A TWO-STREAM FUSION FRAMEWORK FOR LIDAR AND HYPERSPECTRAL IMAGERY

Christian Debes†, Andreas Merentitis†, Roel Heremans†, Jürgen Hahn‡, Nikolaos Frangiadakis† and Tim van Kasteren†

†AGT International
64295 Darmstadt, Germany
E-mail: {cdebes, amerentitis, rheremans, nfrangiadakis, tkasteren}@agtinternational.com

‡Signal Processing Group, Institute of Telecommunications
Technische Universität Darmstadt
64283 Darmstadt, Germany
E-mail: jhahn@spg.tu-darmstadt.de

ABSTRACT

Data fusion can significantly increase accuracy of automated classification in remote sensing applications by combining data from different types of sensors. Particularly for hyperspectral imagery (HSI), complementing the hyperspectral data with topographical information in the form of a Digital Surface Model (DSM) generated by LiDAR data is promising to address problems with artifacts or distortion in difficult areas. In this paper we introduce a novel framework for combining the HSI and LiDAR data, which enables handling identified objects as uniform entities rather than as independent pixels. Further contributions include an initial spectral unmixing step that segregates noise and significantly improves the benefit of adding LiDAR, as well as the application of ensemble learning in the form of Random Forest algorithms that inherently support feature selection.

Index Terms—Hyperspectral imaging, LiDAR, sensor fusion, classification algorithms, remote sensing

1. INTRODUCTION

Since 2006, the annually organized IEEE GRSS Data Fusion Contest deals with important data fusion problems in the field of remote sensing, e.g. [1]. The 2013 contest tackles the fusion of HSI and LiDAR data. HSI is a well known tool for classification of vast areas, however, there are several drawbacks associated with it. First, the pixels usually contain a mixture of materials making reliable classification difficult. Second, as HSI is a passive system, the images are often affected by noise or artifacts, because the scene may not be homogeneously illuminated, altering the spectral responses of the material. Data obtained from LiDAR can alleviate some of these problems as it is an active acquisition system that provides information about the geometrical structure of the scene. The fusion of LiDAR and hyperspectral data is, however, a rather new field of research. In [2] and [3], classification of complex tree areas is explored. Both classification methods fuse the information of the sources in the feature space of the classifier and by doing so achieve higher performances compared to only using hyperspectral data. In the context of urban environments, Kanaev et al. [4] investigate the detection of man-made objects. A hypothesis test based on the geometry of the objects in the scene is formulated to identify potential man-made objects that are finally classified based on hyperspectral information.

In this work, we present a novel classification framework for the fusion of hyperspectral and LiDAR data. In our framework (Fig. 1), an unsupervised object detection and a supervised classification stream are run in parallel. The outcome of the unsupervised stream, as explained in Section 2, is a list of objects assigned to predefined categories. The supervised stream results in a pixel-wise classification map detailed in Section 3. Section 4 explains then how the outcomes of both streams are merged to obtain an accurate classification map. In Section 5, we show results of the proposed method before a conclusion is drawn.

2. UNSUPERVISED OBJECT DETECTION

The aim of the ‘Unsupervised Object Detection’ block in Fig. 1 is to fuse hyperspectral and LiDAR data in order to extract objects such as buildings and streets in an unsupervised fashion. The individual steps of the ‘Unsupervised Object Detection’ module are elaborated in the sequel.
2.1. Extraction and segmentation

Given the aligned hyperspectral image and the LiDAR-derived elevation map, the first step is to extract characteristic maps representing the spatial distribution of a feature that is of relevance for object extraction. Examples of relevant characteristics include vegetation index, elevation and flatness. These characteristics allow us to extract objects as contiguous pixels that have similar properties. For example, the pixels representing a commercial building have low vegetation, high elevation and high flatness.

Each characteristic map \( y^g \) is segmented in order to obtain a binary map \( x^g, g = 0, ... , G-1 \) where a binary ‘1’ denotes the presence of the respective characteristic (e.g. ‘vegetation’) and a binary ‘0’ denotes its absence (e.g. ‘no vegetation’). Neighbourhood relations are taken into account by using a Markov Random Field (MRF)-based approach, namely the ICM algorithm [5] which iteratively maximizes the posterior as,

\[
\hat{x}^g = \arg\max_{x^g} \{p(x^g)p(y^g | x^g)\}. \tag{1}
\]

