



From pixels to patches: a cloud classification method based on a bag of micro-structures

Qingyong Li¹, Zhen Zhang¹, Weitao Lu², Jun Yang², Ying Ma², and Wen Yao²

¹Beijing Key Lab of Transportation Data Analysis and Mining, Beijing Jiaotong University, Beijing, 100044, China

²State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, 100081, China

Correspondence to: Qingyong Li (qingyongli@gmail.com) and Weitao Lu (wtlu@cams.cma.gov.cn)

Received: 6 July 2015 – Published in Atmos. Meas. Tech. Discuss.: 5 October 2015

Revised: 3 February 2016 – Accepted: 19 February 2016 – Published: 1 March 2016

Abstract. Automatic cloud classification has attracted more and more attention with the increasing development of whole sky imagers, but it is still in progress for ground-based cloud observation. This paper proposes a new cloud classification method, named bag of micro-structures (BoMS). This method treats an all-sky image as a collection of micro-structures mapped from image patches, rather than a collection of pixels. It represents the image with a weighted histogram of micro-structures. Based on this representation, BoMS recognizes the cloud class of the image by a support vector machine (SVM) classifier. Five classes of sky condition are identified: cirriform, cumuliform, stratiform, clear sky, and mixed cloudiness. BoMS is evaluated on a large data set, which contains 5000 all-sky images captured by a total-sky cloud imager located in Tibet (29.25° N, 88.88° E). BoMS achieves an accuracy of 90.9% for 10-fold cross-validation, and it outperforms state-of-the-art methods with an increase of 19%. Furthermore, influence of key parameters in BoMS is investigated to verify their robustness.

1 Introduction

Clouds play an important role in the hydrological cycle and the energy balance of the atmosphere–earth surface system because of the interaction with solar and terrestrial radiation (Stephens, 2005). Cloud type is an important cloud macroscopic parameter and plays an essential role in meteorological research. Classification of cloud types is extensively studied based on both satellites and ground-based weather stations. Cloud classification is first investigated based on satellite images (Ameur et al., 2004; Tahir, 2011; Hu et al.,

2015). Most of these methods apply texture features and classifier models to recognize cloud type. However, the information provided by large-scale satellite images is not sufficient enough. For example, these images have too low resolution to capture detailed characteristics of local clouds; thin clouds and earth surface are frequently confused in satellite images because of their similar brightness and temperature (Ricciardelli et al., 2008). By contrast, ground-based cloud observation can obtain more accurate characteristics for local clouds, and ground-based cloud classification has attracted more and more attention (Tapakis and Charalambides, 2013).

Traditionally, cloud type is generally classified by professional observers in ground-based cloud observation. Human observation, however, is somewhat subjective and inconsistent. For example, different observers may obtain different cloud types according to a same sky condition due to different levels of professional skill. Nowadays ground-based imaging devices, which take advantage of new embedded hardware technology and digital image processing techniques, are commonly applied for automated cloud observation. A number of sky-imaging systems have been developed in recent years (Tapakis and Charalambides, 2013). Currently, there are two frequently referred imaging systems: the first one is the whole-sky imager (WSI) series developed by the Scripps Institute of Oceanography, University of California, San Diego. WSIs measure radiances at distinct wavelength bands across the hemisphere and retrieve cloud characteristics (Shields et al., 1998; Kassianov et al., 2005; Urquhart et al., 2015). The second one is the total-sky imager (TSI) series, which are manufactured by Yankee Environmental Systems, Inc. TSIs provide color images for the daytime hemispheric sky conditions and derive frac-

tional sky cover and other useful meteorological information (Long et al., 2006; Calbo and Sabburg, 2008; Mantelli Neto et al., 2010; Jayadevan et al., 2015). In addition, a number of other ground-based sky imagers have also been developed in other countries and institutes, such as the whole-sky camera (Calbo and Sabburg, 2008; Heinle et al., 2010) and the total-sky cloud imager (Li et al., 2011; Yang et al., 2012, 2015). Most of these sky imagers capture sky conditions with red–green–blue (RGB) color images named all-sky images, and are successfully applied to estimate cloud cover. Automatic cloud type classification, however, is still under development.

The modern cloud classification methods are often built upon specific cloud characteristics with the help of certain machine learning models, such as k nearest neighbor (KNN), artificial neural networks, and support vector machine. Singh and Glennen (2005) investigated most well-known texture features for cloud classification, which recognizes common digital images (without 180° field of view) to five different sky conditions. These features include autocorrelation, co-occurrence matrices, edge frequency, Law's features, and primitive length. They pointed out that no single feature is sufficient enough. Calbo and Sabburg (2008) first studied cloud classification for all-sky images. They represented a cloud image by statistical measurements of texture, frequency characteristics of Fourier transform, and others. However, this method achieves an accuracy of only 62%. Afterwards, Heinle et al. (2010) categorized all-sky images by KNN classifier based on a set of statistical features and gray-level co-occurrence matrices (GLCMs). They divided sky conditions into seven types and achieved high accuracy for leave-one-out cross-validation. Kazantzidis et al. (2012) improved the method of Heinle et al. (2010) by combining traditional features and extra characteristics, such as solar zenith angle, cloud coverage, and the existence of raindrops in sky images. Recently, texture features based on salient local binary patterns are applied for cloud classification, which achieves competitive performance (Liu et al., 2013; Liu and Zhang, 2015). Kliangsuwan and Heednacram (2015) proposed a new technique called fast Fourier transform projection on the x axis. This method extracts features by projecting logarithmic magnitude of fast Fourier transform coefficients of a cloud image on the x axis in frequency domain. Cheng and Yu (2015) presented a block-based cloud classification method, which divides an image into multiple blocks and identifies the cloud type for each block based on both statistical features and distribution of local texture features.

The features, which represent a cloud image with a numerical vector, are essential for cloud classification. The features applied in literature for cloud classification can be roughly divided into three categories: physical, spectral, and textural. Physical features concern the physical properties of a sky condition, such as brightness, temperature, whiteness, and cloud coverage (Kazantzidis et al., 2012). Spectral features describe the average color and tonal variation of a cloud image (Heinle et al., 2010; Xia et al., 2015). Textural fea-

tures refer to the spatial distribution of pixel intensity within a cloud image, i.e., homogeneity, randomness, and contrast of the gray level differences of pixels (Singh and Glennen, 2005; Cheng and Yu, 2015; Liu and Zhang, 2015). Essentially, all these features are built upon pixels, which are encoded by RGB vectors as shown in Fig. 1. They are not sufficient enough for cloud classification considering the following aspects.

