

Intrinsic Scene Properties from Hyperspectral Images and LiDAR

Xudong Jin

Harbin Institute of Technology
Harbin, China

hitjxd@foxmail.com

Yanfeng Gu

Harbin Institute of Technology
Harbin, China

guyf@hit.edu.cn

Abstract

In this paper, we propose a joint intrinsic decomposition framework to recover intrinsic scene properties from a remote sensing hyperspectral images (HSIs) and the aligned Light Detection And Ranging (LiDAR) dataset. Our motive is similar to that of Scene-SIRFS for RGB-D images [4, 5], but our method differs significantly from it both in theory and practice, due to the challenges following the complexity of the data considered. More specifically, we propose a novel HSI rendering model with sub-pixel reflectances and shapes which are provided by unorganized LiDAR points, rather than basing on pixel-level rendering model as previously done. Since our method operates directly on point clouds, the output of our model is an intrinsic hyperspectral point cloud (IHSPC) where each point possesses not only coordinates and normals, but also the reflectance over each wavelength. As illustration of applications, we show how to relight the IHSPC to re-render a new HSI with higher spatial resolution and image quality, or under different illuminations, as well as to increase the performance of classification tasks.

1. Introduction

Typically, intrinsic image decomposition (IID) refers to the separation of reflectance and shading component from a single RGB [25, 29, 30, 31]. Some more ambitious methods aims to recover at the same time shape and illumination in addition to reflectance and shading from a single image, this can be done by depth prediction using a novel convolutional neural network (CNN) [19, 28], or by natural image statistics analysis [3], the latter usually requires the inputs to be images with only single segmented objects. As for the more general case, i.e., recovering the shape, illumination, and reflectance from the whole image scene where objects may occlude or support one another and spatially-varying illumination exists, additional sources, either image sequences illuminated under varying illumination [6, 21, 26], depth cues acquired by depth sensor [8, 12, 14, 20], or incident

light direction [1] are needed. Though in computer vision there exists some efficient IID models for RGB or RGB-D images, these methods usually can not be extended easily to HSIs which may contain hundreds of bands, and when they do, they usually work poorly as a remote sensing scene is much more complicated than that in computer vision. Moreover, though the HSI "intrinsic image" problem has already investigated by some previous works [16, 17, 18], the more general "shape, illumination, and reflectance" problem for HSI has seen little success. In this paper, we would consider the problem in the context of remote sensing image processing: given HSI and aligned LiDAR, recover the shape, illumination and reflectance via joint intrinsic image decomposition of the two.

Nowadays, a variety of remote sensing technologies and imaging sensors can be used for Earth observation and to acquire abundant land cover information. What's more, the simultaneous acquisition of multi-sensor data (or the repeated observations of an overlapping region with different sensors) becomes more easily accessible. Each imaging method has its own advantages and limitations, e.g., hyperspectral imaging produces an image with rich spectral information but relatively lower spatial resolution; whereas, the LiDAR can make a dense 3D laser scanning of ground surface and provide 3D representations of the object, but it is weak in the discrimination of different materials with same height. Hence, considering the complementarity, it may be desirable to jointly analyze two or more heterogeneous datasets. Our work can methodologically be deemed as a novel fusion method of HSI and LiDAR, but differs from the existing works in following ways: 1) instead of utilizing a LiDAR derived Digital Surface Model (DSM) [11] which has the same spatial resolution and organized grid structure with HSI, we use directly the unorganized point clouds and explore the relation between HSI pixel and somewhat subpixel LiDAR points, by doing which we can make full use of the information contained in the data; 2) instead of extracting human-designed features from the two datasets and considering how to combine them together [15, 27], we try to physically infer the properties of a scene (coordinates,

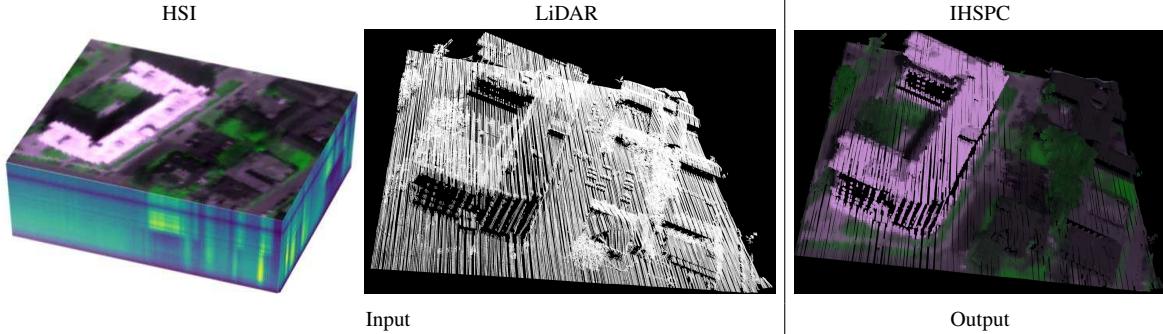


Figure 1. The proposed model takes as inputs a HSI and LiDAR point cloud, and produce as output an IHSPC where each point possesses not only coordinates and normals, but also the reflectance over each wavelength.

height, surface, illumination, reflectance, etc) that together reproduce a HSI and LiDAR.

