Article

Hyperspectral and LiDAR Data Fusion Classification Using Superpixel Segmentation-Based Local Pixel Neighborhood Preserving Embedding

Yunsong Li 1,2,*, Chiru Ge 1,2, Weiwei Sun 3, Jiqtang Peng 4, Qian Du 5 and Keyan Wang 1,2

1 The State Key Laboratory of Integrated Services Networks, Xidian University, Xi’an 710071, China; gchirumsu@gmail.com (C.G.); kywang@mail.xidian.edu.cn (K.W.)
2 The School of Telecommunications Engineering, Xidian University, Xi’an 710071, China
3 The Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China; nbsww@outlook.com
4 The Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University, Wuhan 430062, China; pengjt1982@hubu.edu.cn
5 The Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762, USA; du@ece.msstate.edu

* Correspondence: ysli@mail.xidian.edu.cn; Tel.: +86-158-2926-2265

Received: 10 January 2019; Accepted: 1 March 2019; Published: 6 March 2019

Abstract: A new method of superpixel segmentation-based local pixel neighborhood preserving embedding (SSLPNPE) is proposed for the fusion of hyperspectral and light detection and ranging (LiDAR) data based on the extinction profiles (EPs), superpixel segmentation and local pixel neighborhood preserving embedding (LPNPE). A new workflow is proposed to calibrate the Goddard’s LiDAR, hyperspectral and thermal (G-LiHT) data, which allows our method to be applied to actual data. Specifically, EP features are extracted from both sources. Then, the derived features of each source are fused by the SSLPNPE. Using the labeled samples, the final label assignment is produced by a classifier. For the open standard experimental data and the actual data, experimental results prove that the proposed method is fast and effective in hyperspectral and LiDAR data fusion.

Keywords: hyperspectral; light detection and ranging (LiDAR); image fusion; extinction profile (EP); superpixel segmentation; local pixel neighborhood preserving embedding (LPNPE); Goddard’s LiDAR; hyperspectral and thermal (G-LiHT) data

1. Introduction

With the development of sensor technologies, various information about different materials on the ground can be collected from various image sensor sources, e.g., hyperspectral images (HSI) with spectral characteristics, light detection and ranging (LiDAR) images with shape and elevation information, thermal images with temperature information and synthetic aperture radar (SAR) images with terrain texture information. These multi-source images make it possible to make full use of the complementary and redundant information of each source which can provide reliable image interpretation and improve the performance of intelligent processing algorithms. Although these datasets provide a wealth of information, the automatic interpretation of remote sensing data is still very challenging [1].

Hyperspectral images (HSI) have hundreds of narrow spectral bands throughout the visible and infrared portions of the electromagnetic spectrum. LiDAR is an active remote sensing method, using the pulsed laser to measure distances (ranges). Similar to Radar, the coherent light pulses are transmitted, reflected by a target, and detected by a receiver. LiDAR data products usually include...
LiDAR point cloud, digital terrain model (DTM), canopy height model (CHM) and digital surface model (DSM). LiDAR point cloud contains 3D coordinates, classified ground returns, above ground level (AGL) heights and apparent reflectance. The CHM represents the actual height of buildings, trees, etc. without the influence of ground elevation. The DTM is the bare-earth elevation. The DSM describes the height of all the objects on the surface of the earth, which is the sum of the DTM and the CHM. The passive sensing of hyperspectral systems can describe the spectral characteristics of the observed scenes, whereas the active sensing of LiDAR systems can offer the height and shape information of the scenes. LiDAR also has high accuracy and flexibility, since it can be operated at any time, less sensitive to weather conditions and system parameters are adjustable, for example, flying speed/height, scan angle, pulse rate, and scan rate amongst others.

Currently, various algorithms are only designed for HSI or LiDAR. Several algorithms for classification, feature extraction, and segmentation are proposed for HSI [2–10], while many feature extraction and detection algorithms are only designed for LiDAR [11–16]. However, it is evident that no single type of sensors can always be adequate for reliable image interpretation. For instance, HSI should not be utilized to distinguish objects composed of the same material, such as roofs and roads with the same pavement material. On the other hand, LiDAR data alone cannot be adopted to differentiate objects with the same elevation, such as roofs with the same height, but made of concrete or tile [1].

In many applications, HSI and LiDAR have been used in combination successfully, such as biomass estimation [17], micro-climate modeling [18], mapping plant richness [19] and fuel type mapping [20]. Furthermore, the combined use of HSI and LiDAR results in higher classification accuracies than using each source separately. For example, elevation and shape information in LiDAR data and spectral information acquired by HSI have been investigated [1,21–28]. The above work show that LiDAR and HSI can complement each other well, and by adequately integrating the two data sets, the advantages of both can be fully utilized while the shortcomings of each data set can be solved as well. In the above work, the results have shown that the combined use of LiDAR and optical data can improve classification accuracies in forests and urban areas. The research work sequence on the combined use of LiDAR and HSI led to the 2013 and 2018 data fusion contest organized by the Geoscience and Remote Sensing Society (GRSS).

Hyperspectral and LiDAR fusion classification methods can be feature level fusion or the decision level fusion. In the feature level fusion, features extracted in the HSI and LiDAR are stacked for further processing [21]. However, the stacked features may contain redundant information. With the limited number of labeled samples in many real applications, the stacked features may pose the problem of the curse of dimensionality and therefore result in the risk of overfitting the training data. Therefore, many works in the literature adopt the fusion framework of stacking and dimension reduction [22,23,26–28]. W. Liao et al. Y. Gu et al. and P. Ghamisi et al. use the graph-based method, where calculating the graph needs a lot of calculations and memory [22,23,28]. B. Rasti et al. use a sparse representation based method [26], and B. Rasti et al. use a total variation component analysis based method, in which solving the optimization equation by loops requires much calculation [27]. In the decision level fusion, the classification results of different classifiers are merged with the majority voting strategy [24]. However, voting can lead to rough results. In this paper, we propose an effective fusion method with low computational complexity which is suitable for real-time applications.

