Ground-Based Image Analysis

A tutorial on machine-learning techniques and applications

IMAGE LICENSED BY INGRAM PUBLISHING

Advances in Machine Learning for Remote Sensing and Geosciences

SOUMYABRATA DEV, BIHAN WEN, YEE HUI LEE, AND STEFAN WINKLER

Ground-based whole-sky cameras have opened up new opportunities for monitoring the earth's atmosphere. These cameras are an important complement to satellite images by providing geoscientists with cheaper, faster, and more localized data. The images captured by whole-sky imagers (WSI) can have high spatial and temporal resolution, which is an important prerequisite for applications such as solar energy modeling, cloud attenuation analysis, local weather prediction, and more. Extracting the valuable information from the huge amount of image data by detecting and analyzing the various entities in these images is challenging. However, powerful machine-learning techniques have become available to aid with the image analysis. This article provides a detailed explanation of recent developments in these techniques and their applications in ground-based imaging, aiming to bridge the gap between computer vision and remote sensing with the help of illustrative examples. We demonstrate the advantages of using machine-learning techniques in ground-based image analysis via three primary applications: segmentation, classification, and denoising.

Digital Object Identifier 10.1109/MGRS.2015.2510448
Date of publication: 13 June 2016

0274-6638/16©2016IEEE

TABLE 1. AN OVERVIEW OF VARIOUS GROUND-BASED WSIs AND THEIR INTENDED APPLICATIONS.

APPLICATION	ORGANIZATION	COUNTRY	WSI MODEL
Air traffic control [18]	Campbell Scientific Ltd.	United Kingdom	IR NEC TS9230
Cloud attenuation [21]–[23]	Nanyang Technological University Singapore	Singapore	WAHRSIS
Cloud characterization [13]	Atmospheric Physics Group	Spain	GFAT All-Sky Imager
Cloud classification [12]	Brazilian Institute for Space Research	Brazil	TSI-440
Cloud classification [14]	Laboratory of Atmospheric Physics	Greece	Canon IXUS II with FOV 180°
Cloud macrophysical properties [9]	Pacific Northwest National Laboratory	United States	Hemispheric Sky Imager
Cloud-track wind data monitoring [15]	Laboratoire de Météorologie Dynamique	France	Nikon D100 with FOV 63°
Convection [16]	Creighton University	United States	Digital camera
Radiation balance [17]	Lindenberg Meteorological Observatory	Germany	VIS/NIR 7
Solar-power forecasting [11]	Solar Resource Assessment & Forecasting Laboratory	United States	TSI-440
Weather monitoring [24]	Pacific Northwest National Laboratory	United States	TSI-880
Weather reporting [19]	Ecole Polytechnique Fédérale de Lausanne	Switzerland	Panorama camera

SATELLITES AS A STARTING POINT

Satellite images are commonly used to monitor the earth and analyze its various properties. They provide remote sensing analysts with accurate information about the various earth events. Satellite images are available in different spatial and temporal resolutions and also across various ranges of the electromagnetic spectrum, including visible, near- and far-infrared regions. For example, multitemporal satellite images are extensively used for monitoring forest canopy changes [1] or evaluating sea-ice concentrations [2]. The presence of clouds plays a very important role in the analysis of satellite images. NASA's Ice, Cloud, and Land Elevation Satellite (ICESat) has demonstrated that 70% of the earth's atmosphere is covered with clouds [3]. Therefore, there has been renewed interest among the remote sensing community to further study clouds and their effects on the earth.

Satellite images are a good starting point for monitoring the earth's atmosphere. However, they have either high temporal resolution (e.g., geostationary satellites) or high spatial resolution (e.g., low-orbit satellites) but never both. In many applications, such as solar energy production [4], local weather prediction, tracking contrails at high altitudes [5], studying aerosol properties [6], and the attenuation of communication signals [7], [8], data with high spatial and temporal resolution is needed. This is why ground-based sky imagers have become popular and are now widely used in these and other applications. The ready availability of high-resolution cameras at a low cost facilitated the development of various models of sky imagers.

A WSI consists of an imaging system placed inside a weather-proof enclosure that captures the sky at user-defined intervals. A number of WSI models have been developed over the years, including a commercial WSI [Total Sky Imager (TSI)-440, TSI-880] manufactured by Yankee Environmental Systems that is used by many researchers [9]–[11]. Owing to the high cost and limited flexibility of commercial sky imagers, many research groups have built their own WSI models [12]–[19], e.g., the Scripps Institution of Oceanography at the University of California, San Diego, has been developing and using WSIs as part of its work for many years [20]. Similarly, our group designed the Wide-Angle High-Resolution Sky Imaging System (WAHRSIS) for cloud-monitoring purposes [21]–[23]. Table 1 provides an overview of the types of ground-based sky cameras used by various organizations around the world and their primary applications.

MACHINE LEARNING FOR REMOTE SENSING DATA

The rapid increase in computing power has enabled the use of powerful machine-learning algorithms on large data sets. Remote sensing data fill this description and are typically available in different temporal, spatial, and spectral resolutions. For aerial surveillance and other monitoring purposes, Red-Green-Blue (RGB) images are captured by low-flying aircraft or drones. Multispectral data are used for forest, land, and sea monitoring. Recently, hyperspectral imaging systems with very narrow bands have been employed for identifying specific spectral signatures for agriculture and surveillance applications.

In cloud analysis, one example of such remote sensing data is ground-based images captured by WSIs. With these images, one can monitor the cloud movement and predict the clouds' future location, detect and track contrails, and monitor aerosols. This is important in applications such as cloud attenuation and solar radiation modeling, which require high temporal and spatial resolution data. The requirement for high-resolution data is further exemplified by areas where weather conditions are more localized. Such microclimates are mainly prevalent near bodies of water that may cool the local atmosphere or in heavily urban areas where buildings and roads absorb the sun's energy (Singapore, the authors' home, being a prime example of such conditions). These weather

conditions lead to quicker cloud formation, which can have sudden impacts on signal attenuation or solar radiation. Therefore, high-resolution ground-based imagers are required for continuous and effective monitoring of the earth's atmosphere.

In this article, we show how a number of popular state-of-the-art machine-learning methods can be effectively used in remote sensing in general and ground-based image analysis, in particular. A high-level schematic framework for this is shown in Figure 1.

There are a number of challenges in applying machine-learning techniques in remote sensing. While the high dimensionality of remote sensing data can provide rich information and a complex data model, it is normally expensive and difficult to create a sufficient amount of labeled data for reliable supervised training. Additionally, the influence of atmospheric noise and interference introduces error and variance in the acquired training data. Thus, without effective regularization and feature extraction, overfitting can occur in the learned model that may eventually affect the performance of the method.

Moreover, processing the rich amount of high-dimensional data directly leads to high computational cost and memory requirements, while the large amount of data redundancy fails to facilitate the learning significantly. Therefore, appropriate feature extraction is crucial in machine learning, especially for remote sensing applications. In the "Feature Extraction" section, we discuss some of the most popular types of features, including computer vision features, remote sensing features, dimensionality reduction (DR), and sparse representation features. Instead of the full-dimensional raw input data, these extracted features are used for subsequent analysis in different application domains. Illustrative examples are also provided for these types of features to demonstrate their utility and effectiveness.

Using three primary applications as examples, i.e., segmentation, classification, and denoising, the "Applications" section shows that a learning-based framework can potentially perform better than heuristic approaches. Image segmentation is the task of categorizing pixels into meaningful regions that share similar properties, belong to same group, or form certain objects. Classification is the problem of recognizing objects based on some predefined categories. Denoising estimates the true signals from their corrupted observations. In this article, we show how a number of popular state-of-the-art machine-learning methods can be effectively used in remote sensing, in general, and ground-based image analysis, in particular.

FEATURE EXTRACTION

Effective image features are important for computational efficiency and enhanced performance in different applications. Due to the high dimensionality of the data, it is difficult and inefficient to learn from the raw data directly. Moreover, the effect of collinearity among the input variables and the presence of noise degrade the performance of the algorithms to a great extent. Therefore, discriminative features should be chosen carefully from the input data. It is beyond the scope of this tutorial to encompass and list all existing feature-extraction techniques. We focus on the popular feature extractors that are widely used in the remote sensing community and that show promise for ground-based image analysis. Based on the application domains and the nature of the techniques, four primary categories of feature-extraction techniques are distinguished in this article, which will be discussed in more detail:

- computer vision features
- remote-sensing features
- dimensionality reduction
- sparse representation features.