Our motive is similar to that of the Scene-SIRFS [4, 5] for a single RGB-D image, which models a scene using a mixture of shapes and takes as input a RGB image and a coarse and noisy depth image while yields as output an improved depth map, surface normals, a reflectance image, a shading image, and illumination. But in both theory and practice, there exists significant distinctions. At first sight it may seem trivial to first derive a DSM from LiDAR and then apply a method similar to Scene-SIRFS for RGB-D images. But a further reflection corrects the first impression, by showing that not only the spatial resolution and image quality of HSIs are usually lower than that of RGB images, but also there exists some rough and inconsistent surfaces which Scene-SIRFS can hardly handle in outdoor remote sensing scenes. Consequently, our method operates directly on point clouds, and tries to discover the inner imaging relation between pixels in HSI and LiDAR points. The output of our model is thus an IHSPC where each point possesses not only coordinates and normals, but also the reflectance over each wavelength, as shown in Figure 1.

Another challenge (even more sticky to handle) is how to quantitatively evaluate the accuracy of our model, as there are no pre-existing real-world or Pseudo-Synthetic remote sensing datasets with ground-truth height, surface normals, reflectance, and illumination. It is extremely difficult to produce such ground-truth information, especially for HSIs where reflectance and illumination over each wavelength must be considered. Sean Bell *et al.* proposed a method in [7] to determine ground-truth reflectance for large scale indoor scenes by collecting human judgment, but it can not generalize to outdoor remote sensing scenes as they are much more complicated and only a small datasets are available. Maybe the closest data set is the SHARE 2012 data, which provides ground reflectance of some in-scene materials and ground based LiDAR [10]. However, the reflectance over whole scene is still absence and the geometric distor-

tion of the HSI data is not suitable for our method. Since the point of departure of our investigation is to infer a model that reproduces a HSI or LiDAR, so one way to indirectly evaluate our model is to demonstrate the efficiency of re-rendering the input HSI under different illumination conditions and with different spatial resolutions. Taking as inputs a HSI and a LiDAR and generating a HSI with higher spatial resolution can be deemed as hyperspectral super-resolution method. While traditional super-resolution methods usually work with a high-resolution (HR) RGB and low-resolution HSI (LR) [9, 22], the proposed method provides an alternative when only LR HSI and LiDAR data are offered.

Our paper is as follows: in Section 2 we introduce a new HSI imaging model considering sub-pixel reflectances and shapes, in Section 3, we show how to connect HSI with LiDAR in our model, in Section 4 we show how to estimate light and recover reflectance, in Section 5 we present experiments and in Section 6 we conclude.

2. HSI rendering model

The "intrinsic image" problem usually assumes that an image can be separated as the production of intrinsic component and shading component [13, 23], which can be expressed as:

$$I = sR \quad (1)$$

where I is the image, s denotes the shading component, and R denotes the reflectance component.

While in "shape, illumination, and reflectance from shading" model, shading is no longer treated as a kind of image, but the product of shape and illumination [1, 2, 3]:

$$I = R + S(Z, L) \quad (2)$$

where R is a log-reflectance image, Z is a depth-map, L is a spherical-harmonic (SH) model of illumination, and $S(Z, L)$ is a rendering engine which produces a log-shading image given Z and L .

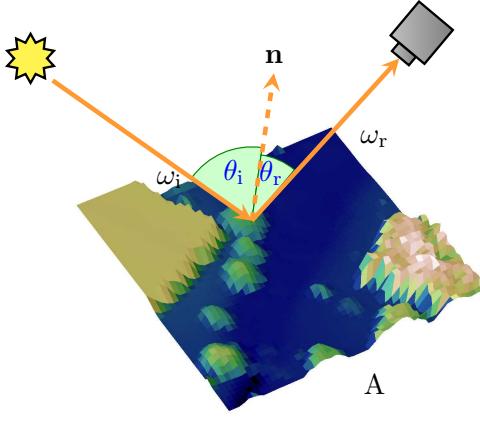


Figure 2. Illustration of our imaging model considering sub-pixel mixing materials and shapes, where the incoming light direction is ω_i , outgoing direction is ω_r , and the BRDF, $f_r(x, y; \omega_i, \omega_r)$ and surface normal, \mathbf{n} are spatially varying functions.

These models do not fit quite well when applied to HSIs. It makes sense in computer vision to assume that a pixel from a RGB image presents single material and possesses smooth shape, while in HSI processing, we must consider sub-pixel mixing materials and shapes due to its lower spatial resolution. Hence, we firstly apply BRDF to how each pixel in HSI is rendered with sub-pixel reflectances and shapes.

By definition, BRDF is:

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{L_i(\omega_i) \cos \theta_i d\omega_i} \quad (3)$$

where L_r is radiance, $L_i(\omega_i) \cos(\theta_i)$ is irradiance, and θ_i is the angle between ω_i and surface normal, \mathbf{n} . It measures the ratio of reflected radiance exiting along ω_r to the irradiance incident on the surface from direction ω_i .

Now we assume a pixel k taken from the given HSI is a sample of ground surface A . As shown in Figures 2, let the incident light direction be ω_i , the observation direction of the sensor be ω_r , the angle between surface normal \mathbf{n} and ω_i be θ_i , and the angle between \mathbf{n} and ω_r be θ_r . Applying the spatially varying version of BRDF, then for a surface element dA of A , the reflected differential of radiance emitted along ω_r is

$$dL_r(x, y; \omega_r) = L_i(\omega_i) \cos \theta_i f_r(x, y; \omega_i, \omega_r) d\omega_i \quad (4)$$

where (x, y) is the spatial coordinates of dA .

Then the differential of flux received by the sensor from dA is:

$$d\Phi = L_r(x, y; \omega_r) dA \cos \theta_r \cdot d\omega_r \quad (5)$$

Integrating over area A we get the radiant flux received by sensor as:

$$\Phi = \int_{\Omega_r} \int_A L_r(x, y; \omega_r) \cos \theta_r \cdot dA dy d\omega_r \quad (6)$$

here we disregard inter-reflections and occlusions for simplicity.