In our fusion framework, we need to extract the spatial features of HSI and LiDAR images. M. Dalla Mura et al. introduced the concept of attribute profile (AP) [29] as a generalization of morphological profile [30]. The AP extracts multi-level representations of an image, which utilizes a sequential application of morphological attribute filters. To further improve the conceptual capability of the AP and the corresponding classification accuracies, the extinction profiles (EPs) was proposed [31,32]. The EP has the following advantages. Firstly, it can remove insignificant details and preserve the geometrical characteristics of the input image simultaneously. Secondly, it can deliver a better recognition performance than the traditional spatial-spectral AP feature extraction method.
For fusion, the spatial features of HSI and LiDAR images as well as spectral features of HSI are stacked first. The stacked features have complete information. Then, to avoid the high dimensionality and the risk of overfitting the training data, we need to find a fast and efficient method to remove redundant information with the HSI and LiDAR data. For the traditional dimension reduction (DR) methods, for example, linear discriminant analysis (LDA) [33,34], local Fisher discriminant analysis (LFDA) [35], local discriminant embedding (LDE) [36], nonparametric weighted feature extraction (NWFE) [37] and so on, they belong to the spectral-based DR methods. However, for the stacked features, the vector-domain similarity is not sufficient to show the intrinsic relationships among different samples. If two samples have a small vector distance, they may have a large spatial pixel distance. The projection based on the vector similarity metric may result in wrong features. A. Plaza et al. have proved that spatial contexture information is useful to increase the classification accuracy of HSI [38]. Therefore, a spatial similarity-based method, a local pixel neighborhood preserving embedding (LPNPE) is utilized [39]. The LPNPE learns the discriminant projections through the local spatial pixel neighborhood. However, it may also have a problem that two samples in the same spatial pixel neighborhood may have a sizeable vector distance. To solve the above problem, an entropy rate superpixel (ERS) segmentation method is utilized [40]. A superpixel created by the ERS contains the pixels with small vector distance. Therefore, we intersect the superpixel and the defined local spatial pixel neighborhood to remove the pixels with large vector distances and get an ideal spatial neighborhood. Finally, the discriminant projections can be learned by the ideal spatial neighborhoods.

Recently, some new clustering algorithms have been proposed [41,42]. These clustering algorithms provide the possibility to design better fusion algorithms and will be discussed in future work.

For the application, our proposed fusion framework can be successfully applied to G-LiHT data. The G-LiHT airborne imager created by the US National Aeronautics and Space Administration (NASA) is an airborne system that simultaneously collects LiDAR data, HSI, and thermal data [43]. The G-LiHT data contains CHM, DTM, Lidar Point Cloud, a hyperspectral reflectance image in the same area at 1m spatial resolution. The G-LiHT data provides new opportunities for the design of new hyperspectral and LiDAR fusion classification algorithms. C. Zhang et al. have evaluated the G-LiHT data for mapping urban land-cover types [24]. However, C. Zhang et al. take merely advantage of LiDAR’s CHM data and HSI, and it does not fully exploit the potential of the G-LiHT data [24].

In this paper, an innovative method, superpixel segmentation based local pixel neighborhood preserving embedding (SSLPNPE) method, is proposed for the fusion of HSI and LiDAR data. In particular, the main contributions of this paper are as follows.

(1) This paper presents a novel fusion method SSLPNPE. Our proposed method has low computational complexity and can significantly improve the classification accuracy of the LiDAR and HSI fusion.

(2) A new workflow is proposed to calibrate the G-LiHT data. With the workflow, our proposed method can be applied for practical applications. Experimental results prove that for the G-LiHT data, the proposed method SSLPNPE is fast and effective in hyperspectral and LiDAR data fusion. The proposed workflow can be generalized to any fusion classification method and the G-LiHT data in any scene.

(3) This paper presents that processing the CHM and the DTM separately can achieve higher classification accuracy than the DSM. To the best of our knowledge, we utilize the CHM and the DTM separately instead of the DSM for the first time in the remote sensing community.

The structure of the paper is as follows. Section 2 introduces the methodology. Section 3 presents the data, experiment setup and experimental results. Section 4 provides the main concluding remarks.

2. Methodology

Figure 1 shows the framework of the proposed method. In this framework, EPs were applied to the hyperspectral and LiDAR images to extract spatial and elevation information. Firstly, the
hyperspectral spectral features, the hyperspectral spatial features, and the LiDAR elevation features are fed to kernel principal component analysis (KPCA) to reduce the noise [44]. Then, because different feature sources typically have a different range of values and different characteristics, all the features are normalized before the image fusion. Next, the proposed fusion method SSLPNPE was adopted to create the fused features, and the specific algorithm is shown in Algorithm 1. Finally, the fused features from SSLPNPE were classified by the support vector machine (SVM) classifier to get the classification map [45].

Figure 1. The framework of the proposed method applies to Goddard’s light detection and ranging (LiDAR), hyperspectral and thermal (G-LiHT) real data. In the fusion strategy, the superpixels intersect with spatial neighbors to delete the samples with large spectral distance.
The EP for the input grayscale image $I$ can be defined as \[31\]:

$$E_{\text{RS}}(x_i) = \{x \in (p,q) \mid p \in [p_i - a, p_i + a], q \in [q_i - a, q_i + a], a = (w-1)/2;$$

Algorithm 1 Feature fusion with superpixel segmentation-based local pixel neighborhood preserving embedding (SSLPNEP).

