Ground-Based Image Analysis: A Tutorial on Machine-Learning Techniques and Applications

Soumyabrata Dev, *Student Member, IEEE,* Bihan Wen, *Student Member, IEEE,*
Yee Hui Lee, *Senior Member, IEEE,* and Stefan Winkler, *Senior Member, IEEE*

***Abstract*—Ground-based whole sky cameras have opened up new opportunities for monitoring the earth's atmosphere. Ground-based cameras are an important complement to satellite images and provide remote sensing analysts with cheaper, faster, and more localized data. The images captured by whole sky imagers can have high spatial *and* temporal resolution, which is an important pre-requisite for applications such as solar energy modeling, cloud attenuation analysis, local weather prediction, etc.**

Extracting 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 walk-through of recent developments in these techniques and their applications in ground-based imaging. We aim 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.

***Index Terms*—whole-sky images, dimensionality reduction, features, segmentation, classification, denoising.**

I. Introduction

SATELLITE images are commonly used to monitor the earth and analyze its various properties. They provide remote sensing analysts with accurate information about 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, multi-temporal 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) [3] has demonstrated that 70% of the world's atmosphere is covered with clouds; in previous studies, this figure had been underestimated as 50%. Therefore, there has been a renewed interest amongst 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 either have high temporal resolution (e.g. geostationary satellies) or high spatial resolution (e.g. low-orbit satellites), but never both. In many applications like solar energy prediction [4], local weather prediction, tracking contrails at high altitudes [5], studying aerosol properties [6], attenuation analysis [7], [8], we need data with high spatial *and* temporal resolution. 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 Whole Sky Imager (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. A commercial WSI called TSI-440, manufactured by Yankee Environmental Systems, is used by many researchers [9], [10]. Owing to the high cost and limited flexibility of commercial sky imagers, many research groups have built their own WSI models [9], [11]–[13]. For example, the Scripps Institution of Oceanography at the University of California San Diego has been developing and using WSIs as part of their work [14]. Likewise, our group has designed the Wide-Angle High-Resolution Sky Imaging System (WAHRSIS) for cloud monitoring purposes [15], [16]. Table I provides an overview of the types of ground-based sky cameras used by various organizations around the world, and their primary applications.

II. Machine Learning for Remote Sensing Data

The rapid increase in computing power has enabled the use of powerful machine learning algorithms on large datasets. Remote sensing data fill this description and are typically available in different temporal, spatial, and spectral resolutions. For aerial surveillance and other monitoring purposes, RGB images captured by low-flying aircraft or drones are often used. Multispectral data are used for forest, land, and sea monitoring. Quite recently, hyperspectral imaging systems with very narrow bands are used in 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

Manuscript received Month Day, 2015; revised Month Day, 2015; accepted Month Day 2015. This work is supported by a grant from Singapore's Defence Science & Technology Agency (DSTA).

S. Dev and Y. H. Lee are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: soumyabr001@e.ntu.edu.sg, EYHLee@ntu.edu.sg).

B. Wen is with the Advanced Digital Sciences Center (ADSC), the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, IL 61801, USA (e-mail: bwen3@illinois.edu).

S. Winkler is with the Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign, Singapore (e-mail: Stefan.Winkler@adsc.com.sg).

Send correspondence to S. Winkler, E-mail: stefan.winkler@adsc.com.sg.

Application	Organization	Country	WSI Model
Air traffic control [17]	Campbell Scientific Ltd.	United Kingdom	IR NEC TS9230
Cloud attenuation [15]	Nanyang Technological University Singapore	Singapore	WAHRSIS
Cloud characterization [18]	Atmospheric Physics Group	Spain	GFAT All-sky imager
Cloud classification [12]	Brazilian Institute for Space Research	Brazil	TSI-440
Cloud classification [19]	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 [20]	Laboratoire de Météorologie Dynamique	France	Nikon D100 with FOV 63°
Convection [21]	Creighton University	United States	Digital Camera
Radiation balance [22]	Lindenberg Meteorological Observatory	Germany	VIS/NIR 7
Solar power forecasting [23]	Solar Resource Assessment & Forecasting Laboratory	United States	TSI-440
Solar power forecasting [23]	Solar Resource Assessment & Forecasting Laboratory	United States	Total Sky Imager 440A
Weather reporting [24]	Ecole Polytechnique Fédérale de Lausanne	Switzerland	Panorama Camera

TABLE I: Overview of the various types of ground-based sky cameras, their associated organization and their intended applications.

clouds' future location, detect and track contrails and monitor aerosols. This is important in applications such as cloud attenuation, solar radiation modeling and others, which require high temporal and spatial resolution data. The requirement for high-resolution data is further exemplified by places where weather conditions are more localized. Such microclimates are prevalent mainly near bodies of water which may cool the local atmosphere, or in heavily urban areas where buildings and roads absorb the sun's energy (Singapore being a prime example of such conditions). This leads to quicker cloud formation which can have sudden impact on signal attenuation or solar radiation. Therefore, high-resolution ground-based imagers are required for a continuous and effective monitoring of the earth's atmosphere.

There are challenges of applying machine learning techniques in remote sensing. The high dimensionality of remote sensing data can provide rich information and a complex data model. However, 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 easily occur in the learned model, which may eventually affect the performance of the applications.

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 Section III, we discuss some of the most popular types of features, including computer vision features, remote-sensing features, dimensionality reduction, 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.