Finally, our imaging model of HSI is calculated as

$$\begin{aligned} I &= \frac{\Phi}{\Omega' A^\perp} \\ &= \frac{\Omega_r \cdot \int_A L_r(x, y; \omega_r) \cos \theta_r \cdot dA dy}{\Omega' A} \\ &= \frac{E_{\omega_i} \cdot \int_A f_r(x, y) \cos \theta_i \cos \theta_r \cdot dA dy}{\int_A \cos \theta_r \cdot dA dy} \end{aligned} \quad (7)$$

where $E_{\omega_i} = \int_{\Omega_i} L_i(\omega_i) \cdot d\omega_i$ denotes the illumination intensity.

3. Connection with LiDAR

In Section 2, we propose a novel imaging model rendering HSI with sub-pixel reflectances and shapes. But it is given in continuous domain, and the integration form is too complicated to calculate in practice. In this section, we would show how to combine HSI with LiDAR and apply Monte Carlo integration to derive the discrete form of our model. Figure 3 shows how the connection between HSI and LiDAR is considered.

The points in LiDAR are discrete and unorganized, and we assume the point cloud is dense enough to provide the sub-pixel shapes for HSI (which is often true since the spatial resolution of HSI is relatively lower, and when repeated observations available, multiple point clouds can be merged into a much denser one). Another assumption we make is that the LiDAR is a random and uniform sample of the ground surface. The sampling points are actually determined by flight path, which possesses some randomness but is not strictly random. Similarly, there could be in a LiDAR point cloud some denser areas on one hand, and some holes on the other, so it is not satisfactorily uniform. Hence, this assumption on which our later analysis is based is not exactly true, but we think the error is acceptable. Additionally, some re-sampling methods can be applied to construct a random and uniform data when required.

We denote the sub-set of LiDAR points, which provides the sub-pixel shapes of pixel k of HSI, as $P_k = \{\mathbf{p}_{kj}\}_{j=1}^{m_k}$, where $\mathbf{p}_{kj} = (x_{kj}, y_{kj}, z_{kj})$ and m_k is the number of points.

We first recall what Monte Carlo integration says, for a multidimensional definite integral

$$F = \int_{\Omega} f(\mathbf{x}) \cdot d\mathbf{x} \quad (8)$$

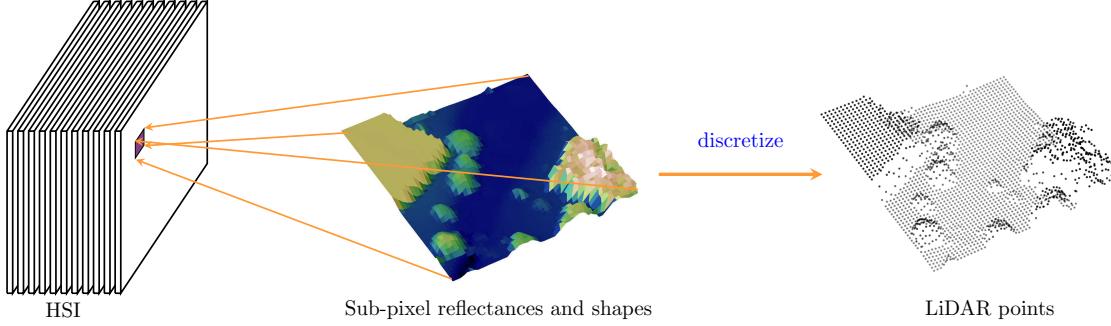


Figure 3. Illustration of how HSI and LiDAR are connected in our model. For each pixel of HSI, we assume some sub-pixel reflectances and shapes which re-render that pixel. Then the sub-pixel shapes are provided and estimated by discrete LiDAR points.

where $\Omega \in \mathbf{R}^m$ has volume

$$V = \int_{\Omega} d\mathbf{x} \quad (9)$$

given N uniform samples, $\{\mathbf{x}_i\}_{i=1}^N$, on Ω , F can be approximated by

$$F \approx V \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i) \quad (10)$$

Now, we rewrite Eq. 7 as

$$I = \frac{E_{\omega_i} \cdot \int_A g_1(x, y) g_2(x, y) \cdot d\mathbf{x} d\mathbf{y}}{\int_A g_2(x, y) \cdot d\mathbf{x} d\mathbf{y}} \quad (11)$$

where $g_1(x, y) = f_r(x, y) \cos \theta_i(x, y)$ and $g_2(x, y) = \cos \theta_r(x, y)$, note θ_i and θ_r are spatially varying functions of (x, y) .

Given a random and uniform (as we assumed previously) sample points of surface A , $P_k = \{\mathbf{p}_{kj}\}_{j=1}^{m_k}$, we can apply an extended form of Monte Carlo integration and approximate I as

$$I_k \approx E_{\omega_i} \frac{\sum_{j=1}^{m_k} g_1(\mathbf{p}_{kj}) g_2(\mathbf{p}_{kj})}{\sum_{j=1}^{m_k} g_2(\mathbf{p}_{kj})} \quad (12)$$

which can be rewrite as

$$I_k = E_{\omega_i} \frac{\sum_{j=1}^{m_k} f_r(\mathbf{p}_{kj})(\vec{\omega}_i \cdot \vec{\mathbf{n}}_{kj})(\vec{\omega}_r \cdot \vec{\mathbf{n}}_{kj})}{\sum_{j=1}^{m_k} (\vec{\omega}_r \cdot \vec{\mathbf{n}}_{kj})} \quad (13)$$

where $\vec{\mathbf{n}}_{kj} = [N_{kj}^x, N_{kj}^y, N_{kj}^z]^T$.