**Input:** HSI: HSI spectral features; EMEP: HSI spatial features; MEP_DTM: LiDAR DTM spatial features; MEP_CHM: LiDAR CHM spatial features;

**Output:** SSLPNEP features, $F$;

1. Stack the features, $F_{\text{sta}} = \text{HSI} + \text{EMEP} + \text{MEP_DTM} + \text{MEP_CHM}$;
2. Randomly select $p$ points in $F_{\text{sta}}$. $x_i$ means one of the $p$ points. Determine the $w \times w$ local pixel neighborhood $N(x_i) = \{x \in (p,q) \mid p \in [p_i - a, p_i + a], q \in [q_i - a, q_i + a], a = (w-1)/2;$$
3. Use the ERS on first PC of HSI to get the superpixel labels. According to the superpixel labels, determine the superpixel segmentation based local pixel neighborhood $S(x_i)$;
4. Obtain the final spatial local neighborhood through $F_i = N(x_i) \cap S(x_i) = \{x_{i1}, x_{i2}, \ldots x_{im}\}$;
5. Get the local pixel neighborhood preserving scatter matrix by $H = \sum_{i=1}^{m} \sum_{g=1}^{q} (x_i - x_{ig})(x_i - x_{ig})^T$;
6. Get the total scatter matrix by $S = \sum_{i=1}^{m} (x_i - m)(x_i - m)^T$;
7. Acquire the optimal projection matrix $V = [v_1, v_2, \ldots v_i]$ by solving $Sv = \lambda Hv$;
8. Get the SSLPNEP features, $F = F_{\text{sta}} \times V$;

$$\text{return } F;$$

2.1. The Extinction Profile

The spatial features of HSI include shape and texture, while the spatial features of LiDAR images include shape, texture and elevation information. These spatial features are essential for classification tasks. Therefore, in our fusion framework, we need a spatial feature extraction method to obtain the spatial features of HSI and LiDAR images.

The extinction profile (EP) \[31\] has recently drawn much attention since it has the following advantages. Firstly, it can remove trivial details and preserve the geometrical characteristics of the input image simultaneously. Secondly, it can deliver a better recognition performance than the traditional spatial-spectral AP \[29\]. Therefore, we utilize the EP for spatial feature extraction in this paper.

The EP consists of a sequence of thinning and thickening transforms obtained by a set of extinction filters (EFs) which are connected filters. The EFs preserve the relative image extrema. For instance, The EP for the input grayscale image $I$ can be defined as \[31\]:

$$\text{EP}(I) = \left\{ \varphi^{p_1}(I), \varphi^{p_2}(I), \ldots \varphi^{p_S}(I) \right\}. \quad (1)$$

The terms $\varphi$ and $\gamma$ are thickening and thinning transforms, and $p_i : \{p_i\} (i = 1, \ldots, S)$, a set of $S$ ordered predicates (i.e., $p_i \subseteq p_{i+1}, i \leq k$). Predicates determine the number of extrema in EPs.

In order to extract the EP from HSI, independent component analysis (ICA) was first utilized to obtain the most informative components, and the three independent components (ICs) are utilized as base images to produce the extended EP (EEM) \[31\]:

$$E_{\text{EM}(I)} = \{E_{\text{IC}1}, E_{\text{IC}2}, E_{\text{IC}3} \}. \quad (2)$$

Since the ICs are relatively independent to each other \[46\], the obtained ICs can characterize the original HSI in different aspects which can offer distinctive and complementary information in the corresponding EP. Compared to morphological profiles (MPs) \[30\] which can only represent the structure and size of observed objects, EPs are more flexible. By these means, the extended multivariate EP (EMEP) joins different EEPs (e.g., height, area, diagonal of bounding box, volume and standard deviation)
into a single stacked vector. The EMEP can be defined as: \( \text{EMEP}(i) = \{ EP_{a_1}, EP_{a_2}, \ldots, EP_{a_\omega} \} \) where \( a_j, j = \{ 1, \ldots, \omega \} \) denotes different types of attributes \([31]\). Considering different extinction attributes can obtain complementary spatial information, the EMEP can better describe spatial information than a single EP.

To extract the EP from the LiDAR CHM and LiDAR DTM, we accepted the term multivariate EP (MEP) for the situation when different types of EPs are used to the LiDAR images because there is only single band available.

2.2. Fusion

In our proposed method, the features in the HSI, the HSI EMEP and the LiDAR MEP were first stacked. However, the stacked features may contain redundant information. To avoid an over-fitting risk in our fusion framework, we needed to reduce the dimension of the stacked features.

Since the adjacent pixels in the spatially homogeneous region were composed of the same material and belong to the same class, we utilized a spatial domain dimension reduction method LPNPE. To solve the problem that two samples in the same spatial pixel neighborhood may have a large vector distance, an entropy rate superpixel (ERS) segmentation method is utilized \([40]\) to remove the pixels with large vector distances and get an ideal spatial neighborhood. The proposed method SSLPNPE can remove redundant information and effectively retain spatial contexture information which is useful to improve the classification accuracy \([38]\).

To segment the HSI efficiently, we first utilize the principal component analysis \([47]\) to obtain the first principal component (PC). Since the first PC contains the most spatial information of HSI, ERS segmentation \([40]\) is then used to create superpixel segmentation map on the first PC. In particular, the first PC, and the edge set \( E \) describes the pairwise similarity between neighboring pixels. Next, by selecting a subset \( A \) from \( E \), the graph \( G \) is clustered into \( K \) connected subgraphs which correspond to superpixels. To get the compact, balanced and homogeneous superpixels, the objective function for the superpixel segmentation is shown as:

\[
\max_A \left( \mathcal{H}(A) + \mu B(A) \right) \text{ s.t. } A \subseteq E, \tag{2}
\]

where \( \mathcal{H}(A) \) is the entropy rate constraint, \( B(A) \) is the balance constraint and \( \mu \geq 0 \) is the weight which controls the contribution of the balance term and the entropy rate term. M. Liu et al. solved this problem by the greedy algorithm \([40]\). Finally, the superpixel patch can be extracted from the segmentation map. The superpixel \( S(x_i) \) represents the superpixel where \( x_i \) is located. It contains the local pixels subset \( Q_i = \{ x_i, x_{i1}, x_{i2}, \ldots, x_{ik} \} \), where \( k \) represents the number of pixels other than \( x_i \) in \( S(x_i) \).