- Physical and spectral features are not accurate themselves, because pixels have great variation and are easy to be noised. Mathematically, a pixel can be regarded as an element of a three-dimensional vector set, which has totally 256^3 elements if each channel of red–green–blue is quantized to 256 levels. Furthermore, RGB values are often influenced by cameras and atmospheric interference. So it is a nontrivial task to accurately measure physical characteristics of clouds. For example, cloud coverage, which refers to the fraction of the sky obscured by clouds, depends on the performance of cloud detection, but it is difficult to estimate cirrus clouds (Li et al., 2012).
- Textural features (such as GLCMs) often represent the global appearance of an image and are sensitive to scale and rotation. All-sky images, however, need representation with rotation invariance since clouds may appear in any direction of an all-sky image. Furthermore, such global textural features would be confusing if an all-sky image contained multiple types of clouds (Cheng and Yu, 2015).
- These features based on pixels describe low-level visual characteristics of cloud images, but they fail to encode middle-level structural information or high-level concepts. Structural information, however, is more useful for classification (Zhang et al., 2007). According to Fig. 1, a pixel is just labeled by a RGB vector, but a patch can be defined as a certain micro-structure, which can be given a meaningful description. In fact, the features based on patches are more popular than those features based on pixels in the community of computer vision (Huang et al., 2014).

Accordingly, we put forward the new cloud image representation, named “bag of micro-structures” (BoMS). BoMS is constructed on image patches, rather than pixels. More specifically, an all-sky image is firstly equally divided into patches (maybe with overlap), and then each patch is mapped to a micro-structure by vector quantization. Finally, the image is regarded as a collection of micro-structures just as a textual document consists of many words, and its features are encoded by a weighted histogram of micro-structures. BoMS outperforms the traditional cloud representations based on two factors: (1) a patch is more informative and robust than a pixel for a cloud image. Micro-structures,

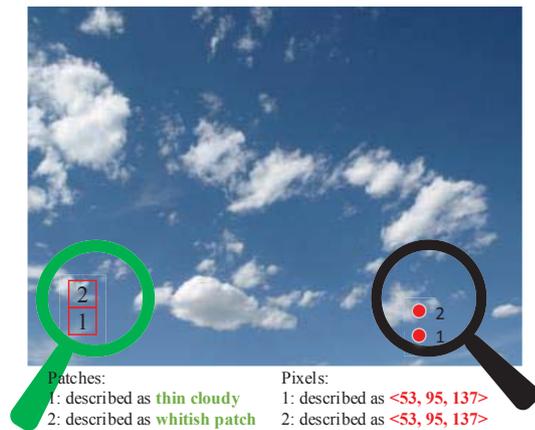


Figure 1. Sketch of the difference between pixels and patches. A pixel is encoded by a RGB vector, whose values are easy to be noised, and a pixel itself is not very meaningful. Meanwhile, a patch is more meaningful than a pixel and can be mapped to certain micro-structures, which are explainable with words.

which are learned offline from an image set, denote general patterns shared by many image patches, so a label of a micro-structure denotes a higher-level concept compared with a RGB vector of a pixel. (2) The holistic histogram representation is high dimensional but sparse, so it is discriminative, even linearly separable for a support vector machine (SVM) classifier.

The remainder of this paper is organized as follows. Section 2 describes the data set and cloud classes. Section 3 introduces the proposed cloud classification method. Section 4 presents experiment results. Finally, Sect. 5 gives our conclusions.

2 Data set and cloud classes

2.1 Data set

Images used for development and evaluation of BoMS were obtained from the total-sky cloud imager (TCI) (Li et al., 2011), which is located in Tibet (29.25° N, 88.88° E). The TCI, developed by the Chinese Academy of Meteorological Sciences, is based on commercially available components. The basic component is a digital camera, which is equipped with a fisheye lens to provide a field of view larger than 180° and is enclosed by a weather protection box. The TCI is programmed to acquire images at fixed intervals, and all images are stored in color JPEG format with a resolution of 1392 × 1040 pixels. Note that these images are rectangular in shape but the mapped whole sky is circular, in which the center is the zenith and the horizon is along the border. Figure 2 displays an example of such an all-sky image. Because the region near the circular border contains certain terrestrial objects, such as trees and buildings on the horizon of the TCI,

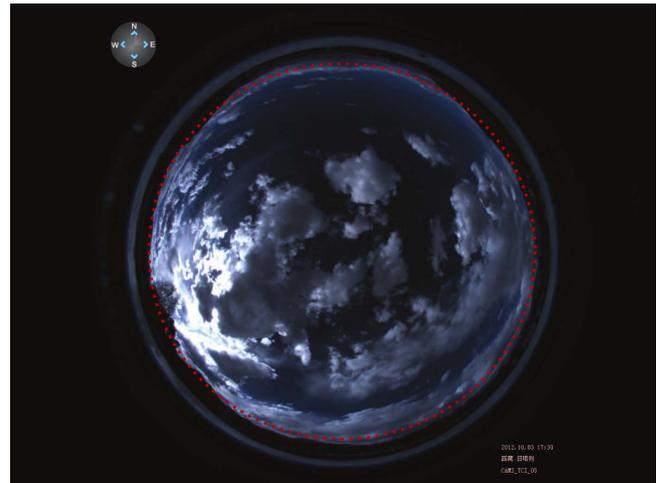


Figure 2. An all-sky image example used in this work (3 October 2012, 17:30 GMT + 8). The area marked by the red circle refers to the circle of interest.

we eliminate the area out of circle of interest (COI). COI is defined by the center (c_x, c_y) and radius r . We set (c_x, c_y) with (718, 536) and r with 442 in this work. Note that COI is the exact area for feature extraction and type identification.

We screened the complete image set that was observed during August 2012 to July 2014 and selected 5000 all-sky images in this work according to our predefined cloud classes (see next section). We did our best to ensure that the data set includes a large variety of different cloud forms. The data set¹ contains 1000 independent images per cloud class.

2.2 Cloud classes

Traditionally, manual cloud classification takes cloud shape as a basic factor, together with shape development and interior micro-structure of the cloud. Clouds are divided into 29 varieties of 10 genera in 3 families with high, mid-, and low levels, according to the “Linnean” system developed by Howard (1803). These criteria are used by surface observers, but they are unsuitable for automatic cloud classification. Calbo and Sabburg (2008) defined eight different sky conditions for automatic cloud classification, while Heinle et al. (2010) considered seven types. Note that there are also other configurations of cloud types for automatic cloud classification, and recent reviews can be found in Tapakis and Charalambides (2013).

Stratiform, cumuliform, and cirriform clouds are the most common sky conditions, and they are primary classes in many cloud classification systems (Tapakis and Charalambides, 2013; Xia et al., 2015). Furthermore, a sky condition obtained by an all-sky imager often contains multiple types of clouds (Cheng and Yu, 2015). We, therefore, define five

¹The data set is available at <http://icn.bjtu.edu.cn/visint/resources/CloudImages>.