Eq. 13 gives an approximation of Eq. 7, from which we can see that in our model a pixel I_k of HSI is re-rendered by illumination intention, E_{ω_i} , sub-pixel distribution of surface normals provided by LiDAR, $\vec{\mathbf{n}}_{kj}$, sub-pixel reflectances remained to be recovered, $f_r(\mathbf{p}_{kj})$, incoming light direction, $\vec{\omega}_i$, and shooting direction of HSI sensors, $\vec{\omega}_r$. Our model explains more clearly how HSI is rendered and describes how HSI and LiDAR are connected by imaging principles.

4. Light estimation and reflectance recovery

Section 3 shows how our model is simplified by connecting HSI with LiDAR, which deals with only multiplication and addition calculations, rather than integration, and can be calculated efficiently. However the question is still not free from difficulty with so many unknown variables to solve. For example, the viewing angle of the sensor is rather challenging to estimate from a single HSI, to make the problem worse, it may change during the image acquisition. But with the observation that the view point of Earth remote sensing images is usually from top the bottom, we would roughly assume $\vec{\omega}_r = (0, 0, 1)$ for simplicity. Hence, we rewrite Eq. 13 as

$$I_k = E_{\omega_i} \frac{\sum_{j=1}^{m_k} f_r(\mathbf{p}_{kj})(\vec{\omega}_i \cdot \vec{\mathbf{n}}_{kj})(N_{kj}^z)}{\sum_{j=1}^{m_k} N_{kj}^z} \quad (14)$$

There is another question to which the solution needs to be found, *i.e.*, the estimation of light direction, $\vec{\omega}_i$. In a remote sensing scene, the light comes from a distant source and be deemed uniformly distributed over surface. Yet the difficulty remains if we previously know nothing about how the material distributes over surface, as we can not tell whether a intensity change of HSI is resulted from material change or surface normal change. To address this problem, we would first extract some critical objects where we assume there is no material change, which are typically some man-made strictures like building surfaces. This is done by segmenting the LiDAR points in terms of smoothness constraint. A set of clusters $\{S_c\}_{c=1}^u$ are obtained, where each cluster is a set of points that are considered to be consist of single material.

For each cluster we have a set of overdetermined equations

$$I_{ck}(\lambda) = E(\lambda) f_c(\lambda) \vec{\omega}_i \cdot \frac{\sum_{j=1}^{m_k} N_{kj}^z \vec{\mathbf{n}}_{kj}}{\sum_{j=1}^{m_k} N_{kj}^z}, \quad k = 1, 2, \dots \quad (15)$$

where λ is wavelength. And recall $\{\mathbf{n}_{kj}\}_{j=1}^{m_k}$ is the normals of a sub-set of LiDAR points, which provides the sub-pixel shapes of pixel k of HSI. Solving these overdetermined system we give a estimation of global light direction, $\vec{\omega}_i$.

Once global light direction is given, our reflectance recovery degenerates to "intrinsic image" problems

$$I(\lambda) = f_r(\lambda)S(\omega_i, \mathbf{n}) \quad (16)$$

of which the particularity is that the shading is rendered by our sub-pixel model

$$S(\omega_i, \mathbf{n}) = E(\lambda)\vec{\omega}_i \cdot \frac{\sum_j N_{kj}^z \vec{\mathbf{n}}_{kj}}{\sum_j N_{kj}^z} \quad (17)$$

And our prior on reflectance is based on the observation that locally, neighboring points share similar reflectance; and globally, the distribution of reflectance appears sparsity. Hence we assume following relations

$$f_i = \sum_{j=1}^n \omega_{ij} f_j \quad (18)$$

where L1-graph coding is used to calculate ω_{ij} :

$$\min_{\alpha^i} \|\alpha^i\|_1 \quad \text{subject to} \quad \mathcal{I}_i = \mathbf{D}^i \alpha^i \quad (19)$$

where matrix $\mathbf{D}^i = [\mathcal{I}_1, \dots, \mathcal{I}_{i-1}, \mathcal{I}_{i+1}, \dots, \mathcal{I}_n, \mathbf{I}_d] \in R^{d \times (d+n-1)}$ and $\alpha^i \in R^{d+n-1}$. Then, set $\omega_{ij} = \alpha_j^i$ if $i > j$, and $\omega_{ij} = \alpha_{j-1}^i$ if $i < j$.

The output of model is thus an IHSPC, $\{\mathbf{P}'_i\}_{i=1}^K$

$$\mathbf{P}'_i = (x_i, y_i, z_i, N_i^x, N_i^y, N_i^z, f_r(\lambda_1), \dots, f_r(\lambda_d)) \quad (20)$$

where each point possesses not only coordinates and normals, but also the reflectance over each wavelength.

With $\{\mathbf{P}'_i\}_{i=1}^K$, we can then apply Eq. 14 to re-render a new HSI with higher spatial resolution and image quality, or under different illuminations.

5. Experiment

As we have mentioned in introduction, it is rather challenging to quantitatively evaluate the accuracy of our model, as there are no pre-existing HSI datasets with ground-truth heights, surface normals, reflectance, and illumination. Even, there are only few open accessible multi-source optical remote sensing datasets that contain both HSI and LiDAR, for the data are expensive to acquire. Usually, these data are provided for classification tasks and only land use and land cover classes would be given (if there is any) as ground truth. Unlike classification labels, which can at least be roughly produced by visual perception, ground-truth reflectance in remote sensing scenes is hardly possible to produce in practice. Hence, we would evaluate our model indirectly.