Suppose the local pixel neighborhood center is \( x_i \) with coordinate \((p_i, q_i)\), the spatial neighborhood can be expressed as

\[
N(x_i) = \{ x(p, q) \mid p \in [p_i - a, p_i + a], q \in [q_i - a, q_i + a] \}, \tag{3}
\]

where \( a = (w-1)/2 \), and \( w \) is the width of the neighborhood window. In the spatial neighborhood \( N(x_i) \), the local pixels subset can be marked as \( P_i = \{ x_i, x_{i1}, x_{i2}, \ldots, x_{is} \} \), where \( s = w^2 - 1 \) denotes the number of neighbors of \( x_i \). In order to optimize the choice of local neighborhoods, the final selected spatial local neighborhood is as

\[
F_i = P_i \cap Q_i = \{ x_i, x_{i1}, x_{i2}, \ldots, x_{im} \}, \tag{4}
\]

where \( m \) is the number of pixels other than \( x_i \). As Figure 1 shows, \( F_i \) removes the point which is very different from other pixels in \( N(x_i) \). The distance scatter in the spatial local neighborhood \( F_i \) denotes \( h_i = \frac{m}{s} \sum_{g=1}^{m} (x_i - x_{ig})(x_i - x_{ig})^T \) where \( x_i \) is the central pixel. Select \( p \) scattered points in the stacked feature space, the local pixel neighborhood preserving scatter matrix defines as
$$H = \sum_{i=1}^{p} h_{i} = \sum_{i=1}^{p} \sum_{g=1}^{m} (x_{i} - x_{ig})(x_{i} - x_{ig})^{T}. \quad (5)$$

The total scatter matrix can be defined as

$$S = \sum_{i=1}^{p} (x_{i} - \overline{m})(x_{i} - \overline{m})^{T}, \quad (6)$$

where $\overline{m}$ is the mean of the selected $p$ scattered points.

SSLPNPE seeks a linear projection matrix to make the local pixel neighborhood keep scatter minimized while the total scatter is maximized in the projection space. The optimal projection matrix $V = [v_{1}, v_{2}, \ldots v_{\iota}]$ can be acquired by solving the following eigenvalue problem.

$$Sv = \lambda Hv \quad (7)$$

3. Experimental Results and Discussion

3.1. Data

3.1.1. 2012 Houston Data

Experiments were conducted by using HSI and DSM data that were acquired on June 2012 across the University of Houston campus and the neighbouring urban area. The Houston data can be requested for free through the webpage (http://www.grss-ieee.org/community/technical-committees/data-fusion/2013-ieee-grss-data-fusion-contest/). The hyperspectral image had 144 spectral bands. The wavelength range of HSI was from 380 to 1050 nm. The same spatial resolution of HSI and DSM data is 2.5 m. The data for the entire scene consisted of $349 \times 1905$ pixels, containing 15 classes which were set by the DFTC via photo-interpretation. Table 1 shows the number of training and testing samples. The LiDAR data was collected on 22nd June 2012, between 14:37:55 and 15:38:10 UTC with 2000 feet the average height of the sensor. The HSI was collected on 23rd June 2012, between 17:37:10 and 17:39:50 UTC with 5500 feet the average height of the sensor. The data contain natural objects (e.g., water, soil, tree and grass) and artificial objects (e.g., parking lot, railway, highway, and road). Figure 2 gives the HSI, the LiDAR DSM, and the positions of the training and testing samples. There was a large cloud shadow in the image with no training samples and a significant number of testing samples to test the algorithms doing with cloud shadow.
Table 1. The 2012 Houston: number of training and testing samples.

<table>
<thead>
<tr>
<th>No</th>
<th>Name</th>
<th>Training</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Grass Healthy</td>
<td>198</td>
<td>1053</td>
</tr>
<tr>
<td>2</td>
<td>Grass Stressed</td>
<td>190</td>
<td>1064</td>
</tr>
<tr>
<td>3</td>
<td>Grass Synthetis</td>
<td>192</td>
<td>505</td>
</tr>
<tr>
<td>4</td>
<td>Tree</td>
<td>188</td>
<td>1056</td>
</tr>
<tr>
<td>5</td>
<td>Soil</td>
<td>186</td>
<td>1056</td>
</tr>
<tr>
<td>6</td>
<td>Water</td>
<td>182</td>
<td>143</td>
</tr>
<tr>
<td>7</td>
<td>Residential</td>
<td>196</td>
<td>1072</td>
</tr>
<tr>
<td>8</td>
<td>Commercial</td>
<td>191</td>
<td>1053</td>
</tr>
<tr>
<td>9</td>
<td>Road</td>
<td>193</td>
<td>1059</td>
</tr>
<tr>
<td>10</td>
<td>Highway</td>
<td>191</td>
<td>1036</td>
</tr>
<tr>
<td>11</td>
<td>Railway</td>
<td>181</td>
<td>1054</td>
</tr>
<tr>
<td>12</td>
<td>Parking Lot 1</td>
<td>192</td>
<td>1041</td>
</tr>
<tr>
<td>13</td>
<td>Parking Lot 2</td>
<td>184</td>
<td>285</td>
</tr>
<tr>
<td>14</td>
<td>Tennis Court</td>
<td>181</td>
<td>247</td>
</tr>
<tr>
<td>15</td>
<td>Running Track</td>
<td>187</td>
<td>473</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2832</td>
<td>12,197</td>
</tr>
</tbody>
</table>

Figure 2. The 2012 Houston—from top to bottom: the hyperspectral images (HSI) using bands 60, 40, and 30, as R, G, and B, respectively; LiDAR digital surface model (DSM) image; testing samples; training samples; and legend of different classes.