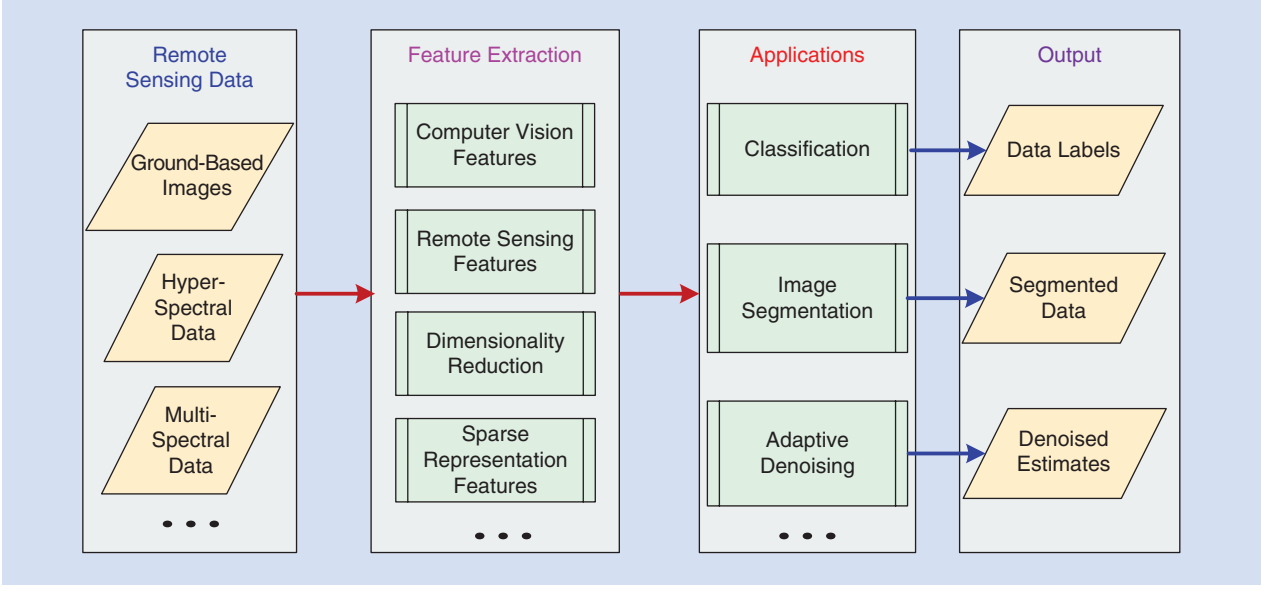

FIGURE 1. A high-level schematic framework of remote sensing data analysis with machine-learning techniques.

COMPUTER VISION FEATURES

Traditional computer vision feature-extraction techniques mainly consist of corner and edge detectors. The term *corner* has varied interpretations. Essentially, a corner denotes a region where there is a sharp variation in brightness. These corner points may not always represent the projection of a three-dimensional (3-D) corner point in the image. In an ideal scenario, the feature detector should detect the same set of corners under any affine transformation of the input images. The most commonly used algorithm is the Harris corner detector [25], which relies on a small window that slides across the image and looks for variations of intensity changes. In automatic satellite image registration, Harris corner detection has been used to extract feature points from buildings and natural terrain [26], [27].

Aside from corners, blobs are also popular discriminatory features. Blobs are small image regions that possess similar characteristics with respect to color and intensity. Some popular blob detectors are the difference of Gaussians (DoG), scale-invariant feature transform (SIFT) [28], and speeded-up robust features (SURF) [29]. These feature descriptors have a high invariability to affine transformations such as rotation. DoG is a bandpass filter that involves the subtraction of two blurred versions of the input image. These blurred versions are obtained by convolving the image with two Gaussian filters of different standard deviations. Due to its attractive property to enhance information at certain frequency ranges, DoG can be used to separate the specular reflection from ground-penetrating radar images [30], which is necessary for the detection of landmines using radar images. DoG also has wide applications in obtaining pan-sharpened images that have high spectral and spatial resolutions [31].

SIFT and SURF are two other very popular blob-based feature-extraction techniques in computer vision that are widely used in remote sensing analysis. SIFT extracts a set of feature vectors from an image that is invariant to rotation, scaling, and translation. They are obtained by detecting extrema in a series of sampled and smoothed versions of the input image. SIFT is mainly applied to the task of image registration in optical remote sensing images [32] and multispectral images [33]. Unlike SIFT, SURF uses integral images to detect the feature points in the input image. SURF's main advantage is its faster execution as compared to SIFT. Image matching on Quickbird images is done using SURF features [34]; Song et al. [35] proposed a robust retrofitted SURF algorithm for remote sensing image registration. These corner and blob detectors are essentially local features, i.e., they have a spatial interpretation, exhibiting similar properties of color, texture, and position in their neighborhood [36]. These local features can help to retain the local information of the image and provide cues for applications, such as image retrieval and image mining.

In addition to corner and blob detectors, local features based on image segmentation are also popular, where the entire image is divided into several subimages by considering the boundaries between different objects in the image. The purpose of segmentation-based features is to find homogeneous regions of the image that can subsequently be used in an image segmentation framework.

Pixel-grouping techniques group pixels with a similar appearance. Popular approaches such as the superpixel method [37] have also been applied for remote sensing image classification. Recently, Vargas et al. [38] presented a bag-of-words model using superpixels for multispectral image classification, and Zhang et al. [39] use superpixel-based feature extraction in aerial image classification. Another popular technique of pixel grouping is graph-based image representation, where pixels with similar properties are connected by edges. Graph-theoretic models allow for encoding the local segmentation cues in an elegant and systematic framework of nodes and edges. The segmented image is obtained by cutting the graph into subgraphs, such that the similarity of pixels within a subgraph is maximized. A good review of the various graph-theoretical models in computer vision is provided by Shokoufandeh and Dickinson [40].

To illustrate the corner and blob detector features in the context of ground-based image analysis, we provide an illustrative example by considering a sample image from the hybrid thresholding algorithm (HYTA) database [41], where the original image is scaled by a factor of 1.3 and rotated by 30°. Figure 2 shows candidate matches between the input

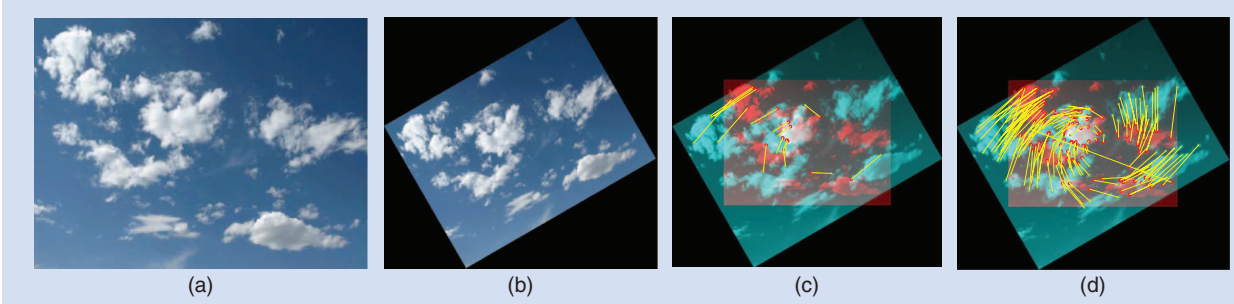

FIGURE 2. Feature matching between (a) the original image and (b) the transformed image that is scaled by a factor of 1.3 and rotated by 30°. (c) The candidate matches using the Harris corner detector. (d) The candidate matches using the SURF detector.

and transformed image using the Harris corner detector and SURF. As clouds do not possess strong edges, the number of detected feature points using the Harris corner detector is far lower than that of the SURF detector. Furthermore, the repeatability of the SURF detector is higher than the corner detector for the same amount of scaling and rotation.

REMOTE SENSING FEATURES

In remote sensing, hand-crafted features exploiting the characteristics of the input data are widely used for image classification [42], involving the generation of a large number of features that capture the discriminating cues in the data. The user makes an educated guess about the most appropriate features. Unlike the popular computer vision feature-extraction techniques given previously, remote sensing features use their inherent spectral and spatial characteristics to identify discriminating cues of the input data. They are not learning based but are derived empirically from the input data and achieve good results in certain applications.

Heinle et al. [43] proposed a 12-dimensional feature vector that captures color, edge, and texture information of a sky/cloud image, which is quite popular in cloud classification. The raw intensity values of RGB aerial images have also been used as input features [44]. In satellite imagery, the normalized difference vegetation index is used in association with the raw pixel intensity values for monitoring land cover, road structures, and so on [45]. In high-resolution aerial images, neighboring pixels are considered for the generation of feature vectors, which results in the creation of, e.g., 3 × 3, 15 × 15, and 21 × 21 pixel neighborhoods. Furthermore, to encode the textural features of the input images, Gabor- and edge-based texture filters are used, e.g., for aerial imagery [46] or landscape image segmentation [47]. Recently, we have used a modified set of Schmid filters for the task of cloud classification [48].

DIMENSIONALITY REDUCTION

Remote sensing data are high-dimensional in nature. Therefore, it is advisable to reduce the inherent dimensionality of the data considerably while capturing sufficient information in the reduced subspace for further data processing. In this section, we discuss several popular DR techniques and point to relevant remote sensing applications. A more detailed review of various DR techniques can be found in [49].

Broadly speaking, DR techniques can be classified as either linear or nonlinear. Linear DR methods represent the original data in a lower-dimensional subspace by a linear transformation, while nonlinear methods consider the nonlinear relationship between the original data and the features. This article focuses on linear DR techniques because of their lower computational complexity and simple geometric interpretation. A brief overview of the different techniques is provided in Table 2, and a detailed treatment of the various methods can be found in [50]. We denote the data as $\mathbf{X} = [x_1 \mid x_2 \mid \ \ldots \ \mid x_n] \in \mathbb{R}^{N \times n}$, where each $x_i \in \mathbb{R}^N$ represents a vectorized data point, N denotes the data dimensionality, and n is the data size. The corresponding features are denoted as $\mathbf{Z} = [z_1 \mid z_2 \mid \ \ldots \ \mid z_n] \in \mathbb{R}^{K \times n}$, where each $z_i \in \mathbb{R}^K$ is the feature representation of x_i, and K denotes the feature dimensionality.

TABLE 2. A SUMMARY OF LINEAR DR TECHNIQUES.