We provide an exhaustive analysis of different applications in ground-based image analysis in this paper. We show that a learning-based framework can potentially perform better compared to heuristic approaches in three primary applications, namely segmentation, classification and denoising. Image segmentation is the task of categorizing pixels into meaningful regions, which share similar properties, belong to same group, or form certain objects. Classification is the problem of recognizing objects based on some pre-defined categories. Denoising estimates the true signals from their corrupted observations.

In this paper, we show how a number of popular, state-of-the-art machine learning methods can be effectively used in remote sensing by illustrating the capabilities of these techniques for ground-based image analysis. A high-level schematic framework of machine learning applications in remote sensing data analysis is shown in Fig. 1.

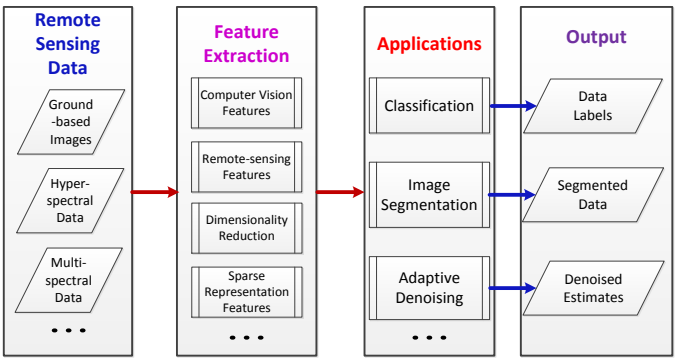


Fig. 1: High-level schematic framework of remote sensing data analysis with machine learning techniques.

III. Feature Extraction

Effective image features are important for computational efficiency and enhanced performance in different applications.

Because of the high dimensionality of the data, it is difficult and inefficient to train using the raw data directly. Moreover, the effect of collinearity amongst 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 those 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, we distinguish four primary categories of feature extraction techniques in this paper, which will be discussed in more detail below:

- Computer vision features;
- Remote-sensing features;
- Dimensionality reduction;
- Sparse representation features.

A. Computer Vision Features

Traditional computer vision feature extraction techniques mainly consist of corner and edge detectors. The term corner has varied interpretations. Intuitively speaking, a corner denotes a region where there is a sharp variation in illumination. These corner points may not always represent the projection of a 3D 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]. It relies on a small window that slides across the image and looks for variation 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], for example.

Aside from corners, blobs are also popular discriminatory features. Blobs are small image regions that possess similar characteristics with respect to color, intensity etc. Popular blob detectors are Difference of Gaussians (DoG), Scale-Invariant Feature Transform (SIFT) [28] and Speeded-Up Robust Features (SURF) [29]. These feature descriptors have high invariability to affine transformations such as rotation.

DoG is a band-pass 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. Because of its attractive property to enhance information at certain frequency ranges, DoG is used to separate the specular reflection from Ground Penetrating Radar (GPR) images [30]. This is necessary for the detection of landmines using radar images. DoG also has wide applications in obtaining pan-sharpened images, which have high spectral and spatial resolutions [31].

SIFT and SURF are two other very popular blob-based feature extraction technique in computer vision that are widely used in remote sensing analysis. SIFT extracts a set of feature vectors from an image that are invariant to rotation, scaling, and translation. It does this by detecting extrema in a series of sampled and smoothed versions of the input image. It 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 feature points in the input image. Its main advantage is its faster execution as compared to SIFT. Image matching on Quickbird images are 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, position, etc. in their neighborhood [36]. These local features help in retaining the local information of the image, and provide insightful cues for different applications as image retrieval and image mining.

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

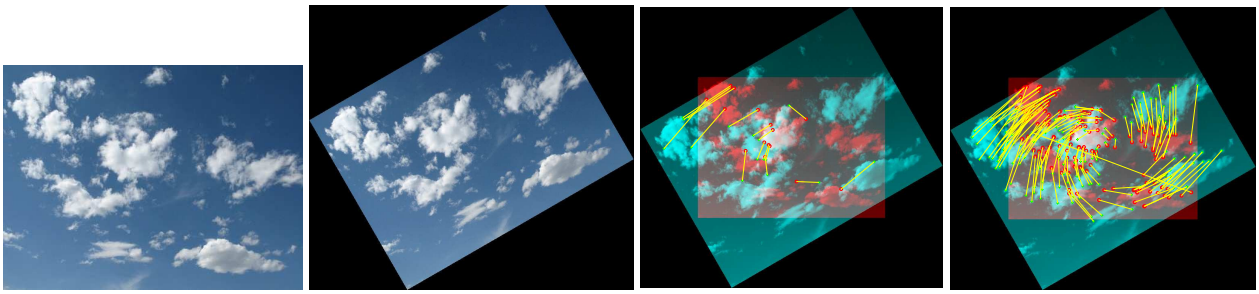
One of the popular pixel-grouping techniques is *superpixel*, that groups pixels having similar appearance. Superpixels were introduced by Ren and Malik [37] in an image classification framework. These techniques are also applied for remote sensing image classification. Recently Vargas et al. [38] presented a Bag Of Words (BOW) model using superpixels for multispectral image classification. Zhang et al. [39] use superpixel-based feature extraction in the 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 sub-graphs, such that the similarity of pixels within a sub-graph is maximized. A good review of the various graph-theoretical models in computer vision is provided by Shokoufandeh and Dickinson [40].

In order 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 HYTA database [11]. The original image is distorted by scaling it with a ratio of 1.3 and rotated by 30°. Fig. 2 shows the matching results between input and the distorted image. We show the candidate matching using the Harris corner detector in Fig. 2(c) and SURF in Fig. 2(d). 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.