2.2. Combinatorial fusion

Given the segmented binary maps \( x^g, g = 0, ... , G-1 \), we perform combinatorial fusion by multiplying all combinations of the binary maps \( x^g \) or their respective counterparts \( 1-x^g \). This yields a total of \( 2^G \) binary maps which contain all combinations of the considered characteristic maps. Consider the example of using two characteristic maps \( G = 2 \): a vegetation index derived from the hyperspectral image and an elevation map derived from LiDAR data. The respective segmentation maps \( x^0 \) and \( x^1 \) can be fused in four different ways as shown in the following table:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Characteristic</th>
<th>Example cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x^0 \cdot x^1 )</td>
<td>High veg, high elev</td>
<td>Trees</td>
</tr>
<tr>
<td>( x^0 \cdot (1-x^1) )</td>
<td>High veg, low elev</td>
<td>Grass</td>
</tr>
<tr>
<td>( (1-x^0) \cdot x^1 )</td>
<td>Low veg, high elev</td>
<td>Buildings</td>
</tr>
<tr>
<td>( (1-x^0) \cdot (1-x^1) )</td>
<td>Low veg, low elev</td>
<td>Streets</td>
</tr>
</tbody>
</table>

Table 1. The four combinatorial fusion maps based on the combinations of vegetation and elevation maps

2.3. Object extraction and selection

Note that the fused binary maps obtained as described in the previous section contain contiguous pixels sets that have similar characteristics. These sets are considered as candidate objects in the ‘Object extraction and selection’ step. A selection step is performed, in which knowledge of specific object classes is embedded, in order to remove outliers. In the sequel, we detail how this is done for three exemplary clusters: commercial buildings, parking lots and streets.

For commercial buildings, the combinatorial fusion of ‘low vegetation’ and ‘high elevation’ is considered. From the potentially high number of candidate objects that are obtained by the binary fused segmentation map a subset is selected. This selection is based on the object area and solidity (the ratio of the pixel area and the convex hull around the object) as commercial buildings typically have a large area occupancy and a high solidity. Fig. 2 gives an example of selecting objects from a binary fused segmentation map, according to these criteria.

Based on the combinatorial fusion of ‘low vegetation’ and ‘low elevation’, objects with a certain area, width and squareness are considered as parking lots (Fig. 3 center). A skeletonized image is derived from the complement of the combinatorial map and the parking map, on which a Hough transform method is applied, to extract the road segments. The Hough-lines are laterally grown until they cover the majority of the street pixels from the combinatorial map (Fig. 3 right).

Note that the output of the ‘Unsupervised Object Detection’ step is not a final classification but rather an initial clustering of candidate objects that will support the actual supervised classification in the following.

3. SUPERVISED CLASSIFICATION

‘Supervised Classification’ constitutes the right stream of the proposed framework (Fig. 1) and is comprised of two parts: feature extraction and classification based on ensemble learning. The outcome of the process is a classification map where a class is assigned to each pixel independently of its neighborhood.

3.1. Hyperspectral and LiDAR feature extraction

Our feature extraction algorithm initially applies blind unmixing based on Automatic Target Generation Procedure [6] to
extract the basic spectra of the materials that are present in the HSI in a way that is resilient to distortions (e.g. from the presence of shadows). In total, fifty basic spectra are automatically extracted and the respective abundance maps are used as features for the pixel-based classification stream. Information from the original spectral bands is also used such that noise impact is minimized through Minimum Noise Fraction (MNF) transformation. Moreover, composite features of the hyperspectral data like vegetation index and water vapor absorption are considered since they are particularly good in discriminating vegetation and water bodies respectively. Finally, on the geographical topology side we use the raw LiDAR-derived elevation map as a feature, as well as composite topology features such as gradients.

It is important to underline at this point that the abundance maps, when combined with the LiDAR information, are very effective features. Fig. 5 (bottom) shows the effect of adding LiDAR on the extracted fifty abundance maps, compared to the effect of adding it on the same number of PCA components (top). It can be observed that with the abundance map-based approach more fine-grained structures and more classes are visible. The plausible explanation is that abundance maps are already noise-segregated and adding LiDAR in this feature set provides exactly the type of topology-related information required to discriminate between some marginal cases (e.g. different types of buildings or vegetation). This is an important observation, given that in literature the use of abundance maps as features is mostly proposed as a dimensionality reduction method (e.g., [7]) and not as a way to enhance classification accuracy (apart from the improvement achieved by mitigating the Hughes phenomenon).