TECHNIQUES	MAXIMIZED OBJECTIVES	SUPERVISED	CONVEX
PCA	Data variance	No	Yes
FA	Likelihood function of underlying distribution parameters	No	No
LDA	Between-class variability over within-class variability	Yes	Yes
NCA	Stochastic variant of the leave one out score	Yes	No

Principal component analysis (PCA) is one of the most common and widely used DR techniques, which projects the N-dimensional data $\mathbf{X}$ onto a lower K-dimensional (i.e., $K \leq N$) feature space as $\mathbf{Z}$ by maximizing the captured data variance or, equivalently, minimizing the reconstruction error. PCA can be represented as

$$\mathbf{Z} = \mathbf{U}^T\mathbf{X}, \tag{1}$$

where $\mathbf{U} \in \mathbb{R}^{N \times K}$ is formed by the principal components that are orthonormal and can be obtained from the eigenvalue decomposition of the data covariance matrix. The objective function is convex, thus convergence and global optimality are guaranteed. In the field of remote sensing, PCA is often used to reduce the number of bands in multispectral and hyperspectral data, and it is also widely used for change detection in forest fires and land-cover studies. Munyati [51] used PCA as a change-detection technique in inland wetland systems using Landsat images, observing that most of the variance was captured in the near-infrared reflectance. Subsequently, the image composite obtained from the principal axes was used in change detection.

Factor analysis (FA) is based on the assumption that the input data $\mathbf{X}$ can be explained by a set of underlying factors. These factors are relatively independent of each other and are used to approximately describe the original data. The input data $\mathbf{X}$ can be expressed as a linear combination of K factors with small independent errors $\mathbf{E}$:

$$\mathbf{X} = \sum_{i=1}^{K} F_i Z_i + \mathbf{E}, \tag{2}$$

where $\{F_i\}_{i=1}^{K} \in \mathbb{R}^N$ are the different derived factors, and $\mathbf{Z}_i$ denotes the ith row of the feature matrix $\mathbf{Z}$. The error matrix $\mathbf{E}$ explains the variance that cannot be expressed by any of the underlying factors. The factors $\{F_i\}_{i=1}^{K}$ can be found by maximizing the likelihood function of the underlying

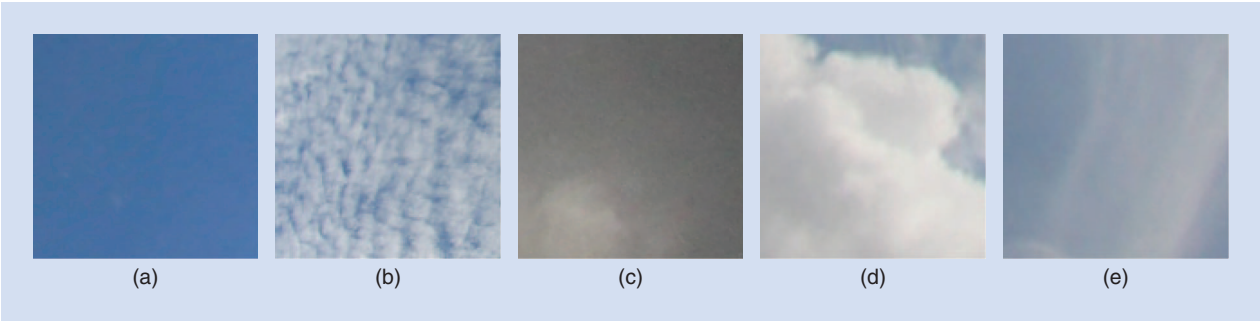

FIGURE 3. Categories for sky/cloud image patches in SWIMCAT: (a) clear sky, (b) patterned clouds, (c) thick dark clouds, (d) thick white clouds, and (e) veil clouds.

distribution parameters. To our knowledge, there is no algorithm with a closed-form solution to this problem. Thus, expectation maximization is normally used, but it offers no performance guarantee due to the nonconvex problem formulation. In remote sensing, FA is used in aerial photography and ground surveys. Doerffer and Murphy [52] have used FA techniques in multispectral data to extract latent and meaningful within-pixel information.

Unlike PCA and FA, which are unsupervised (i.e., using unlabeled data only), linear discriminant analysis (LDA) is a supervised learning technique that uses training data class labels to maximize class separability. Given all training data $\mathbf{X}$ from p classes, the mean of the jth class C_j is denoted as μ_j, and the overall mean is denoted as μ. We define the within-class covariance matrix $\mathbf{S}_W$ as

$$\mathbf{S}_W = \sum_{j=1}^{p} \sum_{i \in C_j} (x_i - \mu_j)(x_i - \mu_j)^T \tag{3}$$

and the between-class covariance matrix $\mathbf{S}_B$ as

$$\mathbf{S}_B = \sum_{j=1}^{p} (\mu_j - \mu)(\mu_j - \mu)^T. \tag{4}$$

Thus, the maximum separability can be achieved by maximizing the between-class variability over within-class variability over the desired linear transform W as

$$\max_{\mathbf{W}} \frac{\mathrm{tr}\{\mathbf{W}\mathbf{S}_B\mathbf{W}^T\}}{\mathrm{tr}\{\mathbf{W}\mathbf{S}_W\mathbf{W}^T\}}, \tag{5}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of the matrix. The solution provides the linear DR mapping $\mathbf{W}$ that is used to produce LDA feature $\mathbf{Z} = \mathbf{WX}$.

LDA is widely used for the classification of hyperspectral images. In such cases, the ratio of the number of training labeled images to the number of spectral features is small because labeled data are expensive, and it is difficult to collect a large number of training samples. For such scenarios, Bandos et al. [53] used regularized LDA in the context of hyperspectral image classification. Du and Nekovel [54] proposed a constrained LDA for efficient real-time hyperspectral image classification.

Finally, neighborhood component analysis (NCA) was introduced by Goldberger et al. [55]. Using a linear transform $\mathbf{A}$, NCA aims to find a feature space such that the average leave-one-out k-nearest neighbor (k-NN) score in the transformed space is maximized. It can be represented as

$$\mathbf{Z} = \mathbf{AX}. \tag{6}$$

NCA aims to reduce the input dimensionality N by learning the transform $\mathbf{A}$ from the data set with the help of a differentiable cost function for $\mathbf{A}$ [55]. However, this cost function is nonconvex in nature, and, thus, the solution obtained may be suboptimal.

The transform $\mathbf{A}$ is estimated using a stochastic neighbor selection rule. Unlike the conventional k-NN classifier that estimates the labels using a majority voting of the nearest neighbors, NCA randomly selects neighbors and calculates the expected vote for each class. This stochastic neighbor selection rule is applied as follows. Each point i selects another point as its neighbor j with the following probability:

$$p_{ij} = \frac{e^{-d_{ij}}}{\sum_{k \neq i} e^{-d_{ik}}}, \tag{7}$$

where d_{ij} is the distance between points i and j, and $p_{ii} = 0$. NCA is used in remote sensing for the classification of hyperspectral images. Weizman and Goldberger [56] demonstrated the superior performance of NCA in the context of images obtained from an airborne visible/infrared imaging spectroradiometer.

We will now illustrate the effect of different DR techniques in the context of ground-based cloud classification. For this purpose, we will use the recently released cloud categorization database called *Singapore Whole-Sky Imaging Categories database* (*SWIMCAT*) [48]. Cloud types are properly documented by the World Meteorological Organization [57]. The SWIMCAT database consists of 784 sky/cloud image patches divided into five visually distinctive categories: clear sky, patterned clouds, thick dark clouds, thick white clouds, and veil clouds. Sample images from each category are shown in Figure 3. (SWIMCAT can be downloaded from http://vintage.winklerbros.net/swimcat.html.)

We extract the 12-dimensional Heinle feature (see the "Remote Sensing Features" section) for each image,

and we randomly select 50 images from each of the five cloud categories. For easier computation, the images are downsampled to a resolution of 32 × 32 pixels using bicubic interpolation. Once the feature vectors are generated, the previously mentioned linear DR techniques, PCA, FA, LDA, or NCA, are applied on the entire input feature space.

Figure 4 visualizes the results obtained with the different techniques. The original high-dimensional feature vector is projected onto the primary two principal axes. The different cloud categories are denoted with different colors. We observe that PCA essentially separates the various cloud categories, but veil clouds are scattered in a random manner. PCA and FA are often confused with one another, as they attempt to express the input variables in terms of latent variables. However, it is important to note that they are distinct methods based on different underlying philosophies, which is exemplified by the results shown in Figure 4. The separation of features in LDA is relatively good as compared to PCA and FA, because LDA aims to increase class separability in addition to capturing the maximum variance. NCA also separates the different classes quite well. To further quantify this separability of different classes in the transformed domain, we will present a quantitative analysis in the "Image Classification" section.

SPARSE REPRESENTATION FEATURES

Features based on sparse representation have been widely studied and used in signal processing and computer vision. Different from DR that provides effective representation in a lower-dimensional subspace, adaptive sparse representation learns a union of subspaces for the data. Compared to fixed sparse models such as the discrete cosine transform (DCT) or wavelets, adaptively learned sparse representation provides improved sparsity and usually serves as a better discriminator in various tasks such as face recognition [58], image segmentation [59], object classification [60], and denoising [61], [62]. Learning-based sparse representation also demonstrates advantages in remote sensing problems such as image fusion [63] and hyperspectral image classification [64].