B. Remote-sensing Features

In remote sensing, hand-crafted features exploiting the characteristics of the input data are widely used for image classification [41]. It involves the generation of a large number of features that capture the discriminating cues in the



(a) (b) (c) (d)

Fig. 2: Feature matching between original image and distorted image (scaled with a ratio of 1.3 and rotated by 30°). (a) Original image (b) Distorted image (c) Candidate matched points using Harris corner detector (d) Candidate matched points using SURF detector.

remote sensing data. The user makes an educated guess of the most appropriate features. Unlike the popular computer vision feature extraction techniques, 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 show promising results in certain applications.

For example, Heinle et al. [42] used a 12-dimensional feature vector that captures the color, edge, and texture information of a sky/cloud image. The Heinle feature is quite popular in cloud classification. The raw intensity values of RGB aerial images are also used as input features. In satellite imagery, Normalized Difference Vegetation Index (NDVI) is used in association with the raw pixel intensity values for land-cover monitoring, road structures etc. [43]. In high-resolution aerial images, the neighborhood pixels are also considered for the generation of feature vectors. This results in the creation of e.g. 3×3, 15×15, 21×21 etc. pixel neighborhoods. Furthermore, in order to encode the textural features of the input images, Gabor- and edge-based texture filters are used, e.g. for aerial imagery [44] or landscape image segmentation [45]. Recently, we have used a modified set of Schmid filters for the task of cloud classification [46].

C. Dimensionality Reduction

Remote sensing data are inherently high-dimensional in nature. Therefore, it is useful 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 will discuss several popular Dimensionality Reduction (DR) techniques, and provide relevant references for further discussions. A more detailed review of various DR techniques can be found in [47].

Broadly speaking, DR techniques can be classified as either linear or non-linear methods. Linear DR methods represent the original data in a lower-dimensional subspace by a linear transformation, while non-linear methods consider the non-linear relationship between the original data and the features. In this manuscript, we will focus on linear DR techniques because of their lower computational complexity and simple geometric interpretation. For a detailed discussion on the various methods, a unified treatment of the various linear DR techniques is done by Cunningham and Ghahramani in [48].

We denote the data as $\mathbf{X} = \left[x_1 \mid x_2 \mid \ \ldots \ \mid x_n\right] \in \mathbb{R}^{N \times n}$, 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 $\mathbf{Z} = \left[z_1 \mid z_2 \mid \ \ldots \ \mid z_n\right] \in \mathbb{R}^{K \times n}$, where each $z_i \in \mathbb{R}^K$ is the feature representation of x_i, and K denotes the feature dimensionality.

Principal Component Analysis (PCA) is one of the most common and widely used DR techniques. It 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 which are orthonormal, and can be obtained from 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. It is also widely used as change detection in forest fires and land-cover changes. Munyati [49] 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 i^{th} 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 underlying distribution parameters. To our knowledge, there is no proposed algorithm with closed-form solution to this problem. Thus expectation-maximization (EM) algorithm is normally used for applications, but there is no performance guarantee due to non-convexity problem formulation. In remote sensing, FA is used in aerial photography and ground surveys. Doerffer and Murphy [50] 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 j^{th} 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

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

which results in the linear DR mapping $\mathbf{W}$ that is used to produce LDA feature $\mathbf{Z}$ as $\mathbf{Z} = \mathbf{W}\mathbf{X}$.

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. This is because labeled data is expensive, and it is difficult to collect a large number of training samples. For such scenarios, Bandos et al. [51] used regularized LDA in the context of hyperspectral image classification. Du and Nekovel [52] proposed a Constrained LDA for efficient real-time hyperspectral image classification.

Finally, **Neighborhood Component Analysis (NCA)** was first introduced by Goldberger et al. [53]. It aims to find the feature space using a linear transform such that the average leave-one-out K-Nearest Neighbor (KNN) score in the transformed space is maximized. Denoting such a transform as $\mathbf{A}$, NCA can be represented as:

$$\mathbf{Z} = \mathbf{A}\mathbf{X}. \tag{6}$$

NCA aims to reduce the input dimensionality N by learning the transform $\mathbf{A}$. It is learned from the data-set by defining a differentiable cost function for $\mathbf{A}$ [53]. However, this cost function is non-convex in nature, and thus the solution obtained may be sub-optimal. NCA attempts to maximize the expected classification performance under a Leave-One-Out (LOO) scenario in the training data set.

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

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

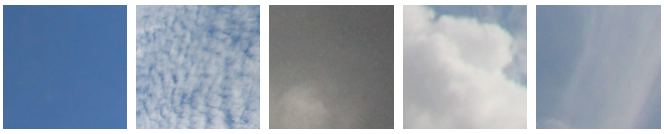
where d_{ij} is the distance between points i and j, and $p_{ii} = 0$.

NCA is widely used in remote sensing for the classification of hyper-spectral images [54]. Weizman and Goldberger [55] have demonstrated the superior performance of NCA in the context of images obtained from Airborne Visible/Infrared Imaging Spectroradiometer.

Technique	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 LOO score	Yes	No

TABLE II: Summary of linear dimensionality reduction techniques.

We now illustrate the effect of different DR techniques in the context of ground-based cloud classification. Cloud types are properly documented by the World Meteorological Organization (WMO) [56]. For this purpose, we use the recently released cloud categorization database called SWIMCAT (Singapore Whole-sky IMaging CATegories database) [46]. The SWIMCAT database[1] consists of a total of 784 sky/cloud image patches divided into 5 categories: clear sky, patterned clouds, thick dark clouds, thick white clouds and veil clouds. Sample images from each category can be found in Fig. 3.