3.1.2. Rochester Data

The G-LiHT airborne imager [43] is a unique system that provides simultaneous CHM, DTM, LiDAR Point Cloud, hyperspectral reflectance image at the 1m spatial resolution on a wide range of airborne platforms and google earth overlay by keyhole markup language (KML) at 0.25m spatial resolution, which are helpful for land cover mapping. The G-LiHT data can be obtained for free through an interactive webpage (https://glihtdata.gsfc.nasa.gov). G-LiHT has been used to collect data over 6500 km² involving various ecological regions in the United States and Mexico. G-LiHT’s LiDAR data is from Riegli VQ-480 Scanning Lidar, and G-LiHT’s hyperspectral data is from the Headwall Hyperspec Imaging Spectrometer. G-LiHT provides already registered hyperspectral and LiDAR data.
Before using G-LiHT data to verify the performance of the land cover mapping algorithms, we need to make ground truth first. General image calibration usually requires a handheld GPS device and manual field measurement. However, the accuracy of the common handheld GPS device is 5 m, which may cause errors in image calibration. For the G-LiHT data, the spatial resolution was 1 m. With 0.25 m google earth overlay, G-LiHT data can be directly calibrated on the image with the high accuracy. A workflow to produce the G-LiHT data ground truth is proposed in this paper as shown in Figure 3. We used five types of data including the G-LiHT hyperspectral image (1 m) with spectral and spatial information, google earth RGB (0.25 m) with supplemental spatial information, G-LiHT LiDAR CHM (1 m) with canopy height information, G-LiHT LiDAR DTM (1 m) with terrain elevation information, and G-LiHT LiDAR slope (1 m) with texture information to determine the category first, that is to determine what is in the scene. For example, in Figure 4, we defined ten categories. Inspired by the calibration method [24], we utilized the superpixel segmentation on HSI first. Then, we utilized the result of the superpixel segmentation and the region of interest (ROI) selection tool to manually determine the initial object areas for each class in HSI. The result of the superpixel segmentation is beneficial to the selection of the boundary of the initial object areas. The spectral curve in HSI has an excellent ability to distinguish between substances, and the spatial resolution of HSI (1 m) is quite high. Subsequently, the selected initial object areas were manually refined by jointly checking HSI, google earth RGB, CHM, DTM, and the slope image. With the variety of information shown in Figure 3, we can eliminate some points with large differences in the same object area. Finally, the object areas were manually labeled, and the G-LiHT ground truth can be obtained as Figure 4d,e.

Experiments are done on the HSI, LiDAR-derived CHM data and LiDAR-derived DTM data that were acquired on June 2015 across the Rochester area (the location is at 43°18′36″N latitude, 70°57′36″W longitude). The hyperspectral image had 114 bands with spectral range 420–950 nm at a 1 m ground sampling distance (GSD). LiDAR CHM and DTM are also at a 1 m GSD. The data of the entire scene consists of 635 × 1169 pixels, including ten classes. In the experiment, we chose less than 200 pixels for each class in ground truth as the training set. The training samples were selected by regions and not a random selection. Therefore, the position of the training sample was concentrated, which is more in line with the way of sample selection in practical applications. Table 2 reports the land cover classes and the corresponding number of training and testing samples. Figure 4 shows the HSI, the LiDAR CHM, the LiDAR DTM, the positions of the testing samples, the positions of the training samples and legend of different classes.

Figure 3. The workflow to produce the G-LiHT data groundtruth.
3.2. Parameter Settings and Comparison of Methods

There are two parameters $\alpha$ and $s$ that need to be set in creating the EP for diagonal of the bounding box, volume, and area. The parameter $s$ represents the number of thresholds. For the parameter $\alpha$, the larger the $\alpha$, the more significant the differences among continuous images in the profile, whereas the smaller the $\alpha$, the fewer extrema there will be, where most image information is usually retained. P. Ghamisi et al. suggested that the proper $\alpha$ value can be determined between 2 and 5 [31]. In our experiments, $\alpha$ and $s$ were set to 5 and 7, respectively. The maximum value was divided into seven equidistant parts to create the EP for standard deviation and height. Considering the original image should also be involved in the profile, the size of the EP is $2s + 1$. The four connected connectivity rule was used to compute the profiles.

To reduce the noise, the KPCA is used for all the features, and the dimension is set to 70.

For the fusion method, 3000 points were randomly chosen as the local pixel neighborhood center. Figure 5 shows that the $3 \times 3$ window is the ideal window width. Therefore the width of the neighborhood window $w$ was set to $3 \times 3$. For the superpixel segmentation method, the balancing