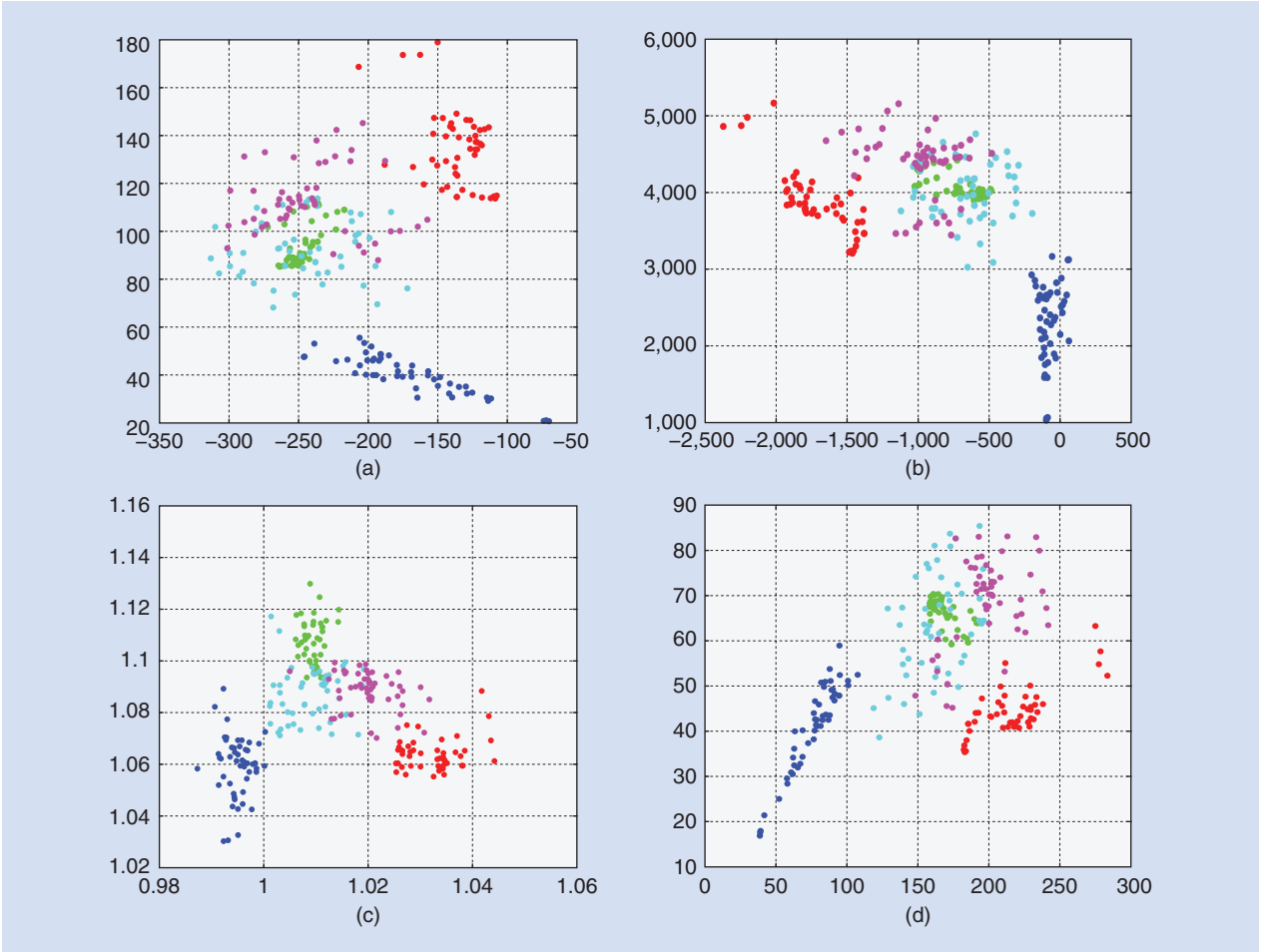

FIGURE 4. A visualization of the results from applying four DR techniques on the SWIMCAT dataset [48]: (a) PCA, (b) FA, (c) LDA, and (d) NCA. The data are reduced from their original 12-dimensional feature space to two dimensions in the projected feature space for a five-class cloud classification problem. The different colors indicate individual cloud classes (i.e., red: clear sky; green: patterned clouds; blue: thick dark clouds; cyan: thick white clouds; magenta: veil clouds).

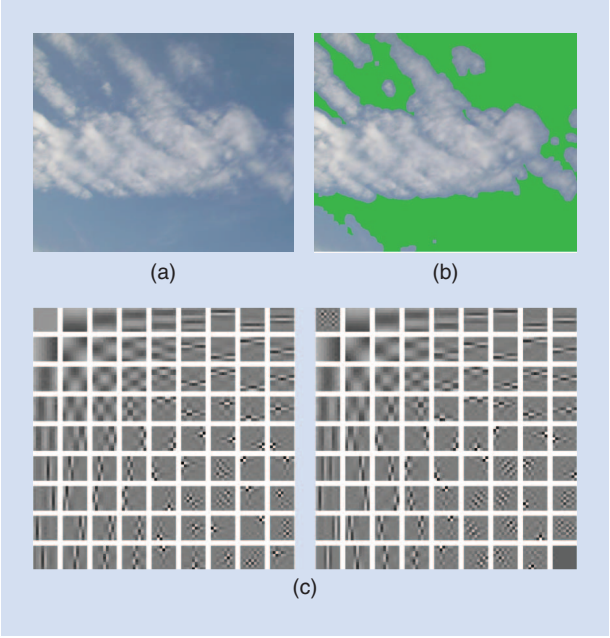

FIGURE 5. Cloud and sky segmentation via learning OCTOBOS sparse representation: (a) the original image, (b) the input image with original pixels clustered as cloud and green pixels clustered as sky, and (c) the learned two-class OCTOBOS with each row visualized as patches in separate blocks.

Several models for sparsity have been proposed in recent years, with the most popular being the synthesis model [61], which suggests that a set of data $\mathbf{X}$ can be modeled by a common matrix $\mathbf{D} \in \mathbb{R}^{N \times K}$ and their respective sparse codes $\mathbf{Z}$:

$$\mathbf{X} = \mathbf{DZ}, s.t. \| z_i \|_0 \leq s \ll K \; \forall i, \tag{8}$$

where $\| \cdot \|_0$ counts the number of nonzeros, which is upper bounded by the sparsity level s. The codes $\{z_i\}_{i=1}^n$ are sparse, meaning that the maximum number of nonzeros s is much smaller than the code dimensionality K. The matrix $\mathbf{D} = [d_1 \,|\, d_2 \,|\, \ldots \,|\, d_K]$ is the synthesis dictionary, with each d_j called an *atom*. This formulation implies that each x_i can be decomposed as a linear combination of only s atoms. For a particular x_i, the selected s atoms also form its basis. In other words, data that satisfies such a sparse model live in a union of subspaces spanned by only a small number of selected atoms of $\mathbf{D}$ due to sparsity. The generalized synthesis model allows for small modeling errors in the data space, which is normally more practical [58], [61].

Given data $\mathbf{X}$, finding the optimal dictionary is well known as the synthesis dictionary learning problem. Since the problem is normally nonconvex, and finding the exact solution is nondeterministic polynomial-time (NP)-hard, various approximate methods have been proposed and have demonstrated good empirical performance. Among those, the K-singular value decomposition (SVD) algorithm [61] has become very popular due to its simplicity and efficiency. For a given $\mathbf{X}$, the K-SVD algorithm seeks to solve the following optimization problem:

$$\min_{\mathbf{D},\mathbf{Z}} \| \mathbf{X} - \mathbf{DZ} \|_F^2 \, s.t. \| z_i \|_0 \leq s \; \forall \boldsymbol{i}, \| d_j \|_2 = 1 \forall j, \tag{9}$$

where $\| \mathbf{X} - \mathbf{DZ} \|_F^2$ represents the modeling error in the original data domain. To solve this joint minimization problem, the algorithm alternates between sparse coding (solving for $\mathbf{Z}$, with fixed $\mathbf{D}$) and dictionary update (solving for $\mathbf{D}$, with fixed $\mathbf{Z}$) steps. K-SVD adopts orthogonal matching pursuit [65] for sparse coding and updates the dictionary atoms sequentially, while fixing the support of the corresponding $\mathbf{Z}$ component by using SVD.

Besides the synthesis dictionary learning, there are learning algorithms associated with other models, such as transform learning [66]. Unlike synthesis dictionary learning, which is normally sensitive to initialization, the transform learning scheme generalizes the use of conventional analytical transforms, such as DCT or wavelets, to a regularized adaptive transform $\mathbf{W}$ as follows:

$$\min_{\mathbf{W},\mathbf{Z}} \| \mathbf{WX} - \mathbf{Z} \|_F^2 + \nu(\mathbf{W}) s.t. \| z_i \|_0 \leq s \forall i, \tag{10}$$

where $\| \mathbf{WX} - \mathbf{Z} \|_F^2$ denotes the modeling error in the adaptive transform domain. Function $\nu(\,.\,)$ is the regularizer for $\mathbf{W}$ [66] to prevent trivial and badly conditioned solutions. The corresponding algorithm [62], [66] provides exact sparse coding and a closed-form transform update with lower complexity and faster convergence, compared to the popular K-SVD.