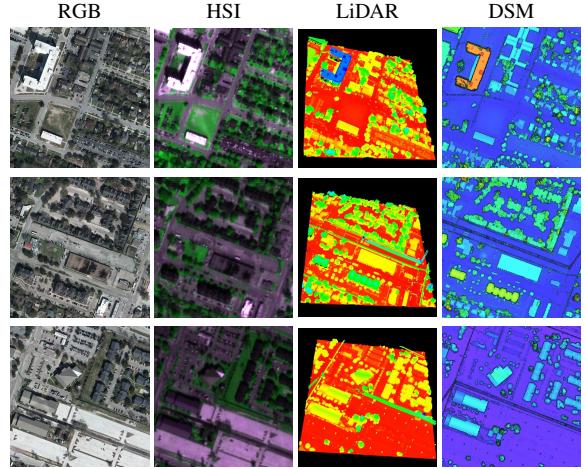


Figure 4. Three scenes of data used in our experiments. The first column: very high-resolution RGBs imagery at a 5-cm GSD. The second column: false-color composite images of HSI at 1-m GSD. The third column: LiDAR point cloud. The fourth column: LiDAR derived DSM at 0.5-m GSD.

5.1. Data description

In our experiment, we use three remote sensing scenes took from the "grss_dfc_2018" dataset, which is used for the 2018 IEEE GRSS Data Fusion Contest [24]. More information can be find at following website: <http://www.grss-ieee.org/community/technical-committees/data-fusion>. The data were acquired by the National Center for Airborne Laser Mapping (NCALM) on February 16, 2017 between 16:31 and 18:18 GMT. The dataset, which are originally meant for multi-source optical remote sensing data fusion and analysis, provide following multi-source data acquired by three different sensors:

- Multispectral LiDAR point cloud data, intensity rasters and DSMs at a 0.5-m ground sampling distance (GSD)
- Hyperspectral data at a 1-m GSD
- Very high-resolution (VHR) RGB imagery at a 5-cm GSD

Though a multispectral LiDAR point cloud data is provided, we do not use the intensity information of it in our model. Figure 4 gives the illustrations of RGB, HSI, LiDAR, and DSM data of three scenes we use. Note, only HSI and LiDAR point cloud are taken as inputs in our model.

5.2. Re-render HSI under different illuminations

One way to evaluate our model is to show that with outputs from our model we can re-render a new HSI under different illuminations or resolution sensibly. The re-render results under three light settings are shown in Figure 5, from which we can see that our model can re-

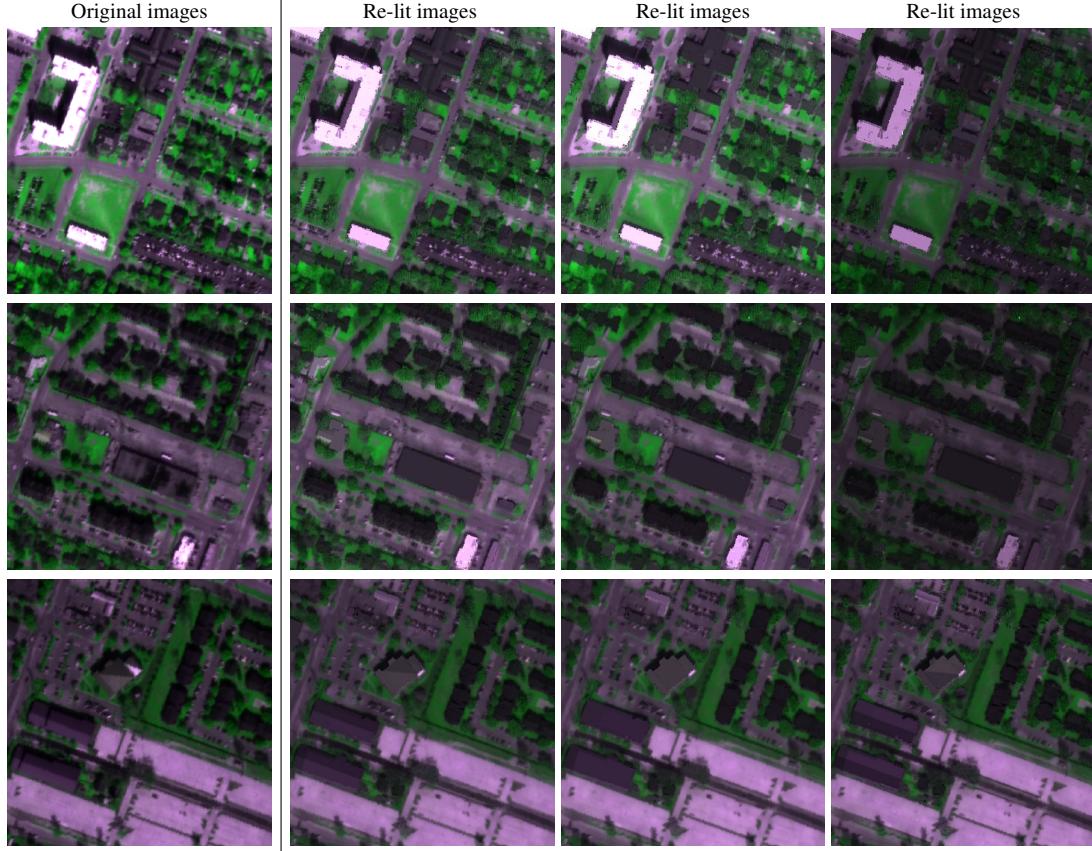


Figure 5. Re-render HSI under different illuminations. The first column: original images. The second column: relight under estimated light, i.e., $\omega_i = (-0.07, -0.31, -0.85), (-0.01, -0.33, -0.77)$, and $(-0.06, -0.34, -0.73)$, respectively. The third column: relight under light perpendicular to ground, i.e., $\omega_i = (0, 0, -1)$. The fourth column: relight under $\omega_i = (0.07, 0.31, -0.85), (0.01, 0.33, -0.77)$, and $(0.06, 0.34, -0.73)$, respectively.