**Table 2. Rochester: number of training and testing samples.**

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Name</td>
</tr>
<tr>
<td>1</td>
<td>Buildings</td>
</tr>
<tr>
<td>2</td>
<td>Trees</td>
</tr>
<tr>
<td>3</td>
<td>Grass/lawn</td>
</tr>
<tr>
<td>4</td>
<td>Water/pools</td>
</tr>
<tr>
<td>5</td>
<td>Residential_gray</td>
</tr>
<tr>
<td>6</td>
<td>Residential_white</td>
</tr>
<tr>
<td>7</td>
<td>Residential_coffee</td>
</tr>
<tr>
<td>8</td>
<td>Parking lot</td>
</tr>
<tr>
<td>9</td>
<td>Roads</td>
</tr>
<tr>
<td>10</td>
<td>Soil</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.** Rochester—from (a–e): the hyperspectral data using bands 40, 30, and 20, as R, G, and B, respectively; LiDAR digital terrain model (DTM) image; LiDAR canopy height model (CHM) image; testing samples; training samples; legend of different classes.
parameter $\mu$ can be automatically adjusted according to M. Liu et al. [40]. In Table 3, $3 \times 3 + \text{superpixel}$ means the spatial neighborhood determined by Equation (4), while $3 \times 3$ means that the spatial neighborhood is $3 \times 3$ window. Table 3 shows that the spatial neighborhood determined by $3 \times 3 + \text{superpixel}$ performs better than the spatial neighborhood determined directly by $3 \times 3$. In Table 3, the number of superpixels is from 10,000 to 200,000 with an interval of 5000. We take three ranges and average the OAs in each range. The results show that too many superpixels and too few superpixels will reduce the performance of the proposed method. To make the superpixels effectively remove the points with large spectral distance in the $3 \times 3$ neighborhoods, we recommend that the number of superpixels can be around $N/n^2$, where $N$ is the total number of pixels in the image and $n$ is the width of the window. Therefore, the number of superpixels in the Houston data was set to 60,000, and the number of superpixels in Rochester data was set to 74,000.

Our proposed method was compared with the generalized graph-based fusion (GGF) algorithm, which was the champion of the 2013 GRSS data fusion contest [1,22]. The numbers of nearest neighbors of features were set to 150. Since the original GGF algorithm only uses DSM, for the fair comparison, we replace the EP DSM features with the stacked features of EP CHM features and EP DTM features as one of the inputs of GGF.

The support vector machine (SVM) classifier was applied to all the methods. The parameters of the SVM classifier used cross-validation to determine parameters $c$ and $g$. Overall accuracy (OA), average accuracy (AA), and Kappa statistic were applied to evaluate the classification results.

![Figure 5](https://via.placeholder.com/150)

**Figure 5.** The classification performance as a function of varying the dimension for superpixel segmentation-based local pixel neighborhood preserving embedding (SSLPNPE) with different width of the neighborhood window.

<table>
<thead>
<tr>
<th></th>
<th>$3 \times 3$</th>
<th>$3 \times 3 + \text{Superpixel}$</th>
<th>OA</th>
<th>Number of Superpixels ($\times 10^3$)</th>
<th>OA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Houston</td>
<td>87.95</td>
<td>10–55</td>
<td>88.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>60–150</td>
<td>88.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>155–200</td>
<td>88.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rochester</td>
<td>97.56</td>
<td>10–65</td>
<td>98.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>70–150</td>
<td>98.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>155–200</td>
<td>98.78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The experimental part uses the following names for simplicity. $\text{HSI}$, $\text{CHM}$, $\text{DTM}$, $\text{DSM}$, $\text{CHM} + \text{DTM}$, $\text{HSI} + \text{DSM}$ and $\text{HSI} + \text{CHM} + \text{DTM}$ present the classification accuracies of HSI, LiDAR CHM, LiDAR DTM, LiDAR DSM, the stack of CHM and DTM, the stack of HSI and DSM and the stack of HSI, CHM and DTM, respectively. $\text{EP}_{\text{HSI}}$, $\text{EP}_{\text{CHM}}$, $\text{EP}_{\text{DTM}}$ and $\text{EP}_{\text{DSM}}$ present the classification accuracies of EPs used to HSI, LiDAR CHM, LiDAR DTM and LiDAR DSM, respectively. $\text{EP}_{\text{HSI}} + \text{EP}_{\text{DSM}}$ presents the classification accuracies of the stack of HSI EPs and LiDAR DSM EPs. $\text{EP}_{\text{CHM}} + \text{EP}_{\text{DTM}}$ presents the classification accuracies of the stack of CHM EPs and LiDAR DTM EPs.
means the same as $EP_{HSI} + EP_{DSM}$. $HSI + EP_{HSI} + EP_{DSM}$ shows the classification accuracies of the stacked features, which is the stack of HSI, HSI EPs and DSM EPs. $HSI + EP_{HSI} + EP_{DSM}$ and $HSI + EP_{HSI} + EP_{CHM} + EP_{DTM}$ mean the same as $HSI + EP_{HSI} + EP_{DSM}$. GGF and SSLPNPE present the classification accuracies of the method GGF for comparison and the proposed method SSLPNPE.

3.3. Results and Discussion

Figure 6 shows the classification performance as a function of varying the dimension for the Houston and Rochester data. For Houston data, when the reduced dimension is from seven to 32, the classification accuracy of the SSLPNPE was higher than the GGF. When the reduced dimension is from 33 to 40, the classification accuracy of the GGF is higher than the SSLPNPE. The dimension with the highest classification accuracy is 37 for GGF and 24 for SSLPNPE. For Rochester data, with varying the dimension, the classification accuracy of the SSLPNPE is a little higher than the GGF. The dimension with the highest classification accuracy is 31 for GGF and 40 for SSLPNPE.

![Figure 6](image)

**Figure 6.** The classification performance as a function of varying the dimension for the Houston and Rochester data.

Table 4 shows the classification accuracies obtained by different approaches for the Houston data. $HSI + DSM$ is higher than $DSM$ and $HSI$, which means that LiDAR DSM and hyperspectral data have complementary information. The use of EPs can significantly improve Kappa, OA, and AA since the EPs can efficiently extract spatial information and model the shape and size of different objects, which are helpful to precisely distinguish different classes of interest [31]. For example, $EP_{DSM}$ is significantly higher than $DSM$ and $EP_{HSI}$ is higher than $HSI$, which confirms the information extraction capability of the EP. The value $EP_{DSM} + EP_{HSI}$ is higher than $EP_{DSM}$ and $EP_{HSI}$, which proves that $EP_{DSM}$ and $EP_{HSI}$ have the complementary information. The stacking features of HSI, HSI EPs and DSM EPs have more information than the stacking features of HSI EPs and DSM EPs, but $HSI + EP_{HSI} + EP_{DSM}$ is lower than $EP_{HSI} + EP_{DSM}$ due to the problem of the curse of dimensionality.