(a) (b) (c) (d) (e)

Fig. 3: Categories for image patches in SWIMCAT: (a) Clear sky, (b) Patterned clouds, (c) Thick dark clouds, (d) Thick white clouds, (e) Veil clouds.

We extract the 12-dimensional Heinle feature (cf. Section III-B) for each image. We randomly select 50 images from each of the 5 cloud categories. For easier computation, images are downsampled to a resolution of 32×32 pixels using bicubic interpolation. Once the feature vectors are generated, the above-mentioned linear DR techniques viz. PCA, FA, LDA, NCA are applied on the entire input feature space.

[1] SWIMCAT can be downloaded from http://vintage.winklerbros.net/swimcat.html

Fig. 4 visualizes the results obtained with the different techniques. The original high-dimensional feature vector is projected to 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 between each other, as they attempt to express the input variables in terms of latent variables. However, we should note that they are distinct methods based on different underlying philosophies, which is exemplified by the results shown in Fig. 4. The separation of features in LDA is relatively good as compared to PCA and FA. This is because LDA aims to increase class-separability in addition to capturing the maximum variance. NCA also essentially separates the different classes quite well. In order to further quantify this separability of different classes in the transformed domain, we will present an quantitative analysis in Section IV-B.

D. Sparse Representation Features

Features based on sparse representation have been widely studied and used in signal processing and computer vision. Different from DR, which 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 serves as a good discriminator in various tasks including face recognition [57], image segmentation [58], object classification [59], and denoising [60]. Learning-based sparse representation also demonstrates advantages in problems such as remote sensing image fusion [61] and hyperspectral image classification [62].

Several models for sparsity have been proposed in recent years. The most popular one is the **synthesis model**, 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 $\mathbf{Z}$:

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

where $\|.\|_0$ counts the number of non-zeros, which is upper-bounded by the sparsity level s. The codes $\{z_i\}_{i=1}^{n}$ are sparse, meaning that the maximum number of non-zeros s is much less than the code dimensionality K. The matrix $\mathbf{D} = \begin{bmatrix} d_1 \mid d_2 \mid & \ldots & \mid d_K \end{bmatrix}$ 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, which also form the basis of x_i. In other words, data that satisfies such a sparse model lives in a union of subspaces spanned by only a small number of selected atoms of D due to sparsity. The generalized synthesis model allows for small modeling errors in the data space, which is normally more practical.

Given data $\mathbf{X}$, finding the "optimal" dictionary is well-known as the synthesis dictionary learning problem. Since the problem is normally non-convex, and the exact solution is NP-hard, various approximate methods have been proposed and have demonstrated good empirical performance. Among those, the K-SVD algorithm [60] 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:

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

where $\|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2$ 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 (OMP) [63] for sparse coding and updates the dictionary atoms sequentially, while fixing the support of corresponding $\mathbf{Z}$ component by using Singular Value Decomposition (SVD).

Besides synthesis dictionary learning, there are learning algorithms associated with other models such as **transform learning** [64]. Different from synthesis dictionary learning, which is normally sensitive to initialization, the transform learning scheme generalizes the use of conventional analytical transforms such as DCT or Wavelet to a regularized adaptive transform $\mathbf{W}$ as follows:

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

where $\|\mathbf{W}\mathbf{X} - \mathbf{Z}\|_F^2$ denotes the modeling error in the adaptive transform domain. Function $\nu\,(.)$ is the regularizer for $\mathbf{W}$ [64], to prevent trivial and badly-conditioned solutions. The corresponding algorithms [64], [65] provide exact sparse coding and a closed-form transform update with lower complexity and faster convergence.

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 (dictionaries, or transforms) as features for applications such as segmentation and classification.

Fig. 5 provides a simple cloud/sky image segmentation example using OCTOBOS [65], which learns a union of transforms as features, to illustrate and visualize the usefulness of sparse features. We extract 9×9 overlapping image patches from the ground-based sky image shown in Fig. 5(a). The color patches are converted to gray-scale and vectorized to form the 81-dimensional data. As shown in Fig. 5(b), the OCTOBOS scheme is capable of learning sparse representations and clustering the data into sky and cloud simultaneously, by comparing the sparsification errors. In the learning process, we restrict the sparsity of each data to be at most 10 out of 81. The distinct sparsifiers, or rows of learned OCTOBOS, are visualized as 9×9 patches in blocks in Fig. 5(c). Both the sparse codes and the learned transform blocks are used as features in this example. Note that we did not use any other remote-sensing features on top of the OCTOBOS clustering scheme [65]. A hybrid version which combines this with cloud-specific features [13] may further enhance the segmentation performance.

IV. 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 primarily

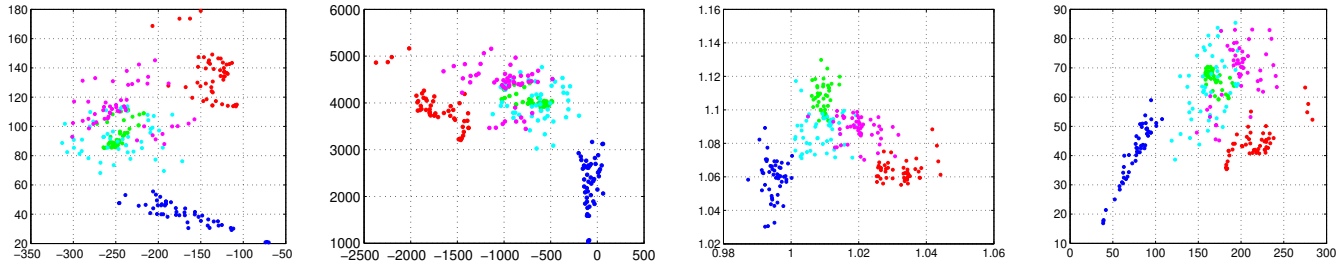


Fig. 4: Visualization of results from applying (a) PCA, (b) FA, (c) LDA, (d) NCA on the SWIMCAT dataset [46]. The data are reduced from their original 12-dimensional feature space to 2 dimensions in the projected feature space, for a five-class cloud classification problem. The different colors indicate individual cloud classes (red: clear sky; green: patterned clouds; blue: thick dark clouds; cyan: thick white clouds; magenta:veil clouds). (*best viewed in color*)