In sparse representation, the sparse codes are commonly used as features for various tasks such as image reconstruction and denoising. More sophisticated learning formulations also include the learned models (i.e., dictionaries or transforms) as features for applications such as segmentation and classification. Figure 5 provides a simple cloud/sky image segmentation example using the overcomplete sparsifying transform model with block cosparsity (OCTOBOS) [62], which learns a union of sparsifying transforms, to illustrate and visualize the usefulness of sparse features. We extract 9×9 overlapping image patches from the ground-based sky image shown in Figure 5(a). The color patches are converted to gray scale and vectorized to form the 81-dimensional data vectors. The OCTOBOS algorithm simultaneously learns a union of two transforms, generates the sparse codes, and clusters the image patches into two classes (i.e., sky class and cloud class) by comparing the modeling errors [67]. Since the overlapping patches are used, each pixel in the image typically belongs to multiple extracted patches. We cluster a pixel into a particular class by majority voting. The image segmentation result, with pixels belonging to the sky class, is visualized in Figure 5(b). In the learning stage, we restrict the sparsity of each vector to be, at most, ten out of 81. The distinct sparsifiers, or rows of learned OCTOBOS, are visualized as 9×9 patches

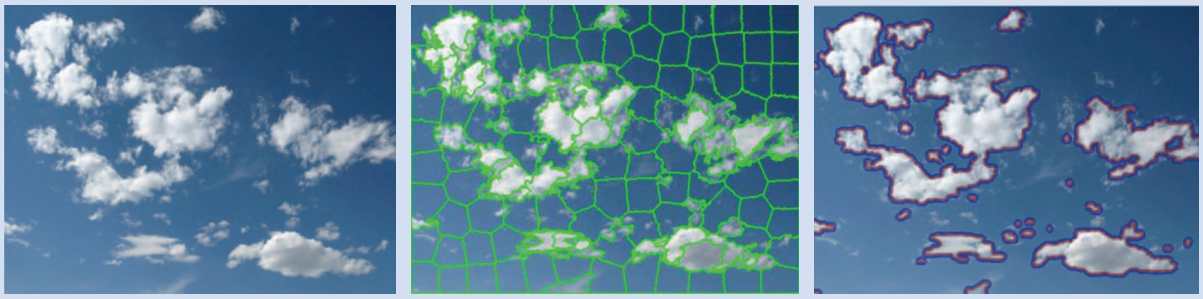

FIGURE 6. Sky/cloud image segmentation using two methods, superpixels and graph cut. (a) A sample image from the HYTA database. (b) An oversegmented image with superpixels. (c) An image segmented using graph cut.

in the blocks in Figure 5(c). Both the sparse codes and the learned transform blocks are used as features for clustering in this example. It is important to note that we did not use any other remote sensing features on top of the OCTOBOS clustering scheme [67]. A hybrid version that combines this with cloud-specific features [68] may further enhance the segmentation performance.

APPLICATIONS

In this section, we present applications of the techniques discussed in the previous section for ground-based sky/cloud image analysis and show experimental results. We focus on three main applications: segmentation, classification, and denoising. We also show that data-driven machine-learning techniques generally outperform conventional heuristic approaches.

IMAGE SEGMENTATION

Image segmentation refers to the task of dividing an image into several segments in an attempt to identify different objects in the image. The problem of image segmentation has been extensively studied in remote sensing for several decades. In the context of ground-based image analysis, image segmentation refers to the segmentation of sky/cloud images obtained by sky cameras. Cloud segmentation is challenging because of the clouds' nonrigid structure and the high degree of variability in sky illumination conditions. In this section, we will provide illustrative examples of several sky/cloud image segmentation methodologies.

Liu et al. [69] use superpixels to identify local homogeneous regions of sky and cloud. Figure 6 illustrates the oversegmented superpixel image of a sky/cloud image from the HYTA database [41]. The generated superpixels respect the image boundaries quite well and are consistent based on the texture and color of sky and cloud regions, respectively. These local regions can thus be used for subsequent machine-learning tasks. The final sky/cloud binary image can be obtained by thresholding this oversegmented image using a threshold matrix [69]. In addition to superpixels, graph-cut-based techniques [70], [71] have also been explored in ground-based image analysis. Liu et al. [72] proposed an automatic graph-cut technique in identifying sky/cloud regions. Figure 6(c) shows the two-level segmented output using automatic graph cut. As clouds do not have any specific shape and cloud boundaries are ill-defined, several approaches have been proposed that use color as a discriminatory feature. The segmentation can be binary [10], [41], multilevel [73], or probabilistic [68]. As an illustration, we show these three cases for a sample image of the HYTA data set. Figure 7(a) shows the binary segmentation of a sample input image from the HYTA database [41]. The process involves thresholding the selected color channel.

Coupled with such binary approaches, a multilevel output image can also be generated. Machine-learning techniques involving Gaussian discriminant analysis can be used for such purposes. In [73], a set of labeled training data is used for a-priori learning of the latent distribution of three labels (clear sky, thin clouds, and thick clouds). We illustrate such three-level semantic labels of the sky/cloud image in Figure 7(b). In addition to two-level and three-level output images, a probabilistic segmentation approach is exploited in [68], wherein each pixel is assigned a confidence value of belonging to the cloud category, which is illustrated in Figure 7(c).

IMAGE CLASSIFICATION

In the most general sense, classification refers to the task of categorizing the input data into two or more classes. We can distinguish between supervised and unsupervised methods, as the latter identifies underlying latent structures in the input data space and thereby makes appropriate decisions on the corresponding labels. In other words, unsupervised methods cluster pixels with similar properties (e.g., spectral reflectance). Supervised methods, on the other hand, rely on a set of annotated training examples. These training data help the system to learn the distribution of the labeled data in any dimensional feature space. Subsequently, the learned system is used in predicting the labels of unknown data points. In remote sensing, *k*-means, Gaussian mixture models (GMM),

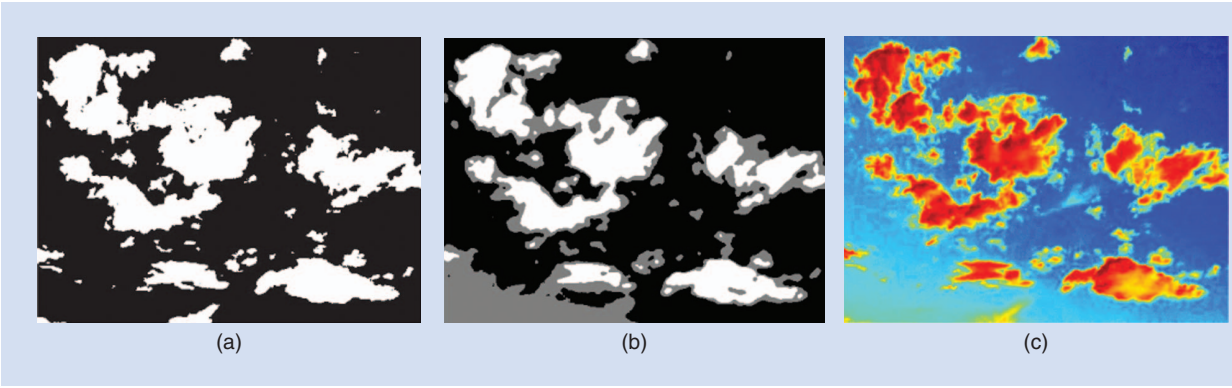

FIGURE 7. Sky/cloud image segmentation: (a) binary (or two-level) segmentation of a sample input image from the HYTA database, (b) three-level semantic segmentation of a sky/cloud image [73], and (c) probabilistic segmentation of a sky/cloud image [68].

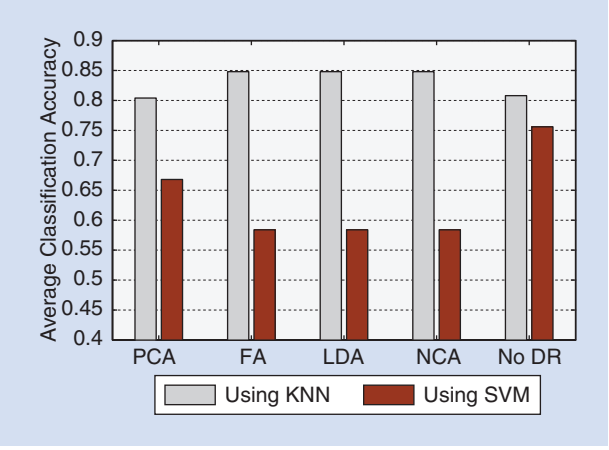

FIGURE 8. The average multiclass classification accuracy using Heinle features for cloud patch categorization for different methods.

and swarm optimization are the most commonly used unsupervised classification, and (clustering) techniques. Ari and Aksoy [74] used GMM and particle swarm optimization for hyperspectral image classification, and Maulik and Saha [75] used a modified differential evolution-based fuzzy clustering algorithm for satellite images. Such clustering techniques are also used in ground-based image analysis.

In addition to supervised and unsupervised methods, semi-supervised learning (SSL) methods are widely used in remote sensing [76]. SSL uses both labeled and unlabeled data in its classification framework, helping to create a robust learning framework that learns the latent marginal distribution of the labels. This is useful in remote sensing, as the availability of labeled data is scarce and manual annotation of data is expensive. One such example is hyperspectral image classification [77]. In addition to SSL methods, models involving sparsity and other regularized approaches are also becoming popular, e.g., Tuia et al. [78] study the use of nonconvex regularization in the context of hyperspectral imaging.

In ground-based image analysis, image classification refers to categorizing sky/cloud types into various kinds, e.g., clear sky, patterned clouds, thick dark clouds, thick white clouds, and veil clouds (see the "Dimensionality Reduction" section). To quantify the accuracy of the separation of data in Figure 4, we use several popular clustering techniques in combination with DR techniques. We use two classifiers for evaluation purposes, i.e., k-NN and the support vector machine (SVM). k-NN is a nonparametric classifier, wherein the output label is estimated using a majority voting of the labels of a neighborhood. The SVM is a parametric method that generates a hyperplane, or a set of hyperplanes, in the vector space by maximizing the margin between classifiers to the nearest neighbor data.