render a new HSI of clearer spatial details. The estimated light directions for three scenes are $(-0.07, -0.31, -0.85)$, $(-0.01, -0.33, -0.77)$, and $(-0.06, -0.34, -0.73)$, respectively. It makes sense that they resemble each other as they are took from a same large scene. Some defects of our model arise from its disregarding occlusions and cast shadows, as a result, it fails to recover accurate reflectance where occlusions affect servilely or to handle shadow areas while re-rendering.

We know that two spectral energy distributions $I_1(\lambda)$ and $I_2(\lambda)$ can either arise from a material change, a surface orientation change, or illumination change. This uncertainty causes inconvenience in some HSI processing tasks, e.g., only material changes may be interested in land cover classification task. The primary goal of our work is to decrease the complexity of these tasks by separating intrinsic component from other component influenced by environment. But on the other hand, some methods (e.g., deep learning) do desire a more complex datasets in the training step to improve model’s generalization ability and reduce overfit-

ting. In order to train a model invariant to translation, viewpoint, resolution, or illumination, data augmentation is desired to enlarge the training dataset, especially for HSI processing in which the available dataset is very small. So our method also shows potential in data augmentation, for it can re-render physically sensible new HSI data under different translation, viewpoint, resolution, or illumination, apart from some popular data augmentation techniques like rotation, flip, scale, and so on.

5.3. Re-render HSI under higher spatial resolutions

While HSI provides a high spectral resolution, its spatial resolution is significantly lower due to hardware limitations. Here we would show via re-rendering the IHSPC that our model produces, we can reconstruct a high-resolution HSI of high quality. In our experiment, we first generated a LR HSI via spectral downsampling of original HR HSI that is used as ground-truth, and then used the LR HSI and LiDAR to obtain an IHSPC, which was thus utilized to re-render a new HR HSI. As comparison, we also generated a normal

Methods	Scene1	Scene2	Scene3
HSPC	10.5	8.4	8.8
IHSPC	8.3	6.3	7.0

Table 1. Quantitative results (on RMSE) of the test methods.

Methods	Scene1	Scene2	Scene3
HSPC	4.4	3.8	3.5
IHSPC	3.2	3.1	2.3

Table 2. Quantitative results (on SAM) of the test methods.

Data	HSPC	MSLPC	IHSPC
OA	66.7	72.8	80.1

Table 3. Overall classification accuracy (%) of HSPC, MSLPC, IHSPC.

hyperspectral point cloud (HSPC) via matching the nearest pixel of HSI to each LiDAR point. To evaluate the quality of the reconstructed HSIs, indexes as root mean square error (RMSE), where the images are on 8-bit intensity, and the spectral angle mapper (SAM), which is given in degrees and the smaller the better, are used.

The results in terms of RMSE and SAM are reported in Table 1 and Table 2, respectively, where the proposed method achieves best scores. The reconstructed HSIs in 746 nm by competing methods are shown in Figure 6, from which we could see that the obtained IHSPC can re-render a HR HSI of much higher quality than HSPC. This is benefit from the sub-pixel reflectance and shapes assumed in our model. The proposed method thus provides an alternative super-resolution technique taken as inputs a HSI and a LiDAR point cloud when HR RGB images are not available.

5.4. Classification result

Another possible way of evaluation is to use a downstream task like point cloud classification. Since the intrinsic hyperspectral reflectance could eliminate the impact of illumination and shading, which would follow an increase in material discrimination ability. To demonstrate that, we classified three kind of data, *i.e.*, HSPC, IHSPC and multispectral LiDAR point cloud (MSLPC), using support vector machine (SVM) under same parameters, and compared the classification result of the three.

The ground-truth classification labels of the point cloud were made by human judgment, with the assistance of VHR RGB image. Finally, nine classes were given, *i.e.*, healthy grass, stressed grass, evergreen trees, deciduous trees, residential buildings, non-residential buildings, roads, sidewalks, parking lots. We picked out randomly four small patches per class as training set, that eventually 10,078 points were selected for training, and the remaining 2,088,358 points were used for test. The training and