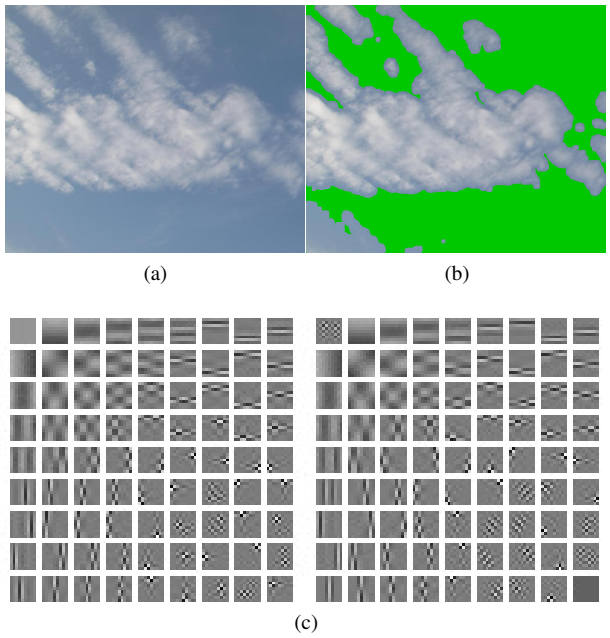


Fig. 5: Cloud and sky segmentation via learning OCTOBOS sparse representation: (a) Original image (b) Input image with original pixels clustered as Cloud, and green pixels clustered as Sky, and (c) Learned two-class OCTOBOS, with each row visualized as patches in separate blocks.

touch upon three main applications: segmentation, classification and denoising. We show that data-driven machine learning techniques generally outperform conventional heuristic approaches.

A. Image Segmentation

Image segmentation refers to the task of dividing the original image into several segments, in an attempt to identify the different objects in the image. In remote sensing, the task of image segmentation has been extensively studied 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' non-rigid 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. use superpixels to identify local homogeneous regions of sky and cloud [66]. In Fig. 6, we illustrate the over-segmented superpixel image of a sky/cloud image from the HYTA database [11]. The generated superpixels respect the image boundaries quite well, and are consistent based on texture and color of sky and cloud regions. These local regions can thus be used for subsequent machine learning tasks.

The final sky/cloud binary image can be obtained by thresholding this over-segmented image using a threshold matrix [66]. In addition to superpixels, graph-cut based techniques [67], [68] have also been explored in ground-based image analysis. Liu et al. proposed an automatic graph-cut technique in identifying sky/cloud regions [69]. As an illustration, we show the two-level segmented output using automatic graph cut in Fig. 6(c).

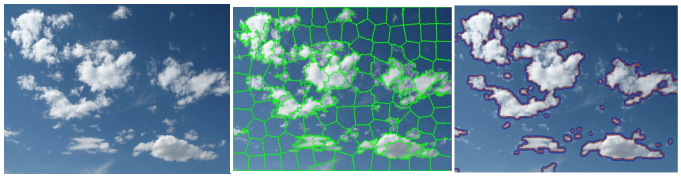
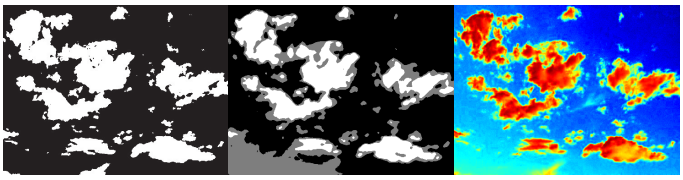


Fig. 6: Illustration of sky/cloud image segmentation using two methods - superpixel and graph-cut. (a) Sample image from HYTA database. (b) Over-segmented image with superpixels. (c) Segmented image using graph-cut technique.

As clouds do not have any specific shape, and the cloud boundaries are ill-defined, several approaches have been proposed that use color as a discriminatory feature. The segmentation can be binary [10], [11], multi-level [70], or probabilistic [13]. As an illustration, we show these three cases for a sample image of HYTA dataset. Fig. 7(a) shows the binary output image of a sample input image from the HYTA

database [11]. The process involves thresholding the selected color channel using a threshold (either fixed or adaptive). In addition to such binary approaches, a multi-level output image can also be generated. Machine learning techniques involving Gaussian discriminant analysis can be used for such purposes. In [70], 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 3-level semantic labels of the sky/cloud image in Fig. 7(b). In addition to 2-level and 3-level output image, a probabilistic segmentation approach is exploited in [13]. In this approach [13], each pixel is assigned a confidence value of belonging to the cloud category. This is illustrated in Fig. 7(c).



(a) (b) (c)

Fig. 7: Illustration of sky/cloud image segmentation. (a) Binary (or 2-level) output image of a sample input image from HYTA database (b) 3-level semantic segmentation of sky/cloud image [70] (c) Probabilistic segmentation of sky/cloud image using [13].