### 3.2. Ensemble learning for classification

When multiple models for classification are available, cross-validation is often deployed in order to select the best performing model. The crucial prior belief in this approach is that it is possible to assess the predictive power of different models simply by comparing them on data that was not used to create these models. Ensemble methods bring this prior belief one step further and instead of simply trying to find the best model, they combine them so as to achieve higher predictive performance than that of any model alone. As a supervised learning technique, an ensemble represents a single hypothesis, but one that is not necessarily contained within the space of the models used in the ensemble. One of the ensemble methods that have gained significant attention in recent years is the Random Forest algorithm, which is based on multiple Classification Tree instances [8]. We propose detecting the presence of shadows in a preprocessing step and then creating two separate Random Forest ensembles - one for the shadow-covered area and one for the shadow-free area.

In addition to attractive theoretical properties [8], the Random Forest algorithm offers also the advantage that it can be used to perform feature selection. This is particularly useful when the training set is very small (in this challenge about 0.4% of the total number of pixels). The reason is that as the volume of the feature space increases the data becomes increasingly sparse in the space it occupies and this scarcity makes it difficult to achieve statistical significance for many learning methods, an aspect of the curse of dimensionality. The random forest Out of Bag error was used to automatically evaluate features according to their impact, resulting in 45 features selected for the shadow-free and 55 for the shadow-covered part. The outcome of the two ensembles and their uncertainties are then synthesized based on the shadow index to create the classification outcome.

### 4. OBJECT-BASED CLASSIFICATION

The unsupervised object detection from Section 2 and the supervised pixel-wise classification result from Section 3 are combined yielding object-based classification. Practically, this means that pixels are processed collectively, rather than individually. Let \( f_i \) denote the \( i \)th pixel of the observed scene with label \( \ell_i \) which can take values \( c \in C \) where \( C \) is the set of labels. Note that the outcome of the pixel-based classification in our case is a soft decision, meaning that the class probability distribution \( p(\ell_i = c) \) is obtained for all \( i \). With \( O_k \) denoting the set of pixels in the \( k \)th object extracted from the unsupervised object extraction step in Section 2 we perform a reassignment of the classification for all objects as,

\[
\ell_i^k = \underset{c}{\arg \max} \sum_{f_j \in O_k} p(\ell_j = c) \quad \forall \{i | f_i \in O_k\} \quad (2)
\]
meaning that for every object we search for the label maximizing the average class certainty. To avoid single-labeling of objects consisting of more than one class, Equation (2) can be conditioned on \( \frac{1}{N_k} \sum_{f_i \in O_k} p(\ell_j = c) > p_0 \) where \( N_k \) is the number of pixels in the \( k^{th} \) object and \( (1 - p_0) \) is the average error one is willing to accept. Finally, for pixels not belonging to an extracted object, a MRF approach similar to Section 2 is applied as,

\[
\ell_{MRF} = \arg \max_\ell \{ p(\ell)p(f|\ell) \}
\]

where \( f \) and \( \ell \) denote the vectorized versions of \( f_i \) and \( f_j \), respectively. The conditional probability \( p(f|\ell) \) can be obtained from the class probability distribution \( p(\ell_i = c) \). Equation (3) can be maximized iteratively using e.g. the ICM algorithm.

5. EXPERIMENTAL RESULTS

The introduced framework for the fusion of hyperspectral and LiDAR data provides improvements over traditional automated methods in several folds. This includes the feature extraction part (combining abundance maps, regularized MNF of the HSI data, as well as composite HSI and LiDAR features), the supervised pixel-based classification (use of ensemble learning in the form of Random Forests that inherently support feature selection) and finally treating identified objects as higher level entities. Since the feature extraction part was briefly discussed in Section 3, this Section focuses only on the improvement provided by the Random Forest algorithm compared to using an SVM and elaborates on the additional benefit of object-based classification.

Pixel-based classification of an SVM classifier and of the Random Forests is shown in Fig. 6 (top); in the latter approach more fine-grained structures are visible in the shadowed area and objects appear smoother. The additional improvement by using object-based classification is depicted in Fig. 6 (bottom). The benefit in this case is that both the highway and railway do not have missing parts and large buildings are classified as one category (i.e. no residential pixels inside commercial buildings).

6. CONCLUSION

Augmenting HSI data through data fusion with topographical information in the form of a DSM is effective for dealing with distortion in difficult areas of the map. In this paper we introduced a novel framework comprised of two parallel streams, ‘Unsupervised Object Detection’ and ‘Supervised Classification’. This framework enables handling of pixels collectively as objects. Further contributions include a spectral unmixing step that segregates noise and significantly improves the benefit of DSM, as well as application of ensemble learning in the form of Random Forest algorithms that inherently support feature selection.

7. REFERENCES