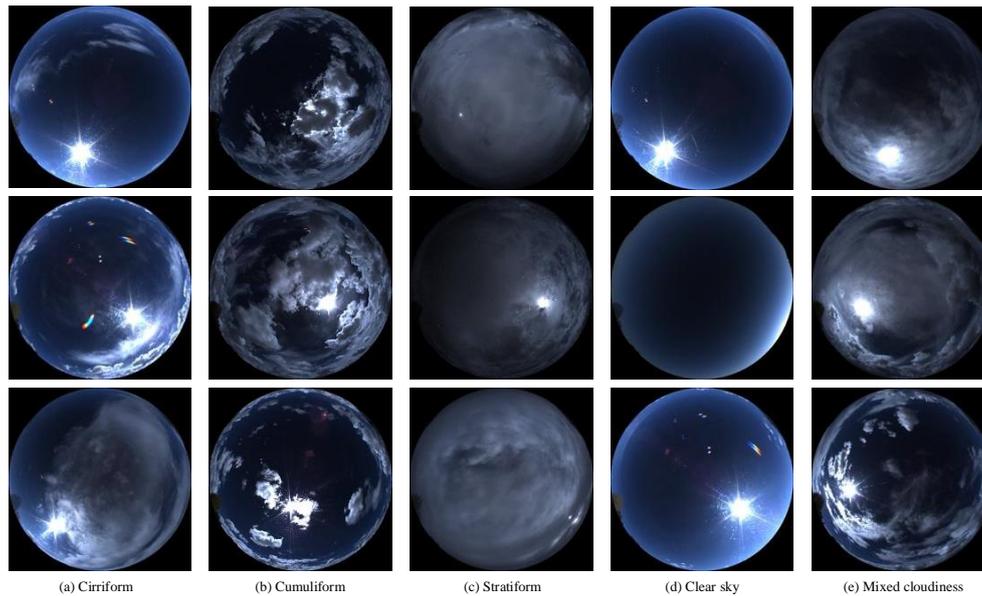


Figure 3. Classic all-sky images of the five sky condition classes.

Table 1. Sky condition classes proposed in this work.

Sky condition classes	Description	Cloud types
Cirriform	Thin clouds that are wispy and feathery-like	Ci, Cc and Cs
Cumuliform	Thick clouds that are puffy and cotton-like	Cu, Cb and Ac
Stratiform	Layered clouds that stretch out across the sky	St, As, Sc and Ns
Clear sky	Clear sky without cloud	No clouds
Mixed cloudiness	Mixed sky conditions with more than one cloud type that covers the sky more than 20%	Co-occurrence

sky conditions for cloud classification as demonstrated in Table 1. In our data set, most images of Cc and Cs are bright and light blue, because the aerosol optical depth is small in Tibet (29.25° N, 88.88° E) where the data set was collected. These features are very similar to those of Ci as shown in Fig. 3a. In addition, Cc, Cs, and Ci all belong to high-level clouds. So we sort these cloud types into the same category. Cumuliform clouds are usually puffy in appearance, similar to large cotton balls; while stratiform clouds are horizontal and layered clouds that stretch out across the sky like a blanket. An image of these classes contains a single cloud type, but an image of mixed condition contains more than one cloud type together. Figure 3 displays some typical all-sky images of these five sky conditions. This configuration of cloud classes is similar to those used by Liu et al. (2011) and Xia et al. (2015), except for the augmented class of mixed cloudiness, which is never investigated in the literature but often occurs in all-sky images.

3 Cloud classification based on a bag of micro-structures

In this section, we first introduce the background of BoMS and the pipeline of the cloud classification method. Then we describe the details of BoMS, including patch descriptor, dictionary of micro-structures, and holistic image representation. At last, the classifier with SVM is presented.

3.1 Overview of the proposed cloud classification method

3.1.1 Review of the bag-of-words model

The bag-of-words model is a simplifying representation used in natural language processing and information retrieval. In this model, a document is represented as the bag of its words, disregarding grammar and even word order but keeping multiplicity (Joachims, 1998). The document is first parsed into words. These words are represented by their stems, for example, “work”, “working”, and “works” would be designated by the stem “work”. Moreover, a stop list is often used to

reject very common words (such as “the”, “a” and “does”), because they occur in most documents and are not meaningful enough to discriminate different documents. After that, the remaining words are assigned with a unique label, and the document is represented by a vector that indicates the occurrences of these words. Note that each component of the vector is often weighted in various ways in order to improve its degree of discrimination (Baeza-Yates and Ribeiro-Neto, 1999). Finally, such vectors are used as features for document classification or used to build an index for information retrieval.

3.1.2 Pipeline of the cloud classification method

Inspired by the bag-of-words model, we propose the new cloud classification method, which treats a cloud image as a collection of micro-structures. More specifically, the proposed method includes two aspects: learning and recognition (as shown in Fig. 4). The learning procedure is offline and carries out two tasks: learning the dictionary of micro-structures and training the SVM model. Recognition procedure analyzes an input image and identifies its cloud class. It includes four main procedures.

1. It divides an input image into patches (maybe with overlap) and extracts a description for each patch according to its appearance. In other words, the input image is treated as a collection of patches, rather than raw pixels.
2. It assigns patch descriptors to a set of predetermined micro-structures by vector quantization. Each patch is mapped to a label of certain micro-structure in a learned dictionary, so the input RGB image can be transformed into a label matrix. Each element of the label matrix refers to an index of a micro-structure. Accordingly, the image is regarded as a bag of micro-structures, just as a document is represented by its words.
3. It constructs a holistic image representation based on BoMS. The histogram of micro-structures is calculated and used as the feature vector of the input image.
4. It applies a SVM classifier to identify the cloud type of the input image, which is represented by a histogram of micro-structures.

We refer to the quantized patch descriptors as micro-structures, because each micro-structure represents a common pattern or appearance shared by many patches. Micro-structures for a cloud image play the same role as words for a text document, though they do not necessarily have an actual meaning as “wispy cloud”, or “puffy cloud”.

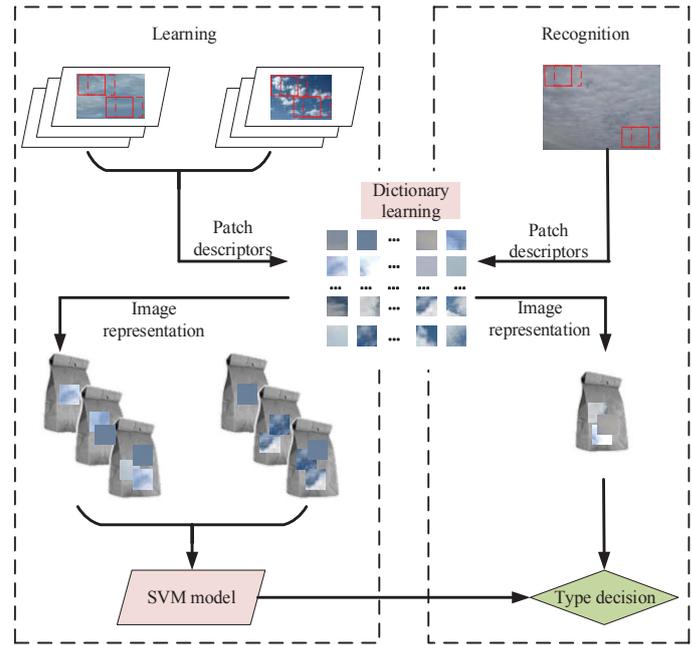


Figure 4. The pipeline of the cloud classification method based on BoMS.