B. Image Classification

In the most general sense, classification refers to the task of categorizing an input data into two (or more) output labels. We can distinguish between supervised and unsupervised methods. The latter identify underlying latent structures in the input data space, and thereby make 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. This training data helps 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) and swarm optimization are the most commonly used unsupervised classification (clustering) techniques. Ari and Aksoy [71] used GMM and particle swarm optimization for hyperspectral image classification. Maulik and Saha [72] 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 also widely used in remote sensing [73]. SSL uses both labeled and unlabeled data in its classification framework. It helps in creating a robust learning framework, which 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 [74]. In addition to SSL methods, models involving sparsity and other regularized approaches are also becoming popular. For example, Tuia et al. [75] study the use of non-convex regularization in the context of hyperspectral imaging.

In ground-based image analysis, image classification refers to categorizing sky/cloud types into various kinds - clear sky, patterned clouds, thick dark clouds, thick white clouds and veil clouds. In order to quantify the accuracy of the separation of data in Fig. 4, we use several popular clustering techniques in combination with DR techniques. We use two classifiers for evaluation purposes, namely KNN and Support Vector Machine (SVM). KNN is a non-parametric classifier, wherein the output label is estimated using a majority voting of the labels of the neighborhood. Support Vector Machine (SVM) is a parametric method that generates a hyperplane or a set of hyperplanes in the vector space, by maximizing the margin between the classifiers to the nearest neighbor data.

We evaluate for five distinct scenarios: (a) PCA, (b) FA, (c) LDA, (d) NCA, (e) No DR technique, and report the classification performances of both KNN and SVM in each of these cases. We again use the SWIMCAT [46] database for evaluation purposes. The training set and testing set consists of a random selection of 50 distinct images. All images are downsampled to 32×32 for easier 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 2-dimensional Heinle feature in training a KNN/SVM classifier for the scenarios (a), (b), (c) and (d). We use the original 12-dimensional vector for training the classifier model for scenario (e). In the testing stage, we obtain the projected 2-D feature points using the computed projection matrix. Thereby, we use KNN/SVM classifier for classifying the test images into individual categories. The average classification accuracy across the 5 classes is shown in Fig. 8.

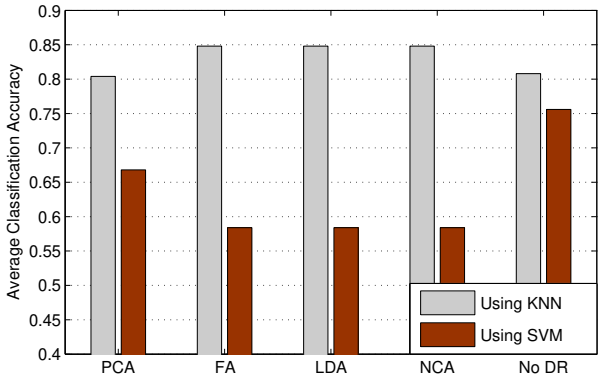


Fig. 8: Average multi-class classification accuracy using Heinle features for cloud patch categorization for different methods.

The KNN classifier clearly achieves better performance than the SVM classifier in all of these cases. It can be clearly interpreted from the 2-D projected feature space (cf. Fig. 4)

that the data points belonging to an individual category lie close to each other. However, it is difficult to separate the different categories using separating hyperplanes in 2-D space. We observe that the complexity of linear SVM classifier is not sufficient to separate the individual classes. KNN performs relatively better in this example. Amongst the different DR techniques, LDA and NCA work the best with the KNN classifier. This is because these methods also use the class labels to obtain maximum inter-class separability. Moreover, the performance without prior dimensionality reduction performs comparably well. In fact, the SVM classifier provides increasingly better results when the feature space has higher dimensionality. This shows that further applications of DR on top of extracting remote sensing features may not be necessary in a classification framework. Of course, dimensionality reduction significantly reduces the computational complexity.

C. Adaptive Denoising

Image and video denoising problems have been heavily studied in the past, with various denoising methods proposed [76]. Denoting the true signal (i.e., clean image or video) as x, the measure y is normally corrupted by additive noise e as

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

The goal of denoising is to obtain the 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 $\tilde{x}$.

Early approaches of denoising used fixed analytical transforms, simple probabilistic models [77], or neighbourhood filtering [78]. Recent non-local methods such as BM3D [79] have been shown to achieve excellent performance, by combining some of these conventional approaches. In the field of remote sensing, Liu et al. [80] used partial differential equations for denoising multi-spectral and hyper-spectral images. Yu and Chen [81] introduced Generalized Morphological Component Analysis (GMCA) for denoising satellite images.

Very recently, machine learning based denoising methods have received increasing interest. Compared to fixed models, adaptive sparse models [60], [64], [82], or probabilistic models [83], [84] have been shown to be more beneficial in image reconstruction. The proposed sparsity-based methods, such as K-SVD [60] and OCTOBOS [65], were introduced in Section III. Adaptive GMM-based denoising [84] also provides promising performance by learning a GMM from the training data as 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 [11] shown in Fig. 6(a) is used as an example and serves as ground truth. We synthetically generate i.i.d. Gaussian noise with $\sigma = 20$ to the clean data. The obtained noisy image is shown in Fig. 9(a).

(a) (b)

Fig. 9: Ground-based image denoising result: (a) Noisy cloud image (PSNR = 22.1 dB), (b) Denoised image (PSNR = 33.5 dB) obtained by using GMM-based algorithm.