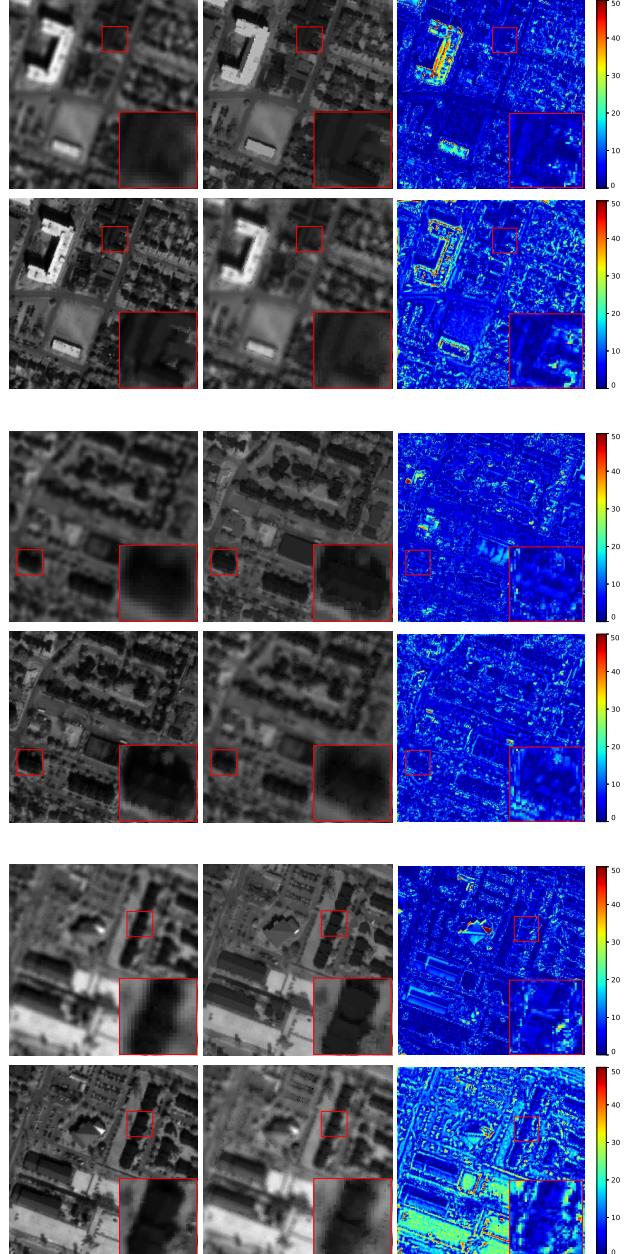


Figure 6. Reconstructed images of three scenes (top two rows, middle two rows, and bottom two rows) at wavelength 746 nm. For every two rows, the first column shows the LR image (top) and the ground-truth image (bottom); the second column shows the reconstructed results of IHSPC (top) and HSPC (bottom); and the third column shows the absolute difference of IHSPC (top) and HSPC (bottom).

test data is given in Figure 7. The overall classification accuracy (OA) scores are shown in Table 3, from which we can see that IHSPC obtains the best score, outperforms both HSPC that is generated by directly assigning HSI spectra to LiDAR and MSLPC that is obtained by single hyperspectral

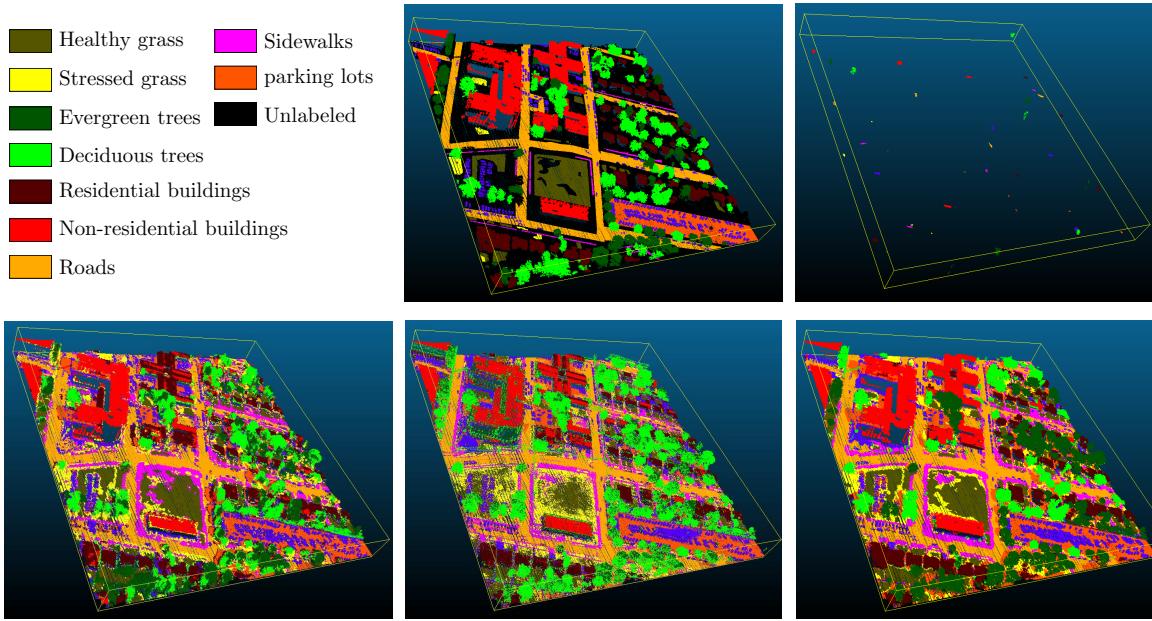


Figure 7. The first row: legend (left), test data (middle), training data (right). The second row: the classification maps of HSPC (left), MSLPC (middle) and IHSPC (right).

lidar sensor.

6. Conclusion

We have presented a joint intrinsic decomposition model, which recovers intrinsic scene properties from a remote sensing HSI and a LiDAR point cloud. We have done this by proposing a novel HSI rendering model with sub-pixel reflectances and shapes provided by LiDAR points, and thus combine HSI and LiDAR together.

The output of our model is a intrinsic hyperspectral point cloud , which we show can be applied to re-render a new HSI with higher spatial resolution and image quality, or under different illuminations, and it is easy to imagine other applications such as HSI or LiDAR classification and segmentation. We have also discussed some challenges remained in the sense of remote sensing intrinsic image decomposition, which still require further efforts, *e.g.*, it is rather difficult to generalize some existing IID methods to HSI which may contain hundreds of bands, or to produce HSI and LiDAR dataset with ground-truth reflectance, surface normals, and illuminations.