3.2 Cloud representation of BoMS

3.2.1 Patch descriptor based on appearance

A cloud image is regarded as a collection of local patches, rather than simple pixels. Image patches should be firstly described by certain feature vectors (named descriptors) based on their visual appearance. Of course, this descriptor should be discriminative for cloud patches. We apply statistical measurements of color and contrast to describe image patches, because color and contrast are the most important appearance features to distinguish cloud patches from others patterns.

Firstly, a cloud image is equally divided into several patches (maybe with overlap). Given an image I with width w and height h , a patch refers to a square area defined by the top-left point (p_x, p_y) and size s , $1 \leq s \leq \min(w, h)$. Furthermore, all patches are indirectly specified by the sampling step τ . Of course, if τ equals s , a cloud image would be segmented into grids without overlap. Note that the border patches that are partly beyond the scope of COI are discarded because they contain nonsense pixels.

Secondly, statistical measurements of color and contrast are extracted as a descriptor for each patch. The mean, standard deviation, and skewness of blue component are considered as used by Heinle et al. (2010). In addition, similar measurements for the ratio of red and blue components are applied as a Supplement, since such a ratio is powerful to distinguish cloud from sky (Calbo and Sabburg, 2008). Furthermore, the difference between color components is verified to be useful for cloud classification (Heinle et al., 2010),

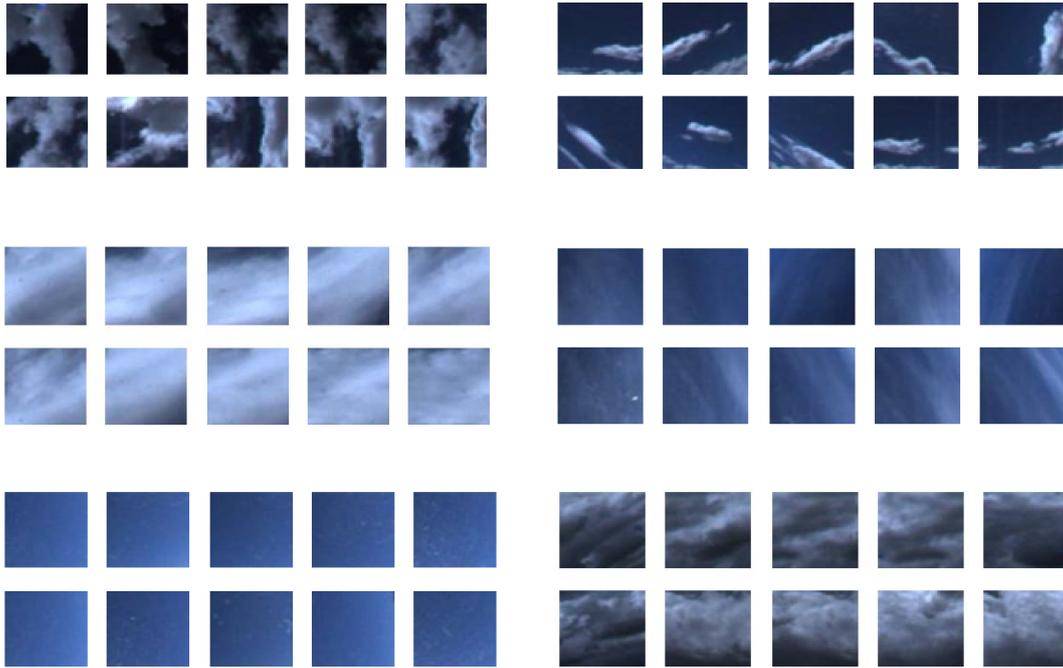


Figure 5. Samples of image patches randomly selected from certain clusters. All images in a cluster are assigned with an identical label of micro-structure.

so improved contrast features based on such difference are included in the patch descriptor as well. More specifically, the patch descriptor is calculated as follows.

Mean, standard deviation, and skewness of blue component

Color is one of the most important characteristics to distinguish clouds from sky. Especially, the blue component has the highest discrimination power. So the mean M_b , standard deviation D_b and skewness S_b of blue components are used in the patch descriptor. M_b encodes the main color appearance, while D_b and S_b partly describe the texture characteristics of an image patch.

$$M_b = \sum_{i=1}^{s^2} B_i / s^2, \quad (1)$$

$$D_b = \sqrt{\sum_{i=1}^{s^2} (B_i - M_b)^2 / (s^2 - 1)}, \quad (2)$$

$$S_b = \frac{1}{s^2 - 1} \sum_{i=1}^{s^2} \left(\frac{B_i - M_b}{D_b} \right)^3, \quad (3)$$

where B_i refers to the intensity value of blue channel for the pixel i in a patch with size s .

Mean, standard, and skewness of the ratio of red and blue components

The ratio of red and blue components is a popular feature used to classify cloud from sky, because a clear sky scatters

more blue than red light and appears blue, whereas clouds scatter blue and red light with similar extent and appear white or gray (Calbo and Sabburg, 2008). So the mean M_t , standard deviation D_t , and skewness S_t of the ratio values are adopted as supplements for the above measurements of blue component.

$$M_t = \sum_{i=1}^{s^2} R t_i / s^2, \quad (4)$$

$$D_t = \sqrt{\sum_{i=1}^{s^2} (R t_i - M_t)^2 / (s^2 - 1)}, \quad (5)$$

$$S_t = \frac{1}{s^2 - 1} \sum_{i=1}^{s^2} \left(\frac{R t_i - M_t}{D_t} \right)^3, \quad (6)$$

where $R t_i (= R_i / B_i)$ represents the ratio of red component and blue component.