Since GGF utilizes the complementary information of HSI, HSI EPs, and LiDAR EPs and maintains spectral similarity while reducing the dimensions of stacked features, GGF has higher classification accuracy than $HSI + EP_{HSI} + EP_{DSM}$. Figure 7 shows the classification map of the Houston data.

Figure 7 shows that the SSLPNPE performs better on the “parking lot 1” category and the “running track” category. From Figure 6, Table 4 and Figure 7, we can see that the classification performance of the SSLPNPE is similar to the GGF. The reason why SSLPNPE performs well is that SSLPNPE can remove redundant information of the stacked features and effectively retain spatial contexture information by optimizing spatial neighborhoods with the HSI superpixels. Table 6 shows the time of fusion and classification between the GGF algorithm and the SSLPNPE algorithm. The experimental device had a 2.3 GHz Intel i5 CPU. For the Houston data, the fusion time of the GGF and the SSLPNPE were 25.77 and 12.48 s. The SSLPNPE was faster than the GGF because the GGF needs to calculate
the norm distance between each feature to produce the graph, which needs a lot of calculations and memory. For the SSLPNPE, it intersected the superpixel and $3 \times 3$ local spatial neighborhoods to determine the final local spatial neighborhoods, which is much more efficient. Especially, the ERS superpixel method is a low computational algorithm [40]. As a result, our proposed method SSLPNPE was fast and effective for the fusion of HSI and LiDAR data.

Table 4. The 2012 Houston: classification accuracies achieved by different approaches.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>DSM</th>
<th>HSI</th>
<th>HSI + DSM</th>
<th>$E_{DSM}$</th>
<th>$E_{HSI}$</th>
<th>$E_{HSI} + E_{DSM}$</th>
<th>GGF</th>
<th>SSLPNPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54.26</td>
<td>99.05</td>
<td>81.25</td>
<td>99.62</td>
<td>98.77</td>
<td>99.15</td>
<td>99.62</td>
<td>98.77</td>
</tr>
<tr>
<td>2</td>
<td>55.24</td>
<td>99.30</td>
<td>99.30</td>
<td>66.43</td>
<td>95.10</td>
<td>95.10</td>
<td>99.30</td>
<td>95.10</td>
</tr>
<tr>
<td>3</td>
<td>58.58</td>
<td>86.01</td>
<td>86.01</td>
<td>67.16</td>
<td>70.71</td>
<td>82.65</td>
<td>75.47</td>
<td>92.54</td>
</tr>
<tr>
<td>4</td>
<td>14.16</td>
<td>72.90</td>
<td>72.99</td>
<td>24.65</td>
<td>83.76</td>
<td>88.29</td>
<td>80.26</td>
<td>86.48</td>
</tr>
<tr>
<td>5</td>
<td>7.24</td>
<td>59.94</td>
<td>59.36</td>
<td>64.77</td>
<td>68.05</td>
<td>68.34</td>
<td>66.12</td>
<td>68.05</td>
</tr>
<tr>
<td>6</td>
<td>42.50</td>
<td>78.84</td>
<td>78.84</td>
<td>98.10</td>
<td>87.76</td>
<td>92.98</td>
<td>87.48</td>
<td>99.11</td>
</tr>
<tr>
<td>7</td>
<td>5.48</td>
<td>74.93</td>
<td>75.50</td>
<td>41.50</td>
<td>89.15</td>
<td>90.01</td>
<td>86.84</td>
<td>96.64</td>
</tr>
<tr>
<td>8</td>
<td>15.09</td>
<td>71.23</td>
<td>71.58</td>
<td>49.82</td>
<td>68.07</td>
<td>72.98</td>
<td>80.70</td>
<td>78.25</td>
</tr>
<tr>
<td>9</td>
<td>81.38</td>
<td>99.60</td>
<td>99.60</td>
<td>98.79</td>
<td>100.00</td>
<td>99.60</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>10</td>
<td>69.13</td>
<td>97.46</td>
<td>97.46</td>
<td>97.04</td>
<td>98.73</td>
<td>98.94</td>
<td>97.25</td>
<td>98.94</td>
</tr>
<tr>
<td>11</td>
<td>86.28</td>
<td>83.30</td>
<td>83.52</td>
<td>65.22</td>
<td>84.15</td>
<td>86.95</td>
<td>85.81</td>
<td>90.70</td>
</tr>
<tr>
<td>12</td>
<td>33.42</td>
<td>80.42</td>
<td>80.68</td>
<td>61.08</td>
<td>81.90</td>
<td>85.22</td>
<td>83.14</td>
<td>89.25</td>
</tr>
<tr>
<td>13</td>
<td>0.2632</td>
<td>0.7983</td>
<td>0.7911</td>
<td>0.5784</td>
<td>0.8035</td>
<td>0.8401</td>
<td>0.8177</td>
<td>0.8833</td>
</tr>
</tbody>
</table>

Concerning Table 5, we can easily notice that $CHM + DTM$ performed better than $CHM$ and $DTM$, which shows the CHM and the DTM have complementary information for classification. Although the DTM is derived from the sum of the CHM and the DTM, $CHM + DTM$ is higher than $DSM$, and $HSI + CHM + DTM$ is higher than $HSI + DSM$, which means that the stack of the CHM and the DTM can reserve more information than the DSM. Same as the Houston data, the use of EPs can considerably improve the classification accuracy of the hyperspectral data and the LiDAR data as shown in Table 5. Figure 8 shows the classification map of the 2015 Rochester data. Figure 8 shows
that the SSLPNPE performs better on the “grass/lawn” category, the “residential_gray” category and the “road” category. As shown in Figure 6, Table 5 and Figure 8, the classification performance of the SSLPNPE is a little higher than the GGF. However, the computing time of SSLPNPE is lower than the computing time of the GGF in Table 6.