Fig. 10 provides the denoising results comparison using several popular learning-based denoising schemes, including GMM [84], OCTOBOS [65], and K-SVD [60]. The quality of the denoised image is measured by Peak Signal-to-Noise Ratio (PSNR) as the objective metric (the clean image has infinite PSNR value). As a comparison, we also include the denoising result by applying a fixed overcomplete DCT dictionary [60]. DCT is an analytical transform commonly used in image compression technique such as JPEG. To give a fair comparison, we maintain the same sparse model richness, by using 256×64 transform in OCTOBOS, and 64×256 dictionaries in K-SVD and DCT. For GMM, we follow the default settings in the publicly available software.

As illustrated in Fig. 10, learning-based denoising methods clearly provide better denoised PSNRs than DCT-based method, with an average improvement of 0.9 dB. Among all of the learning-based denoising algorithms, K-SVD and OCTOBOS are unsupervised learning methods using image sparsity. OCTOBOS additionally combines clustering procedure, in order to learn a structured overcomplete sparse model. GMM is a supervised learning method which is pre-trained with a standard image corpus. In our experiment, OCTOBOS and GMM perform slightly better than K-SVD, since they are using either a more complicated model or supervised learning. The denoising example using the GMM-based method is shown in Fig. 9.

WSIs continuously generate large-scale cloud image data which need to be processed efficiently. Though learning-based algorithms can provide promising performance in applications such as denoising, most of them are batch algorithms. Consequently, batch methods such as K-SVD and OCTOBOS, require correspondingly large storage, and also introduce latency if the data are generated in real time. Thus, online versions of the learning-based methods [85], [86] are needed to process high-resolution WSI data. These learning schemes are scalable to big data problems, by taking advantage of stochastic learning techniques. Here, we show an example of denoising a color image of size 3000×3000 pixels generated by WAHRSIS at night, shown in Fig. 11. Note that such a method is also capable of processing real-time and

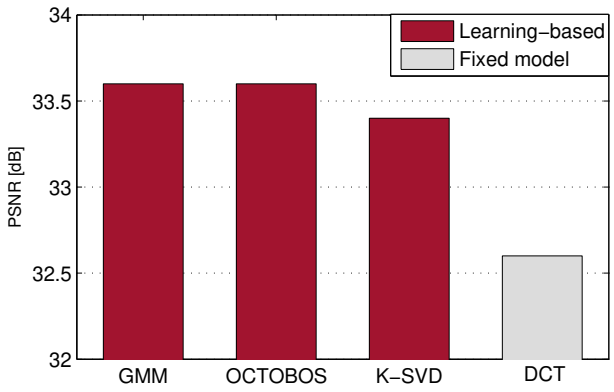
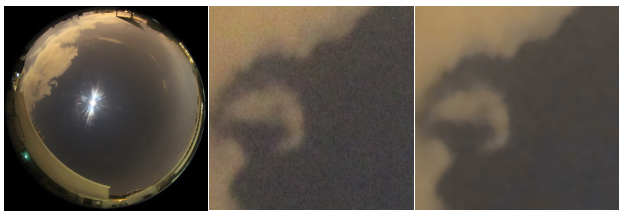


Fig. 10: PSNR values for denoising with OCTOBOS, GMM method, K-SVD, and DCT dictionary

high-dimensional data [87]. Thus it can be easily extended to applications involving multi-temporal satellite images and multispectral data in remote sensing.



(a) (b) (c)

Fig. 11: Real large-scale nighttime cloud image denoising result: (a) Noisy cloud image, (b) Zoom-in of the noisy image (c) Zoom-in of the denoised image obtained by using the online transform learning based denoising scheme.

V. Conclusion

In this tutorial paper, we have provided a walk-through of recent developments of machine learning techniques in ground-based image analysis. Sensing the earth's atmosphere using high-resolution ground-based sky cameras provides a cheaper, faster, and localized manner of data acquisition. Because of 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 show promising performance in comparison to heuristic approaches.

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 Sensing of Environment*, vol. 56, no. 1, pp. 66–77, April 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 Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 1905–1912, July 2008.

[3] NASA, "Nasa's icesat satellite," http://www.nasa.gov/vision/earth/lookingatearth/icesat_light.html, 2015, [Online; accessed 2015].

[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, November 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," *Atmospheric Measurement Techniques*, vol. 6, no. 12, pp. 3597–3612, December 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," *Journal of Geophysical Research*, vol. 115, no. D20, October 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 International Symposium on Antennas and Propagation*, 2014.

[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 International Symposium on Antennas and Propagation*, 2014.

[9] C. N. Long, J. M. Sabburg, J. Calbó, and D. Pages, "Retrieving cloud characteristics from ground-based daytime color all-sky images," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 5, pp. 633–652, 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," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 3, pp. 437–447, 2006.

[11] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *Journal of Atmospheric and Oceanic Technology*, vol. 28, no. 10, pp. 1286–1296, October 2011.

[12] S. L. M. 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," *Journal of Atmospheric and Oceanic Technology*, vol. 27, no. 9, pp. 1504–1517, 2010.

[13] S. Dev, Y. H. Lee, and S. Winkler, "Systematic study of color spaces and components for the segmentation of sky/cloud images," in *Proc. International Conference on Image Processing (ICIP)*, 2014.

[14] 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," *Applied Optics*, vol. 52, no. 8, pp. 1605–1616, March 2013.

[15] 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*, 2014.

[16] 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. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015.

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

[18] A. Cazorla, F. J. Olmo, and L. Alados-Arboledas, "Development of a sky imager for cloud cover assessment," *Journal of the Optical Society of America A*, 2008.