Contrast between color components

Heinle et al. (2010) pointed out that D_{GB} (referring to the difference between green and blue channels) and D_{RB} are the most weighted features. Essentially, such a difference is a simple measurement for color contrast, but it ignores the intensity level of each component. For example, the D_{RB} of a dark cloud would be same with that value of a bright cloud, but they are different according to their appearance. So we define contrast as the normalized difference between color

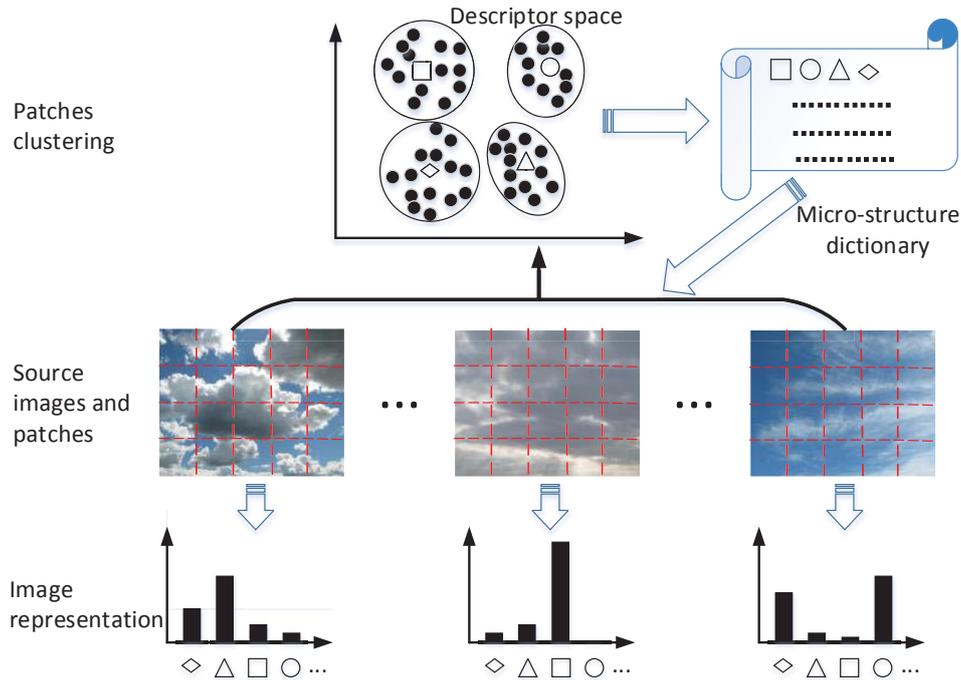


Figure 6. Image representation based on micro-structures, in which patch descriptors are grouped to construct micro-structure dictionary, and histogram of such micro-structures in an image is used as its representation.

components.

$$C_1 = \frac{M_r - M_g}{M_r + M_g}, \quad (7)$$

$$C_2 = \frac{M_r - M_b}{M_r + M_b}, \quad (8)$$

$$C_3 = \frac{M_g - M_b}{M_g + M_b}, \quad (9)$$

where M_r , M_g , and M_b refer to the mean values of red, green, and blue components in the image patch, respectively.

Consequently, each patch can be represented by a nine-dimensional feature vector:

$$D = \langle M_b, D_b, S_b, M_t, D_t, S_t, C_1, C_2, C_3 \rangle. \quad (10)$$

A sampled collection of such descriptors are used to construct a micro-structure dictionary (see the following section). Note that each dimension of D is normalized to $[0, 1]$, in order to eliminate the effect of magnitude.

3.2.2 Learning dictionary of micro-structures with k means algorithm

Dictionary of micro-structures is one of the most important aspects of BoMS, and it is learned offline by the k means algorithm (Han et al., 2006). First, a large number of image patches are equally sampled from the data set, and their descriptors are extracted to form a collection. Second, the descriptor collection is clustered by k means algorithm and

divided into k clusters. Finally, the centroids of these clusters are regarded as micro-structures, and they form the micro-structure dictionary. A micro-structure represents a specific local pattern shared by all patches assigned to it. Figure 5 shows examples of image patches belonging to particular clusters.

K -means clustering is a method of vector quantization, which is popular for cluster analysis in data mining. Given a set of patch descriptors $\{D_1, D_2, \dots, D_n\}$, k means algorithm groups the n descriptors into $k (\leq n)$ sets $S = \{S_1, S_2, \dots, S_k\}$ by minimizing the sum of squared Euclidean distances between descriptors and their corresponding centroids. The objective function of k means is formulated as follows:

$$\operatorname{argmin}_S \sum_{i=1}^k \sum_{D \in S_i} \|D - \mu_i\|^2, \quad (11)$$

where μ_i represents the centroid of cluster i . Given an initial set of k centroids $\{\mu_1^0, \mu_2^0, \dots, \mu_k^0\}$, k means algorithm alternates between assignment step and update step.

- *Assignment step.* It assigns each descriptor to its nearest cluster centroid and obtains grouped clusters as follows:

$$S_i^t = \left\{ D_p \mid \|D_p - \mu_i^t\|^2 \leq \|D_p - \mu_j^t\|^2, \forall j, 1 \leq j \leq k \right\}.$$

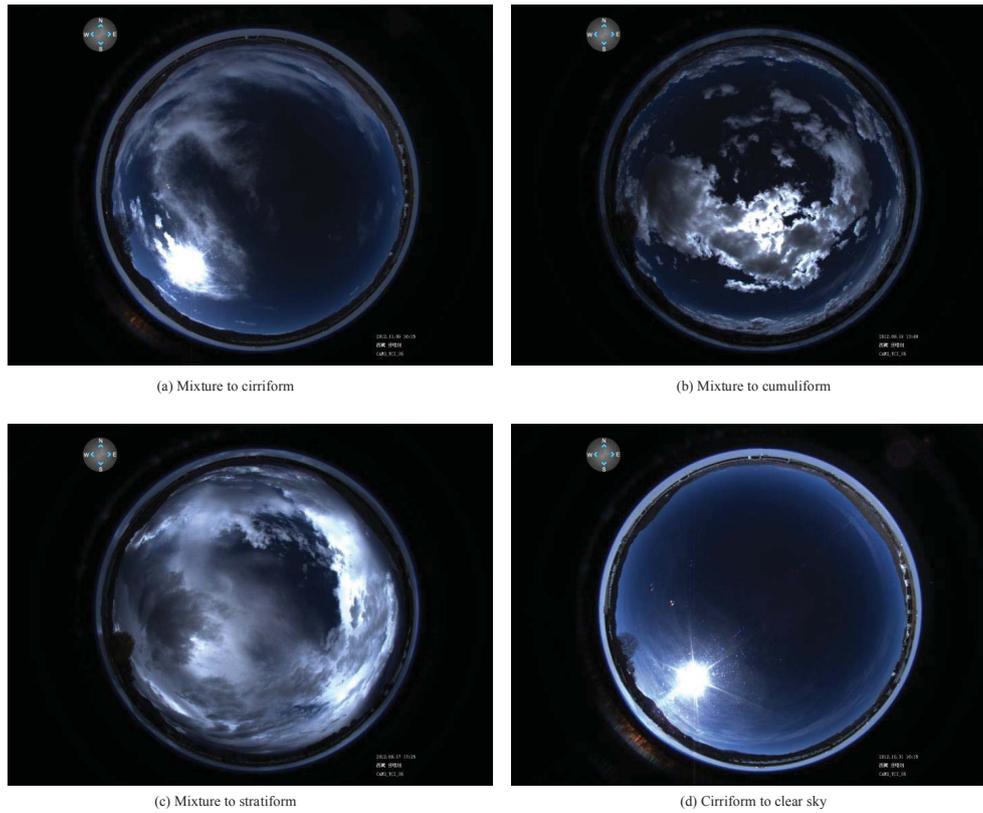


Figure 7. Some misclassified all-sky images.