We evaluate five distinct scenarios, 1) PCA, 2) FA, 3) LDA, 4) NCA, and 5) no DR, and report the classification performances of both k-NN and SVM in each of these cases. We again use the SWIMCAT [48] database for evaluation purposes. The training and testing sets consist of random selections of 50 distinct images, all of which are downsampled to 32×32 pixels for faster computation. Using the 50 training images for each of the categories, we compute the corresponding projection matrix for PCA, FA, LDA, and NCA. We use the reduced two-dimensional (2-D) Heinle feature for training a k-NN/SVM classifier for scenarios 1–4. We use the original 12-dimensional vector for training the classifier model for scenario 5. In the testing stage, we obtain the projected 2-D feature points using the computed projection matrix, followed by a k-NN/SVM classifier for classifying the test images into individual categories. The average classification accuracies across the five classes are shown in Figure 8.

The k-NN classifier achieves better performance than the SVM classifier in all of the cases. From the 2-D projected feature space (see Figure 4), it is clear that the data points belonging to an individual category lie close to each other. However, it is difficult to separate the different categories using hyperplanes in 2-D space. We observe that the complexity of the linear SVM classifier is not sufficient to separate the individual classes, as k-NN performs relatively better in this example. Among the different DR techniques, LDA and NCA work best with the k-NN classifier, because these methods also use the

class labels to obtain maximum interclass separability. Moreover, the performance without prior DR performs comparably well. In fact, the SVM classifier provides increasingly better results when the feature space has higher dimensionality, which shows that further applications of DR on top of extracting remote sensing features may not be necessary in a classification framework. Of course, DR significantly reduces the computational complexity.

ADAPTIVE DENOISING

Image and video denoising problems have been heavily studied in the past, with various denoising methods proposed [79]. Denoting the true signal (i.e., clean image or video) as x, the measurement y is usually corrupted by additive noise e as

$$y = x + e. \tag{11}$$

The goal of denoising is to obtain an estimate $\tilde{x}$ from the noisy measurement y such that $\|\tilde{x} - x\|$ is minimized. Denoising is an ill-posed problem. Thus, certain regularizers, including sparsity, underlying distribution, and self-similarity, are commonly used to obtain the best estimate $\tilde{x}$.

The early approaches of denoising used fixed analytical transforms, simple probabilistic models [80], or neighborhood filtering [81]. Recent nonlocal methods such as block-matching and 3-D filtering [82] have been shown to achieve excellent performance by combining some of these conventional approaches. In the field of remote sensing, Liu et al. [83] used partial differential equations for denoising multispectral and hyperspectral images. Yu and Chen [84] introduced the generalized morphological component analysis for denoising satellite images.

Recently, machine-learning-based denoising methods have received increasing interest. Compared to fixed models, adaptive sparse models [61], [66] or probabilistic models [85] have been shown to be more powerful in image reconstruction. The popular sparsity-based methods, such as K-SVD [61] and OCTOBOS [62], were introduced in the "Feature Extraction" section. Besides, adaptive GMM-based denoising [85] also provides a promising performance by learning a GMM from the training data as a regularizer for denoising, especially in denoising images with complicated underlying structures. While these data-driven denoising methods have become popular in recent years, the usefulness of signal-model learning has rarely been explored in remote sensing or ground-based image analysis, which normally generates data with certain unique properties. Data-driven methods can potentially be even more powerful for representing such signals than conventional analytical models.

We now illustrate how various popular learning-based denoising schemes can be applied to ground-based cloud images. The same cloud image from the HYTA database [41] shown in Figure 6(a) is used as an example and serves as ground truth. We synthetically add zero-mean Gaussian noise with $\sigma = 20$ to the clean data. The obtained noisy image has a peak signal-to-noise ratio (PSNR) of 22.1 dB and is shown in Figure 9(a).

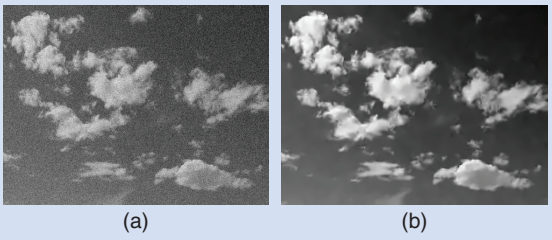

FIGURE 9. The ground-based image denoising result: (a) a noisy cloud image (PSNR = 22.1 dB) and (b) a denoised image (PSNR = 33.5 dB) were obtained using a GMM-based algorithm.

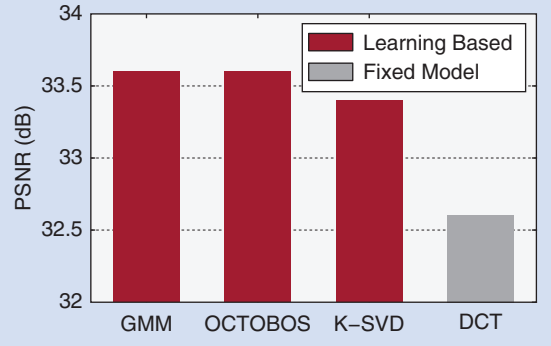

FIGURE 10. The PSNR values for denoising with OCTOBOS, GMM, K-SVD, and DCT dictionary.

Figure 10 provides the denoising performance comparison using several popular learning-based denoising schemes, including GMM [85], OCTOBOS [62], and K-SVD [61]. The quality of the denoised image is measured by PSNR as the objective metric (the clean image has infinite PSNR value). As a comparison, we also include the denoising PSNR by applying a fixed overcomplete DCT dictionary [61]. DCT is an analytical transform commonly used in image compression. For a fair comparison, we maintain the same sparse model richness by using a 256×64 transform in OCTOBOS and 64×256 dictionaries in the K-SVD and DCT methods. For GMM, we follow the default settings in the publicly available software [85].

As illustrated in Figure 10, learning-based denoising methods clearly provide better denoised PSNRs than the DCT-based method, with an average improvement of 1.0 dB. Among all of the learning-based denoising algorithms, K-SVD and OCTOBOS are unsupervised learning methods using image sparsity. In addition, OCTOBOS features a clustering procedure to learn a structured overcomplete sparse model. GMM is a supervised learning method that is pretrained with a standard image corpus. In our experiment, OCTOBOS and GMM perform slightly better than

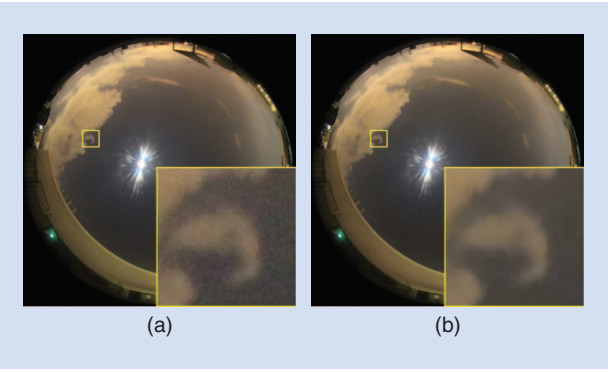

FIGURE 11. The real large-scale night-time cloud/sky image denoising result, with a regional zoom-in for comparison: (a) a real noisy cloud image and (b) a denoised image obtained using an online transform learning-based denoising scheme [88].

K-SVD since they are using either a more complicated model or supervised learning. The denoising result using the GMM-based method is shown in Figure 9(b).

WSIs continuously generate large-scale cloud image data that need to be processed efficiently. Although learning-based algorithms can provide a promising performance in applications such as denoising, most of them are batch algorithms. Consequently, the storage requirements of batch methods such as K-SVD and OCTOBOS increase with the size of the data set; furthermore, processing real-time data in batch mode translates to latency. Thus, online versions of learning-based methods [86], [87] are needed to process high-resolution WSI data. These online learning schemes are more scalable to big-data problems by taking advantage of stochastic learning techniques.

Here, we show an example of denoising a color image measurement of 3,000 × 3,000 pixels that is generated by WAHRSIS at night using online transform learning [88]. The denoising results are illustrated in Figure 11. It is important to note that such a method is also capable of processing real-time high-dimensional data [89]. Thus, it can easily be extended to applications involving multitemporal satellite images and multispectral data in remote sensing.

CONCLUSION

In this article, we have provided an overview of recent developments in machine learning for remote sensing, using examples from ground-based image analysis. Sensing the earth's atmosphere using high-resolution ground-based sky cameras provides a cheaper, faster, and more localized manner of data acquisition. Due to the inherent high-dimensionality of the data, it is expensive to directly use raw data for analysis. We have introduced several feature-extraction techniques and demonstrated their properties using illustrative examples. We have also provided extensive experimental results in segmentation, classification, and denoising of sky/cloud images. Several techniques from machine learning and computer vision communities have been adapted to the field of remote sensing and often outperform conventional heuristic approaches.

ACKNOWLEDGMENT

This work is supported by a grant from Singapore's Defense Science and Technology Agency. B. Wen and S. Winkler are also supported by the research grant for the Human-Centered Cyber-Physical Systems Program at the Advanced Digital Sciences Center from Singapore's Agency for Science, Technology, and Research, A*STAR.

AUTHOR INFORMATION

Soumyabrata Dev (soumyabr001@e.ntu.edu.sg) graduated summa cum laude from the National Institute of Technology Silchar, India, with a B.Tech. in electronics and communication engineering in 2010. From 2010 to 2012, he worked with Ericsson as a network engineer. Currently, he is pursuing a Ph.D. degree in the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From August to December 2015, he was a visiting student at the Audiovisual Communication Laboratory, École Polytechnique Fédérale de Lausanne, Switzerland. His research interests include remote sensing, statistical image processing, and machine learning. He is a Student Member of the IEEE.