Acknowledgment

The authors would like to thank the National Center for Airborne Laser Mapping and the Hyperspectral Image Analysis Laboratory at the University of Houston for acquiring and providing the data used in this study, and the

IEEE GRSS Image Analysis and Data Fusion Technical Committee.

This work was supported in part by the National Science Fund for Excellent Young Scholars under Grant 61522107 and the National Natural Science Foundation of key international cooperation under the Grant 61720106002. This work was also supported by key research and development project of Ministry of Science and Technology (No.2017YFC1405100). (Corresponding author: Yanfeng Gu.)

References

- [1] J. T. Barron and J. Malik. High-frequency shape and albedo from shading using natural image statistics. In *CVPR*, pages 2521–2528, June 2011.
- [2] J. T. Barron and J. Malik. Color constancy, intrinsic images, and shape estimation. In *ECCV*, pages 57–70, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- [3] J. T. Barron and J. Malik. Shape, albedo, and illumination from a single image of an unknown object. In *CVPR*, pages 334–341, June 2012.
- [4] J. T. Barron and J. Malik. Intrinsic scene properties from a single rgb-d image. In *CVPR*, pages 17–24, June 2013.
- [5] J. T. Barron and J. Malik. Intrinsic scene properties from a single rgb-d image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(4):690–703, April 2016.
- [6] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, May 2007.

- [7] S. Bell, K. Bala, and N. Snavely. Intrinsic images in the wild. *ACM Trans. Graph.*, 33(4):159:1–159:12, July 2014.
- [8] Q. Chen and V. Koltun. A simple model for intrinsic image decomposition with depth cues. In *ICCV*, December 2013.
- [9] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Transactions on Image Processing*, 25(5):2337–2352, May 2016.
- [10] A. M. Giannandrea, N. Raqueno, D. W. Messinger, J. Faulring, J. P. Kerekes, J. van Aardt, K. Canham, S. Hagstrom, E. Ontiveros, A. Gerace, J. Kaufman, K. M. Vongsy, H. Griffith, B. D. Bartlett, E. Ientilucci, J. Meola, L. Scarff, and B. Daniel. The share 2012 data campaign, 2013.
- [11] Y. Gu, Q. Wang, X. Jia, and J. A. Benediktsson. A novel mkl model of integrating lidar data and msi for urban area classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(10):5312–5326, Oct 2015.
- [12] M. Hachama, B. Ghanem, and P. Wonka. Intrinsic scene decomposition from rgbd images. In *ICCV*, pages 810–818, Dec 2015.
- [13] B. K. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(4):277 – 299, 1974.
- [14] J. Jeon, S. Cho, X. Tong, and S. Lee. Intrinsic image decomposition using structure-texture separation and surface normals. In *ECCV*, pages 218–233, Cham, 2014. Springer International Publishing.
- [15] S. Jia, J. Hu, Y. Xie, L. Shen, X. Jia, and Q. Li. Gabor cube selection based multitask joint sparse representation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6):3174–3187, June 2016.
- [16] X. Jin and Y. Gu. Superpixel-based intrinsic image decomposition of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8):4285–4295, Aug 2017.
- [17] X. Jin, Y. Gu, and T. Liu. Intrinsic image recovery from remote sensing hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):224–238, Jan 2019.
- [18] X. Kang, S. Li, and J. A. Benediktsson. Intrinsic image decomposition for feature extraction of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2666–2677, May 2015.
- [19] S. Kim, K. Park, K. Sohn, and S. Lin. Unified depth prediction and intrinsic image decomposition from a single image via joint convolutional neural fields. In *ECCV*, pages 143–159, Cham, 2016. Springer International Publishing.
- [20] N. Kong and M. J. Black. Intrinsic depth: Improving depth transfer with intrinsic images. In *ICCV*, December 2015.
- [21] P. Y. Laffont and J. C. Bazin. Intrinsic decomposition of image sequences from local temporal variations. In *ICCV*, pages 433–441, Dec 2015.
- [22] C. Lanaras, E. Baltsavias, and K. Schindler. Hyperspectral super-resolution by coupled spectral unmixing. In *ICCV*, December 2015.
- [23] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, Jan 1971.
- [24] B. Le Saux, N. Yokoya, R. Hänsch, and S. Prasad. 2018 IEEE GRSS Data Fusion Contest: Multimodal Land Use Classification [Technical Committees]. *IEEE geoscience and remote sensing magazine*, 6(1):52–54, Mar. 2018.
- [25] C. Li, K. Zhou, and S. Lin. Intrinsic face image decomposition with human face priors. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *ECCV*, pages 218–233, Cham, 2014. Springer International Publishing.
- [26] Z. Li and N. Snavely. Learning intrinsic image decomposition from watching the world. In *CVPR*, June 2018.
- [27] W. Liao, A. Piuropa, R. Bellens, S. Gautama, and W. Philips. Generalized graph-based fusion of hyperspectral and lidar data using morphological features. *IEEE Geoscience and Remote Sensing Letters*, 12(3):552–556, March 2015.
- [28] E. Shelhamer, J. T. Barron, and T. Darrell. Scene intrinsics and depth from a single image. In *ICCV*, December 2015.
- [29] J. Shen, X. Yang, Y. Jia, and X. Li. Intrinsic images using optimization. In *CVPR*, pages 3481–3487, June 2011.
- [30] L. Shen and C. Yeo. Intrinsic images decomposition using a local and global sparse representation of reflectance. In *CVPR*, pages 697–704, June 2011.
- [31] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9):1459–1472, Sept 2005.