- *Update step.* It recalculates the centroids of new clusters. A new centroid is updated as follows:

$$\mu_i^{t+1} = \frac{1}{|S_i^t|} \sum_{D_j \in S_i^t} D_j.$$

The algorithm is regarded as converged when the assignments no longer change or the number of iteration has reached a predefined value. After the algorithm is converged, the dictionary is constructed and denoted as $V = \{\mu_1, \mu_2, \dots, \mu_k\}$, where μ_i refers to the prototype of micro-structure with label s_i . The number of clusters k determines the size of the dictionary, which can vary from hundreds to thousands.

3.2.3 Image representation based on BoMS

Image representation refers to the characteristics of an all-sky image and is encoded by a numeric feature vector. Image representation is different from the patch descriptor, but it is built on patch descriptors as shown in Fig. 6. Inspired by the bag-of-words model, we bring forward the “bag of micro-structure” model to extract the image representation.

Firstly, an all-sky image is divided into several patches (see Sect. 3.2.1), and each patch maps to certain type of micro-structure in the dictionary. Assume that the image is

divided into m patches denoted as $\{D_1, D_2, \dots, D_m\}$. $D_i (1 \leq i \leq m)$ is a nine-dimensional vector and is mapped to certain micro-structure label $s_i (\in [1, k])$ by searching for the nearest micro-structure among the dictionary V . Thereby, the descriptor set $\{D_1, D_2, \dots, D_m\}$ is transformed into a label set $\{L_1, L_2, \dots, L_m\}$, whose value of each element refers to the index of a certain micro-structure. The label set is composed of many micro-structures, just as a document consists of words.

After that, we transform the image, which is denoted as a label set, into a representation suitable for the learning algorithm. We apply an attribute-value representation for all-sky images. Basically, each distinct micro-structure s_i corresponds to a feature, assigned with the value t_i that counts the number of occurrences of micro-structure s_i . In order to highlight some important micro-structures, we apply a weighting strategy and represent the image by a vector:

$$F = \langle t_1, t_2, \dots, t_k \rangle. \quad (12)$$

$t_i (1 \leq i \leq k)$ refers to the weighted frequency of micro-structure s_i and is calculated by

$$t_i = n_i \log \frac{N}{N_i}, \quad (13)$$

where n_i is the number of occurrences of micro-structure s_i in the image, N_i is the number of documents containing

s_i , and N is the number of images in the whole data set. In essence, TF (term frequency) and IDF (inverse document frequency) are two major factors in the weighting strategy (Baeza-Yates and Ribeiro-Neto, 1999). TF weights micro-structures more highly if they occur more often in an image, while IDF decreases the weights of the micro-structures that appear often in the data set because these micro-structures do not help to discriminate between different images.

This representation is analogous to the bag-of-words model for a document in terms of form and semantics. Micro-structures reveal characteristics of local patterns for all-sky cloud images, just as words convey meanings of a document. Note that both representations are sparse and high dimensional.

3.3 Classifier with SVM

Support vector machines are based on the structural risk minimization principle from computational learning theory (Vapnik, 2000). Basically, the SVM classifier learns a hyperplane that separates two-class data with maximal margin. The margin is defined as the distance of the closest training sample to the separating hyperplane. Given a training set \mathbf{X} and corresponding class labels \mathbf{Y} that takes value ± 1 , SVM finds a decision function:

$$f(x) = \text{sign}(w^T x + b),$$

where w and b refer to the parameters of the hyperplane.

The SVM applies two strategies to address the data set that is not linearly separable. Firstly, it introduces a regularization term that penalizes misclassification of samples in proportion to their distance from the hyperplane, and this regularization term is weighted by the parameter C . Secondly, a mapping Φ is considered to transform the original data space of \mathbf{X} into another feature space. The feature space may have a high or even infinite dimension. Because SVM can be formulated by the terms of scalar products in the mapped feature space, it is avoidable to directly define such mapping by introducing the *kernel* function $K(u, v) = \Phi(u)^T \times \Phi(v)$. In the kernel formulation, the decision function can be written as

$$f(x) = \text{sign}(y_i \alpha_i K(x, x_i) + b), \quad (14)$$

where x_i is a feature vector from the training set \mathbf{X} and y_i is the label of x_i . The parameters α_i are learned by SVM, and they are typically zero for most i . The feature vectors x_i corresponding to nonzero α_i are known as *support vectors*.

In the cloud classification method, the input features refer to the BoMS representation in Eq. (12). We take the one-against-all approach for the multi-class problem. That is, there are five classes for all-sky images, so we train five SVM classifiers. Classifier model i distinguishes images between category i and all the other four categories. Given a test image, we assign it to the class with the largest SVM output.

There are two reasons motivating us to select SVM rather than other methods, such as KNN and artificial neural net-

works. On the one hand, the image representation based on BoMS is high dimensional (more than 100 dimensions in our experiments). SVM has the potential to handle large feature space because it embraces overfitting protection, and it is more efficient for space and time compared to KNN. On the other hand, the image representation of BoMS is sparse. In other words, the feature vector contains only few entries that are not zero. SVM is proven to be well suited for problems with dense concepts and sparse instances (Melgani and Bruzzone, 2004; Tong and Koller, 2002).

4 Results and discussion

In this section we evaluate the performance of BoMS, compared with the baseline (Heinle et al., 2010), and investigate the effect of the key parameters in BoMS.

We apply 10-fold cross-validation to estimate the performance of classification methods. The data set is randomly partitioned into 10 equal sized subsets. One single subset is retained for validation, and the remaining nine subsets are used as training data. The cross-validation process is then repeated 10 times. During the cross-validation, each subset is used exactly once as validation; meanwhile, each image in the data set is also used for validation exactly once. Finally, the measure of performance is defined by accuracy (Ac), which is given by

$$\text{Ac} = \frac{\text{Correctly classified image number}}{\text{Total image number}}. \quad (15)$$

We conduct each experiment three times and take average values as final results.

4.1 Performance of BoMS

Table 2 demonstrates the confusion matrix of BoMS method, in which the patch size s equals 12 with a step $\tau = 6$ and the dictionary size k equals 500. The kernel of SVM selects linear kernel function, and the regularization C is set to 62.5, which is optimized by a search strategy.

According to Table 2, clear sky is the easiest class to be identified with an accuracy of 99.5%, and it is just a little confused with cirriform because some cirrus is very thin and similar to clear sky. On the contrary, mixed cloudiness is the most difficult class with a low accuracy of 79.5%, and it is misclassified to all other classes. This result is not difficult to be understood since mixed sky condition contains multiple cloud types, and its distribution of micro-structures is easy to be confused with others. Figure 7 displays some misclassified images of mixed cloudiness. Stratiform obtains a good accuracy as well, because it is notably different from cirriform and cumuliform according to color and contrast.