Bihan Wen (bwen3@illinois.edu) received a B.Eng. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2012 and an M.S. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign in 2015. He is currently pursuing a Ph.D. degree at the University of Illinois at Urbana–Champaign. His current research interests include signal and image processing, machine learning, sparse representation, and big data applications. He is a Student Member of the IEEE.

Yee Hui Lee (EYHLee@ntu.edu.sg) received the B.Eng. (honors) and M.Eng. degrees from the School of Electrical and Electronics Engineering at Nanyang Technological University, Singapore, in 1996 and 1998, respectively, and a Ph.D. degree from the University of York, United Kingdom, in 2002. Currently, she is an associate professor and assistant chair (students) at the School of Electrical and Electronic Engineering, Nanyang Technological University, where she has been a faculty member since 2002. Her research interests include channel characterization, rain propagation, antenna design, electromagnetic bandgap structures, and evolutionary techniques. She is a Senior Member of the IEEE.

Stefan Winkler (stefan.winkler@adsc.com.sg) received B.Eng. and M.Eng. degrees from the University of Technology Vienna, Austria, and the Ph.D. degree from the École Polytechnique Fédérale de Lausanne, Switzerland, in 1996 and 2001, respecitively. He cofounded Genista, worked in several large corporations in Europe and the United States, and held faculty positions at the National University of Singapore and the University of Lausanne, Switzerland. He is currently a principal scientist and director of the Video and Analytics Program at the University of Illinois' Advanced Digital Sciences Center in Singapore. He has published over 100 papers and the book *Digital Video Quality.*

He is an associate editor of *IEEE Transactions on Image Processing*; a member of the Image, Video, and Multidimensional Signal Processing Technical Committee of the IEEE Signal Processing Society; and chair of the IEEE Singapore Signal Processing Chapter. His research interests include video processing, computer vision, perception, and human-computer interaction. He is a Senior Member of the IEEE.

REFERENCES

[1] J. B. Collins and C. E. Woodcock, "An assessment of several linear change detection techniques for mapping forest mortality using multitemporal landsat TM data," *Remote Sens. Environ.*, vol. 56, no. 1, pp. 66–77, Apr. 1996.

[2] H. Lee and H. Han, "Evaluation of SSM/I and AMSR-E sea ice concentrations in the Antarctic spring using KOMPSAT-1 EOC images," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 7, pp. 1905–1912, July 2008.

[3] D. Wylie, E. Eloranta, J. D. Spinhirne, and S. P. Palm, "A comparison of cloud cover statistics from the GLAS Lidar with HIRS," *J. Climate*, vol. 20, no. 19, pp. 4968–4981, Oct. 2007.

[4] C. L. Fua and H.-Y. Cheng, "Predicting solar irradiance with all-sky image features via regression," *Solar Energy*, vol. 97, pp. 537–550, Nov. 2013.

[5] U. Schumann, R. Hempel, H. Flentje, M. Garhammer, K. Graf, S. Kox, H. Losslein, and B. Mayer, "Contrail study with ground-based cameras," *Atmos. Measurement Techniques*, vol. 6, no. 12, pp. 3597–3612, Dec. 2013.

[6] A. Chatterjee, A. M. Michalak, R. A. Kahn, S. R. Paradise, A. J. Braverman, and C. E. Miller, "A geostatistical data fusion technique for merging remote sensing and ground-based observations of aerosol optical thickness," *J. Geophys. Res.*, vol. 115, no. D20, Oct. 2010.

[7] F. Yuan, Y. H. Lee, and Y. S. Meng, "Comparison of cloud models for propagation studies in Ka-band satellite applications," in *Proc. IEEE Int. Symp. Antennas and Propagation*, Kaohsiung, Taiwan, 2014, pp. 383–384.

[8] F. Yuan, Y. H. Lee, and Y. S. Meng, "Comparison of radio-sounding profiles for cloud attenuation analysis in the tropical region," in *Proc. IEEE Int. Symp. Antennas and Propagation Society*, Memphis, TN, 2014, pp. 259–260.

[9] C. N. Long, J. M. Sabburg, J. Calbó, and D. Pages, "Retrieving cloud characteristics from ground-based daytime color all-sky images," *J. Atmos. Ocean. Tech.*, vol. 23, no. 5, pp. 633–652, May 2006.

[10] M. P. Souza-Echer, E. B. Pereira, L. S. Bins, and M. A. R. Andrade, "A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera," *J. Atmos. Oceanic Tech.*, vol. 23, no. 3, pp. 437–447, Mar. 2006.

[11] M. S. Ghonima, B. Urquhart, C. W. Chow, J. E. Shields, A. Cazorla, and J. Kleissl, "A method for cloud detection and opacity classification based on ground based sky imagery," *Atmos. Measurement Techniques*, vol. 5, no. 11, pp. 2881–2892, Nov. 2012.

[12] S. L. Mantelli Neto, A. von Wangenheim, E. B. Pereira, and E. Comunello, "The use of Euclidean geometric distance on RGB color space for the classification of sky and cloud patterns," *J. Atmos. Oceanic Tech.*, vol. 27, no. 9, pp. 1504–1517, Sept. 2010.

[13] A. Cazorla, F. J. Olmo, and L. Alados-Arboledas, "Development of a sky imager for cloud cover assessment," *J. Opt. Soc. Amer. A*, vol. 25, no. 1, pp. 29–39, Jan. 2008.

[14] A. Kazantzidis, P. Tzoumanikas, A. F. Bais, S. Fotopoulos, and G. Economou, "Cloud detection and classification with the use of whole-sky ground-based images," *Atmos. Res.*, vol. 113, pp. 80–88, Sept. 2012.

[15] F. Abdi, H. R. Khalesifard, and P. H. Falmant, "Small scale cloud dynamics as studied by synergism of time lapsed digital camera and elastic LIDAR," presented at the Int. Laser Radar Conf., Nara, Japan, 2006.

[16] J. A. Zehnder, J. Hu, and A. Razdan, "A stereo photogrammetric technique applied to orographic convection," *Monthly Weather Rev.*, vol. 135, no. 6, pp. 2265–2277, June 2007.

[17] U. Feister, J. Shields, M. Karr, R. Johnson, K. Dehne, and M. Woldt, "Ground-based cloud images and sky radiances in the visible and near infrared region from whole sky imager measurements," in *Proc. Climate Monitoring Satellite Application Facility Training Workshop*, Dresden, Germany, 2000, pp. 20–22.

[18] E. Rumi, D. Kerr, J. M. Coupland, A. P. Sandford, and M. J. Brettle, "Automated cloud classification using a ground based infrared camera and texture analysis techniques," in *Proc. SPIE Remote Sensing of Clouds and the Atmosphere*, Dresden, Germany, 2013, vol. 8890.

[19] Z. Chen, Y. Feng, A. Lindner, G. Barrenetxea, and M. Vetterli, "How is the weather: Automatic inference from images," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Orlando, FL, 2012, pp. 1853–1856.

[20] J. E. Shields, M. E. Karr, R. W. Johnson, and A. R. Burden, "Day/night whole sky imagers for 24-h cloud and sky assessment: History and overview," *Appl. Optics*, vol. 52, no. 8, pp. 1605–1616, Mar. 2013.

[21] S. Dev, F. Savoy, Y. H. Lee, and S. Winkler, "WAHRSIS: A low-cost, high-resolution whole sky imager with near-infrared capabilities," in *Proc. IS&T/SPIE Infrared Imaging Systems*, Baltimore, MD, 2014, vol. 9071.

[22] S. Dev, F. Savoy, Y. H. Lee, and S. Winkler, "Design of low-cost, compact and weather-proof whole sky imagers with HDR capabilities," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Milan, Italy, 2015, pp. 5359–5362.

[23] F. M. Savoy, J. C. Lemaitre, S. Dev, Y. H. Lee, and S. Winkler, "Cloud base height estimation using high-resolution whole sky imagers," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Milan, Italy, 2015, pp. 1622–1625.

[24] C. N. Long, D. W. Slater, and T. Tooman, "Total sky imager (TSI) model 880 status and testing results," U.S. Dept. Energy, Washington, D.C., Tech. Rep. DOE/SC-ARM/TR-006, Atmospheric Radiation Measurement (ARM), 2001.

[25] C. Harris and M. Stephens, "A combined corner and edge detector," *in Proc. Fourth Alvey Vision Conf.*, Manchester, United Kingdom, 1988, pp. 147–151.

[26] I. Misra, S. M. Moorthi, D. Dhar, and R. Ramakrishnan, "An automatic satellite image registration technique based on Harris corner detection and Random Sample Consensus (RANSAC) outlier rejection model," in *Proc. Int. Conf. Recent Advances in Information Technology*, Dhanbad, India, 2012, pp. 68–73.

[27] X. Ying, Z. Linjun, L. Xiaobo, and Y. B. Hae, "An automatic registration method for AVHRR remote sensing images," *Int. J. Multimedia Ubiquitous Eng.*, vol. 9, no. 8, pp. 355–366, 2014.

[28] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. Computer Vision (ICCV)*, Kerkyra, Greece, 1999, pp. 1150–1157.

[29] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Computer Vision Image Understanding*, vol. 110, no. 3, pp. 346–359, June 2008.

[30] X. Xiaoyin and E. L. Miller, "Adaptive difference of Gaussians to improve subsurface imagery," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Toronto, Canada, 2002, pp. 3441–3443.

[31] K. P. Upla, M. V. Joshi, and P. P. Gajjar, "Pan-sharpening: Use of difference of Gaussians," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Quebec City, Canada, 2014, pp. 4922–4925.