For HSI and LiDAR fusion classification, there were a few methods proposed in recent years. However, many fusion methods are very complicated. In the follow-up work, we will add a comparison of more fusion methods. According to the comparison of the real samples in G-LiHT data, the experimental results have been able to demonstrate the effectiveness of SSLPNPE.

Table 5. The 2015 Rochester: classification accuracies achieved by different approaches.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>CHM</th>
<th>DTM</th>
<th>DSM</th>
<th>CHM + DTM</th>
<th>HSI</th>
<th>HSI + DSM</th>
<th>HSI + CHM + DTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29.31</td>
<td>92.94</td>
<td>90.24</td>
<td>90.08</td>
<td>70.49</td>
<td>88.57</td>
<td>88.65</td>
</tr>
<tr>
<td>2</td>
<td>99.87</td>
<td>36.53</td>
<td>84.12</td>
<td>99.94</td>
<td>96.75</td>
<td>96.75</td>
<td>97.13</td>
</tr>
<tr>
<td>3</td>
<td>0.08</td>
<td>53.16</td>
<td>62.21</td>
<td>71.02</td>
<td>89.99</td>
<td>90.07</td>
<td>90.87</td>
</tr>
<tr>
<td>4</td>
<td>13.81</td>
<td>14.93</td>
<td>15.67</td>
<td>21.27</td>
<td>94.40</td>
<td>94.40</td>
<td>94.40</td>
</tr>
<tr>
<td>5</td>
<td>49.40</td>
<td>33.23</td>
<td>46.43</td>
<td>59.03</td>
<td>73.81</td>
<td>80.65</td>
<td>82.44</td>
</tr>
<tr>
<td>6</td>
<td>84.38</td>
<td>40.26</td>
<td>54.93</td>
<td>58.40</td>
<td>92.67</td>
<td>97.87</td>
<td>97.82</td>
</tr>
<tr>
<td>7</td>
<td>5.11</td>
<td>20.04</td>
<td>20.81</td>
<td>35.98</td>
<td>85.89</td>
<td>86.42</td>
<td>86.40</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>64.59</td>
<td>8.37</td>
<td>66.51</td>
<td>80.38</td>
<td>81.10</td>
<td>81.10</td>
</tr>
<tr>
<td>9</td>
<td>94.98</td>
<td>29.16</td>
<td>53.01</td>
<td>69.16</td>
<td>77.24</td>
<td>79.75</td>
<td>79.99</td>
</tr>
<tr>
<td>10</td>
<td>0.01</td>
<td>77.66</td>
<td>85.46</td>
<td>82.23</td>
<td>99.33</td>
<td>99.73</td>
<td>99.73</td>
</tr>
</tbody>
</table>

AA  37.69  46.75  52.13  65.36  86.10  89.53  89.88
OA  73.56  43.91  69.71  84.11  88.84  91.64  91.99
Kappa  0.6227  0.3292  0.5981  0.7778  0.8463  0.8849  0.8896

Table 6. Compare the time between generalized graph-based fusion (GGF) algorithm and superpixel segmentation-based local pixel neighborhood preserving embedding (SSLPNPE) algorithm.

<table>
<thead>
<tr>
<th>Data</th>
<th>Evaluation Index</th>
<th>GGF</th>
<th>SSLPNPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dimension</td>
<td>37</td>
<td>24</td>
</tr>
<tr>
<td>Houston</td>
<td>fusion time(s)</td>
<td>25.77</td>
<td>12.48</td>
</tr>
<tr>
<td></td>
<td>classification time(s)</td>
<td>2.09</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>dimension</td>
<td>31</td>
<td>40</td>
</tr>
<tr>
<td>Rochester</td>
<td>fusion time(s)</td>
<td>39.92</td>
<td>16.73</td>
</tr>
<tr>
<td></td>
<td>classification time(s)</td>
<td>0.98</td>
<td>1.00</td>
</tr>
</tbody>
</table>
4. Conclusions

This paper introduces a fusion method SSLPNPE for the classification of LiDAR and HSI using EPs, superpixel segmentation and LPNPE. In this paper, the feature extraction methods EP have been utilized for spatial and elevation information extraction from HSI and LiDAR data. Then, the derived features of each source are fused by SSLPNPE. Using the labeled samples created by our proposed workflow that suits to calibrate the G-LiHT data, the final classification map is produced by SVM classifier. Our proposed method is not only fast and effective in hyperspectral and LiDAR data fusion, but also can be successfully applied to the G-LiHT data. The proposed workflow can be generalized to any fusion classification method and the G-LiHT data in any scene. This paper presents that processing the CHM and the DTM separately can achieve higher classification accuracy than the DSM for the first time in the remote sensing community.


Funding: This work was supported in part by the China Scholarship Counsil program (201706960055). This work was also partially financed in part by the National Natural Science Foundation of China (41671342, 61871177, 61801357, 61571345, 91538101, 61501346, 61502367 and 61701360) and the 111 project (B08038). It was also partially supported by the Natural Science Basic Research Plan in Shaanxi Province of China (No. 2016JQ6023), the Yangtze River Scholar Bonus Schemes of China (No. CJT160102), and the Ten Thousand Talent Program. It was also partially supported by the Fundamental Research Funds for the Central Universities JB180104, and General Financial Grant from the China Postdoctoral Science Foundation (No. 2017M620440). This work was also partially funded by Zhejiang Provincial Natural Science Foundation of China (LR19D010001), by Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (18R05) and by the K. C. Wong Magna Fund in Ningbo University.

Acknowledgments: The authors would like to thank the IEEE GRSS Image Analysis and Data Fusion Technical Committee that support the Houston data. The authors would like to thank NASA for providing the G-LiHT data.

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

References


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).