Table 2. Confusion matrix of BoMS.

Ground truth	Classified					
	Cirriiform	Cumuliform	Stratiform	Clear sky	Mixed cloudiness	Ac
Cirriiform	0.872	0.011	0.013	0.023	0.081	
Cumuliform	0.013	0.920	0.007	0.001	0.059	
Stratiform	0.008	0.005	0.965	0.000	0.022	
Clear sky	0.005	0.000	0.000	0.995	0.000	
Mixed cloudiness	0.077	0.092	0.035	0.001	0.795	0.909

Table 3. Comparison of BoMS and baselines according to accuracy of each class.

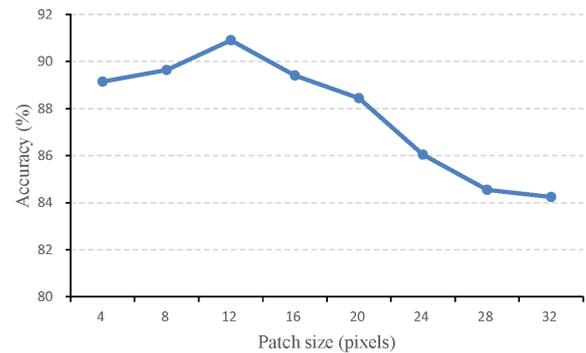
Methods	Cirriiform	Cumuliform	Stratiform	Clear sky	Mixed cloudiness	Overall
F12+KNN	0.438	0.780	0.892	0.939	0.515	0.713
F12+SVM	0.525	0.843	0.922	0.904	0.274	0.694
BoMS _{F12}	0.853	0.896	0.952	0.997	0.736	0.887
BoMS	0.872	0.920	0.965	0.995	0.795	0.909

4.2 Performance comparisons with the baselines

In order to verify the advantage of BoMS, we compare its classification performance with the following three methods:

- F12+kNN is the original method proposed by Heinle et al. (2010), and it acts as the baseline in this comparison. We set k with 9 by an optimized search procedure and set the weight vector of features with [1, 1, 1, 1, 1, 1, 2, 1, 2, 2, 3, 1], which is suggested by Heinle et al. (2010).
- F12+SVM method applies the 12-dimensional features used in (Heinle et al., 2010) but replaces kNN with SVM, in order to investigate the influence of classifier. The kernel of SVM selects linear kernel as well.
- BoMS_{F12} method applies the framework of BoMS, but its patch descriptor is replaced by Heinle's 12-dimensional features, rather than 9-dimensional features described in Sect. 3.

Table 3 presents the comparison results. BoMS outperforms all other methods. Especially, BoMS outperforms F12+KNN with regards to all five classes and achieves an increase of 19% overall accuracy. There are two main differences between BoMS and F12+KNN: feature representation and classifier model. What makes sense for such improvement? We first compare F12+KNN and F12+SVM and observe that F12+SVM does not make an improvement. This result indicates that SVM does not lead to better performance just based on traditional features. The comparison between F12+SVM and BoMS_{F12} shows that BoMS_{F12} is notably better than F12+SVM. So it can be deduced that the image representation based on BoMS contributes to the excellent performance.

**Figure 8.** The curve of overall accuracy of BoMS with different patch sizes.

4.3 Parameter analysis

In the framework of BoMS, patches are the fundamental objects, and they are mainly determined by patch size s . What is the influence of parameter s ?

Figure 8 displays the accuracy curve of BoMS according to different s . Note that the sampling step τ is set to $s/2$ in this experiment. Generally, the accuracy for most values of patch size s is greater than 86%, and smaller patch size can result in better accuracy. Especially, we get the best performance setting s with 12. This result reveals that structure information of image patches is significant for cloud classification, but patches with too small size maybe do not encode enough structure information. Meanwhile, large patches have too large variations to construct efficient micro-structures.

The patch number of an all-sky image is partly determined by sampling set τ . Sampled patches would overlap if τ were smaller than s , and a smaller τ results in more patches. Our

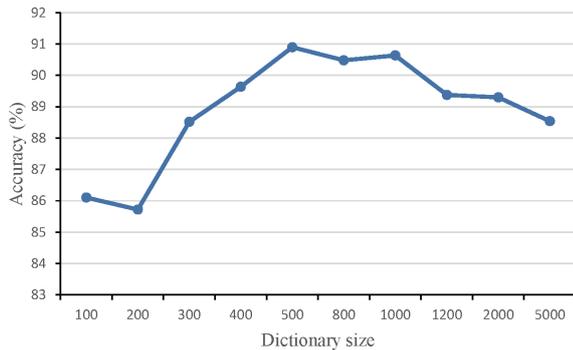


Figure 9. The curve of overall accuracy of BoMS with different dictionary sizes.

experiment results show that classification accuracy is robust for τ , and it is the best choice to set τ with $s/2$, regarding the accuracy and computational complexity.

The dictionary of micro-structures is another core factor for BoMS, and its size k not only determines the dimension of the image representation but also influences the classification performance. Figure 9 displays the accuracy curve of BoMS according to different dictionary size k . Generally, BoMS achieves stable performance with an accuracy greater than 88%, when dictionary size k is more than 300. Especially, BoMS obtains the best performance with $k = 500$. We can account for this result with two aspects. Firstly, a small dictionary of micro-structures just contains limited distinct patterns, since one micro-structure represents one common pattern that is shared by many image patches. As a result, the BoMS with small dictionary is not discriminative enough for cloud classification. Secondly, the dictionary would be saturated with micro-structures when its size is large enough, and a proper micro-structure would be divided into multiple sub-patterns if the dictionary size further increased. However, such sub-patterns cannot promote classification performance.

Moreover, the dimension of BoMS equals with size k . In other words, larger dictionary size results in a higher dimension of cloud representation. Consequently, a medium dictionary size is a good choice, considering the computational complexity.

5 Conclusions

This study presents the new cloud classification method based on a bag of micro-structures, whereas most state-of-the-art methods (Heinle et al., 2010; Liu and Zhang, 2015; Kliangsuwan and Heednacram, 2015; Cheng and Yu, 2015) apply traditional features based on pixels. In this method, an all-sky image is treated as a collection of micro-structures just as a document consists of words, and it is represented by a high-dimensional histogram of micro-structures. Subsequently, the SVM classifier is used to identify the cloud

type of the all-sky image. A large data set is constructed with actual all-sky images captured by the TCI located in Tibet (29.25° N, 88.88° E), and evaluation is carried out to verify the performance of BoMS. The experiment results demonstrate that BoMS achieves a high accuracy of 90.9%, and it outperforms the state-of-the-art method proposed by Heinle et al. (2010). Moreover, the experiments on the influence of key parameters, including patch size s and dictionary size k , are carried out to verify the robustness of BoMS.