[32] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale-invariant feature matching for optical remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 11, pp. 4516–4527, Oct. 2011.

[33] Q. Li, G. Wang, J. Liu, and S. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geosci. Remote Sensing Lett.*, vol. 6, no. 2, pp. 287–291, Apr. 2009.

[34] C. Wu, C. Song, D. Chen, and X. Yu, "A remote sensing image matching algorithm based on the feature-extraction," in *Advances in Neural Networks* (Lecture Notes in Computer Science, vol. 7368). Berlin Heidelberg: Springer, 2012, pp. 282–289.

[35] Z. L. Song and J. Zhang, "Remote sensing image registration based on retrofitted SURF algorithm and trajectories generated from lissajous figures," *IEEE Geosci. Remote Sensing Lett.*, vol. 7, no. 3, pp. 491–495, July 2010.

[36] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 8, pp. 1026–1038, Aug. 2002.

[37] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. Int. Conf. Computer Vision (ICCV)*, Nice, France, 2003, pp. 10–17.

[38] J. Vargas, A. X. Falcao, J. A. dos Santos, J. C. Esquerdo, A. Coutinho, and J. F. Antunes, "Contextual superpixel description for remote sensing image classification," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Milan, Italy, 2015, pp. 1132–1135.

[39] G. Zhang, X. Jia, and N. M. Kwok, "Super pixel based remote sensing image classification with histogram descriptors on spectral and spatial data," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Munich, Germany, 2012, pp. 4335–4338.

[40] A. Shokoufandeh and S. Dickinson, "Graph-theoretical methods in computer vision," in *Theoretical Aspects of Computer Science* (Lecture Notes in Computer Science, vol. 2292). Berlin Heidelberg: Springer, 2002, pp. 148–174.

[41] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *J. Atmos. Oceanic Tech.*, vol. 28, no. 10, pp. 1286–1296, Oct. 2011.

[42] P. Tokarczyk, J. D. Wegner, S. Walk, and K. Schindler, "Beyond hand-crafted features in remote sensing," in *Proc. ISPRS Workshop on 3D Virtual City Modeling*, 2013.

[43] A. Heinle, A. Macke, and A. Srivastav, "Automatic cloud classification of whole sky images," *Atmos. Measurement Techniques*, vol. 3, no. 3, pp. 557–567, May 2010.

[44] P. Tokarczyk, J. Montoya, and K. Schindler, "An evaluation of feature learning methods for high resolution image classification," in *Proc. ISPRS Annals Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2012, pp. 389–394.

[45] A. K. Bhandari, A. Kumar, and G. K. Singh, "Feature-extraction using Normalized Difference Vegetation Index (NDVI): A case study of Jabalpur city," in *Proc. Int. Conf. Communication, Computing & Security*, Rourkela, India, 2012, pp. 612–621.

[46] J. Shao and W. Foerstner, "Gabor wavelets for texture edge extraction," in *Proc. SPIE ISPRS Commission III Symp.: Spatial Information from Digital Photogrammetry and Computer Vision*, Munich, Germany, 1994, vol. 2357.

[47] M. Galun, E. Sharon, R. Basri, and A. Brandt, "Texture segmentation by multiscale aggregation of filter responses and shape elements," in *Proc. Int. Conf. Computer Vision (ICCV)*, Nice, France, 2003, pp. 716–723.

[48] S. Dev, Y. H. Lee, and S. Winkler, "Categorization of cloud image patches using an improved texton-based approach," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Quebec City, Canada, 2015.

[49] J. Arenas-Garcia, K. Petersen, G. Camps-Valls, and L. K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods," *IEEE Signal Processing Mag.*, vol. 30, no. 4, pp. 16–29, July 2013.

[50] J. P. Cunningham and Z. Ghahramani, "Linear dimensionality reduction: Survey, insights, and generalizations," *J. Machine Learning Res.*, vol. 16, 2015, pp. 2859–2900.

[51] C. Munyati, "Use of principal component analysis (PCA) of remote sensing images in wetland change detection on the Kafue Flats, Zambia," *Geocarto Int.*, vol. 19, no. 3, pp. 11–22, 2004.

[52] R. Doerffer and D. Murphy, "Factor analysis and classification of remotely sensed data for monitoring tidal flats," *Helgoländer Meeresuntersuchungen*, vol. 43, no. 3–4, pp. 275–293, Sept. 1989.

[53] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[54] Q. Dua and R. Nekoveib, "Implementation of real-time constrained linear discriminant analysis to remote sensing image classification," *Pattern Recognit.*, vol. 38, no. 4, pp. 459–471, Apr. 2005.

[55] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," in *Proc. Neural Information Processing Systems (NIPS)*, Vancouver, Canada, 2005, pp. 513–520.

[56] L. Weizman and J. Goldberger, "Classification of hyperspectral remote-sensing images using discriminative linear projections," *Int. J. Remote Sensing*, vol. 30, no. 21, pp. 5605–5617, Oct. 2009.

[57] World Meteorological Organization, *International Cloud Atlas*, vol. 1, Geneva, Switzerland, 1975.

[58] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[59] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," presented in *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, 2008.

[60] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, 2010, pp. 3501–3508.

[61] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[62] B. Wen, S. Ravishankar, and Y. Bresler, "Structured overcomplete sparsifying transform learning with convergence guarantees and applications," *Int. J. Computer Vision*, vol. 114, no. 2, pp. 137–167, Sept. 2015.

[63] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sensing*, vol. 51, no. 9, pp. 4779–4789, Sept. 2013.

[64] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[65] Y. Pati, R. Rezaiifar, P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, 1993, pp. 40–44.

[66] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Processing*, vol. 61, no. 5, pp. 1072–1086, Mar. 2013.

[67] B. Wen, S. Ravishankar, and Y. Bresler, "Learning overcomplete sparsifying transforms with block cosparsity," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Paris, France, 2014, pp. 803–807.

[68] S. Dev, Y. H. Lee, and S. Winkler, "Systematic study of color spaces and components for the segmentation of sky/cloud images," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Paris, France, 2014, pp. 5102–5106.

[69] S. Liu, L. Zhang, Z. Zhang, C. Wang, and B. Xiao, "Automatic cloud detection for all-sky images using superpixel segmentation," *IEEE Geosci. Remote Sensing Lett.*, vol. 12, no. 2, pp. 354–358, Feb. 2015.

[70] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 9, pp. 1124–1137, Sept. 2004.

[71] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[72] S. Liu, Z. Zhang, B. Xiao, and X. Cao, "Ground-based cloud detection using automatic graph cut," *IEEE Geosci. Remote Sensing Lett.*, vol. 12, no. 6, pp. 1342–1346, June 2015.

[73] S. Dev, Y. H. Lee, and S. Winkler, "Multi-level semantic labeling of sky/cloud images," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Quebec City, Canada, 2015, pp. 636–640.

[74] C. Ari and S. Aksoy, "Unsupervised classification of remotely sensed images using Gaussian mixture models and particle swarm optimization," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Honolulu, HI, 2010, pp. 1859–1862.

[75] U. Maulik and I. Saha, "Modified differential evolution based fuzzy clustering for pixel classification in remote sensing imagery," *Pattern Recognit.*, vol. 42, no. 9, pp. 2135–2149, Sept. 2009.

[76] A. N. Erkan, G. Camps-Valls, and Y. Altun, "Semi-supervised remote sensing image classification via maximum entropy," in *Proc. IEEE Int. Workshop Machine Learning for Signal Processing (MLSP)*, Kittila, Finland, 2010, pp. 313–318.

[77] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression," *IEEE Geosci. Remote Sensing Lett.*, vol. 10, no. 2, pp. 318–322, Mar. 2013.

[78] D. Tuia, R. Flamary, and M. Barlaud, "To be or not to be convex? A study on regularization in hyperspectral image classification," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Milan, Italy, 2015, pp. 4947–4950.

[79] M. Lebrun, M. Colom, A. Buades, and J. Morel, "Secrets of image denoising cuisine," *Acta Numerica*, vol. 21, pp. 475–576, May 2012.

[80] W. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, no. 1, pp. 55–59, Jan. 1972.

[81] J. Lee, "Refined filtering of image noise using local statistics," *Computer Graph. Image Process.*, vol. 15, no. 4, pp. 380–389, Apr. 1981.

[82] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[83] P. Liu, F. Huang, G. Li, and Z. Liu, "Remote-sensing image denoising using partial differential equations and auxiliary images as priors," *IEEE Geosci. Remote Sensing Lett.*, vol. 9, no. 3, pp. 358–362, May 2012.

[84] C. Yu and X. Chen, "Remote sensing image denoising application by generalized morphological component analysis," *Int. J. Appl. Earth Observation Geoinformation*, vol. 33, pp. 83–97, 2014.

[85] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Computer Vision (ICCV)*, Barcelona, Spain, 2011, pp. 479–486.

[86] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learning Res.*, vol. 11, pp. 19–60, Mar. 2010.

[87] S. Ravishankar, B. Wen, and Y. Bresler, "Online sparsifying transform learning–Part I: Algorithms," *IEEE J. Select. Topics Signal Processing*, vol. 9, no. 4, pp. 625–636, June 2015.

[88] S. Ravishankar, B. Wen, and Y. Bresler, "Online sparsifying transform learning for signal processing," in *Proc. IEEE Global Conf. Signal and Information Processing (GlobalSIP)*, Atlanta, GA, 2014, pp. 364–368.

[89] B. Wen, S. Ravishankar, and Y. Bresler, "Video denoising by online 3D sparsifying transform learning," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Quebec City, Canada, 2015, pp. 118–122.

GRS