0.7876 | 0.8577 | 0.8592 | 0.8950 | 0.9181 |

and OTVCA. EPs extract spatial and elevation information from HSI and LiDAR, respectively. Then, OTVCA, which is a non-convex optimization problem, has been utilized to estimate the unknown (low-dimensional) fused features. It has been shown that the low-dimensional fused features obtained by the EPs and OTVCA improve the classification accuracies compared to the integrated features for HSI and LiDAR. The experimental results confirm that the EPs can effectively extract spatial and elevation information from HSI and LiDAR, respectively. In addition, the OTVCA-based fusion technique captures the redundancy of the features while improves the classification accuracies. This has been shown based on the classification accuracies obtained by using both RF and SVM classifiers for both rural and urban data sets. Additionally, the experiments showed that the OTVCA fusion technique provides classification maps having homogeneous regions while preserving the structures which is due to the exploitation of the total variation.

5. REFERENCES

- [1] Pedram Ghamisi, Jon Atli Benediktsson, and Stuart Phinn, "Land-cover classification using both hyperspectral and lidar data," *Int. Jour. Image Data Fus.*, vol. 6, no. 3, pp. 189–215, 2015.
- [2] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas," *IEEE Trans. Geos. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, 2008.
- [3] G.F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT, no. 14, pp. 55 – 63, 1968.
- [4] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*, Artech House Publishers, INC, Boston, USA, 2015.
- [5] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geos. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, 2016.
- [6] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 6976–6985, Dec 2016.
- [7] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geos. Remote Sens. Let.*, vol. 13, no. 11, pp. 1641–1645, 2016.
- [8] W. Liao, R. Bellens, S. Gautama, and W. Philips, "Feature fusion of hyperspectral and lidar data for classification of remote sensing data from urban area," in *EARSeL Special Interest Group on Land Use and Land Cover, 5th Workshop*, Sebastian Van Der Linden, Tobias Kuemmerle, and Katja Janson, Eds., 2014, pp. 34–34.
- [9] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comp.*, vol. 10, no. 5, pp. 1299–1319, 1998.