[19] 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," *Atmospheric Research*, vol. 113, no. 0, pp. 80–88, September 2012.

[20] 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," in *Proc. International Laser Radar Conference*, 2006.

[21] J. A. Zehnder, J. Hu, and A. Razdan, "A stereo photogrammetric technique applied to orographic convection," *Monthly Weather Review*, 2007.

[22] 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*, 2000.

[23] 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," *Atmospheric Measurement Techniques*, vol. 5, no. 11, pp. 2881–2892, November 2012.

[24] Z. Chen, Y. Feng, A. Lindner, G. Barrenetxea, and M. Vetterli, "How is the weather: Automatic inference from images," in *Proc. International Conference on Image Processing (ICIP)*, 2012.

[25] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Fourth Alvey Vision Conference*, 1988.

[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. International Conference on Recent Advances in Information Technology*, 2012.

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

[28] D.G. Lowe, "Object recognition from local scale-invariant features," in *Proc. International Conference on Computer Vision (ICCV)*, 1999.

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

[30] X. Xiaoyin and E.L. Miller, "Adaptive difference of gaussians to improve subsurface imagery," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2002.

[31] K.P. Upla, M.V. Joshi, and P.P. Gajjar, "Pan-sharpening: Use of difference of gaussians," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2014.

[32] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale-invariant feature matching for optical remote sensing images," *IEEE Transactions on Geoscience and 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 Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 287–291, April 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*, vol. 7368 of *Lecture Notes in Computer Science*, pp. 282–289. Springer Berlin Heidelberg, 2012.

[35] Z. L. Song and J. Zhang, "Remote sensing image registration based on retrofitted SURF algorithm and trajectories generated from lissajous figures," *IEEE Geoscience and Remote Sensing Letters*, 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026–1038, August 2002.

[37] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. International Conference on Computer Vision (ICCV)*, 2003.

[38] J. Vargas, A. X. Falc ao, J. A. dos Santos, J. C. Esquerdo, A. Coutinho, and J. F. Antunes, "Contextual superpixel description for remote sensing image classification," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015.

[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. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012.

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

[41] 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.

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

[43] A.K. Bhandari, A. Kumar, and G.K. Singh, "Feature extraction using Normalized Difference Vegetation Index (NDVI): A case study of jabalpur city," *Procedia Technology*, vol. 6, no. 0, pp. 612–621, 2012.

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

[45] M. Galun, E. Sharon, R. Basri, and A. Brandt, "Texture segmentation by multiscale aggregation of filter responses and shape elements," in *Proc. International Conference on Computer Vision (ICCV)*, 2003.

[46] S. Dev, Y. H. Lee, and S. Winkler, "Categorization of cloud image patches using an improved texton-based approach," in *Proc. International Conference on Image Processing (ICIP)*, 2015.

[47] 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 Magazine*, vol. 30, no. 4, pp. 16–29, July 2013.

[48] J. P. Cunningham and Z. Ghahramani, "Unifying linear dimensionality reduction," arXiv e-prints, June 2014.

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

[50] 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, 1989.

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

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

[53] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," in *Proc. Neural Information Processing Systems (NIPS)*, 2005.

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

[55] L. Weizman and J. Goldberger, "A classification-based linear projection of labeled hyperspectral data," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2007.

[56] World Meteorological Organization, *International Cloud Atlas*, vol. 1, 1975.

[57] A. Yang, J. Wright, Y. Ma, and S. Sastry, "Feature selection in face recognition: A sparse representation perspective," Tech. Rep., 2007.

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

[59] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

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

[61] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4779–4789, September 2013.

[62] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3973–3985, October 2011.

[63] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conference on Signals, Systems and Computers*, 1993.

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

[65] B. Wen, S. Ravishankar, and Y. Bresler, "Structured overcomplete sparsifying transform learning with convergence guarantees and applications," *International Journal on Computer Vision*, 2014.

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

[67] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, September 2004.

[68] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, November 2001.

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

[70] S. Dev, Y. H. Lee, and S. Winkler, "Multi-level semantic labeling of sky/cloud images," in *Proc. International Conference on Image Processing (ICIP)*, 2015.

[71] C. Ari and S. Aksoy, "Unsupervised classification of remotely sensed images using gaussian mixture models and particle swarm optimization," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2010.

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

[73] A.N. Erkan, G. Camps-Valls, and Y. Altun, "Semi-supervised remote sensing image classification via maximum entropy," in *Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2010.

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

[75] D. Tuia, R. Flamary, and M. Barlaud, "To be or not to be convex? a study on regularization in hyperspectral image classification," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015.

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

[77] W. Richardson, "Bayesian-based iterative method of image restoration," *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55–59, 1972.

[78] J. Lee, "Refined filtering of image noise using local statistics," *Computer graphics and image processing*, vol. 15, no. 4, pp. 380–389, 1981.

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

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

[81] C. Yu and X. Chen, "Remote sensing image denoising application by generalized morphological component analysis," vol. 33, pp. 83–97, 2014.

[82] M. Yaghoobi, S. Nam, R. Gribonval, and M. E. Davies, "Noise aware analysis operator learning for approximately cosparse signals," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.

[83] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, 2012.

[84] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. International Conference on Computer Vision (ICCV)*, 2011.

[85] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[86] S. Ravishankar, B. Wen, and Y. Bresler, "Online sparsifying transform learning - part i: Algorithms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 625–636, 2015.

[87] B. Wen, S. Ravishankar, and Y. Bresler, "Video denoising by online 3d sparsifying transform learning," in *Proc. International Conference on Image Processing (ICIP)*, 2015.