We will extend our research in future from the following aspects. Firstly, we will extensively investigate different patch descriptors and find out more efficient patch representation for BoMS. Secondly, we are going to study topic models for cloud images in order to reduce the dimension of cloud representation and further improve classification accuracy. Lastly, we will conduct research to establish a configuration of sky conditions, which is suitable for automatic cloud classification systems.

Acknowledgements. This work is partly supported by the National Key Scientific Instrument and Equipment Development Projects of China (2012YQ11020504), Fundamental Research Funds for the Central Universities (2014JBZ003), the National Natural Science Foundation of China (41105121), and the Basic Research Fund of Chinese Academy of Meteorological Sciences.

Edited by: C. von Savigny

References

- Ameur, Z., Ameur, S., Adane, A., Sauvageot, H., and Bara, K.: Cloud classification using the textural features of Meteosat images, *Int. J. Remote Sens.*, 25, 4491–4503, 2004.
- Baeza-Yates, R. and Ribeiro-Neto, B.: *Modern Information Retrieval*, ACM Press, Addison Wesley, USA, 82 pp., 1999.
- Calbo, J. and Sabburg, J.: Feature extraction from whole-sky ground-based images for cloud-type recognition, *J. Atmos. Ocean. Techn.*, 25, 3–14, 2008.
- Cheng, H.-Y. and Yu, C.-C.: Block-based cloud classification with statistical features and distribution of local texture features, *Atmos. Meas. Tech.*, 8, 1173–1182, doi:10.5194/amt-8-1173-2015, 2015.
- Han, J., Kamber, M., and Pei, J.: *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco, CA, USA, 401 pp., 2006.
- Heinle, A., Macke, A., and Srivastav, A.: Automatic cloud classification of whole sky images, *Atmos. Meas. Tech.*, 3, 557–567, doi:10.5194/amt-3-557-2010, 2010.
- Howard, L.: *On the Modifications of Clouds*, 3rd edn., J. Taylor, London, UK, 28 pp., 1803.
- Hu, X., Wang, Y., and Shan, J.: Automatic recognition of cloud images by using visual saliency features, *IEEE Geosci. Remote S.*, 12, 1760–1764, doi:10.1109/LGRS.2015.2424531, 2015.
- Huang, Y., Wu, Z., Wang, L., and Tan, T.: Feature coding in image classification: a comprehensive study, *IEEE T. Pattern Anal.*, 36, 493–506, 2014.

- Jayadevan, V. T., Rodriguez, J. J., and Cronin, A. D.: A New Contrast-Enhancing Feature for Cloud Detection in Ground-Based Sky Images, *J. Atmos. Ocean. Techn.*, 32, 209–219, 2015.
- Joachims, T.: Text categorization with support vector machines: learning with many relevant features, in: Proceedings of the 10th European Conference on Machine Learning, Chemnitz, Germany, 21–23 April 1998, Springer-Verlag Berlin Heidelberg, Germany, 137–142, 1998.
- Kassianov, E., Long, C. N., and Ovtchinnikov, M.: Cloud sky cover versus cloud fraction: whole-sky simulations and observations, *J. Appl. Meteorol.*, 44, 86–98, 2005.
- Kazantzidis, A., Tzoumanikas, P., Bais, A., Fotopoulos, S., and Economou, G.: Cloud detection and classification with the use of whole-sky ground-based images, *Atmos. Res.*, 113, 80–88, 2012.
- Kliangsuwan, T. and Heednacram, A.: Feature extraction techniques for ground-based cloud type classification, *Expert Syst. Appl.*, 42, 8294–8303, doi:10.1016/j.eswa.2015.05.016, 2015.
- Li, Q., Lu, W., and Yang, J.: A hybrid thresholding algorithm for cloud detection on ground-based color images, *J. Atmos. Ocean. Techn.*, 28, 1286–1296, 2011.
- Li, Q., Lu, W., Yang, J., and Wang, J. Z.: Thin cloud detection of all-sky images using Markov random fields, *IEEE Geosci. Remote S.*, 9, 417–421, 2012.
- Liu, L., Sun, X. J., Chen, F., Zhao, S. J., and Gao, T. C.: Cloud classification based on structure features of infrared images, *J. Atmos. Ocean. Techn.*, 28, 410–417, 2011.
- Liu, S. and Zhang, Z.: Learning discriminative salient LBP for cloud classification in wireless sensor networks, *Int. J. Distrib. Sens. N.*, 501, 327290, 1–7, 2015.
- Liu, S., Wang, C., Xiao, B., Zhang, Z., and Shao, Y.: Salient local binary pattern for ground-based cloud classification, *Acta Meteorol. Sin.*, 27, 211–220, 2013.
- Long, C. N., Sabburg, J. M., Calbó, J., and Pagès, D.: Retrieving cloud characteristics from ground-based daytime color all-sky images, *J. Atmos. Ocean. Techn.*, 23, 633–652, 2006.
- Mantelli Neto, S. L., von Wangenheim, A., Pereira, E. B., and Comunello, E.: The use of Euclidean geometric distance on RGB color space for the classification of sky and cloud patterns, *J. Atmos. Ocean. Techn.*, 27, 1504–1517, 2010.
- Melgani, F. and Bruzzone, L.: Classification of hyperspectral remote sensing images with support vector machines, *IEEE T. Geosci. Remote*, 42, 1778–1790, 2004.
- Ricciardelli, E., Romano, F., and Cuomo, V.: Physical and statistical approaches for cloud identification using Meteosat Second Generation-Spinning Enhanced Visible and Infrared Imager Data, *Remote Sens. Environ.*, 112, 2741–2760, 2008.
- Shields, J., Johnson, R., Karr, M., and Wertz, J.: Automated day/night whole sky imagers for field assessment of cloud cover distributions and radiance distributions, in: Tenth Symposium on Meteorological Observations and Instrumentation, 11–16 January 1998, Phoenix, Arizona, USA, American Meteorological Society, 11–16, 1998.
- Singh, M. and Glennen, M.: Automated ground-based cloud recognition, *Pattern Anal. Appl.*, 8, 258–271, 2005.
- Stephens, G. L.: Cloud feedbacks in the climate system: a critical review, *J. Climate*, 18, 237–273, 2005.
- Tahir, A.: A system based on ratio images and quick probabilistic neural network for continuous cloud classification, *IEEE T. Geosci. Remote*, 49, 5008–5015, 2011.
- Tapakis, R. and Charalambides, A.: Equipment and methodologies for cloud detection and classification: a review, *Sol. Energy*, 95, 392–430, 2013.
- Tong, S. and Koller, D.: Support vector machine active learning with applications to text classification, *J. Mach. Learn. Res.*, 2, 45–66, 2002.
- Urquhart, B., Kurtz, B., Dahlin, E., Ghonima, M., Shields, J. E., and Kleissl, J.: Development of a sky imaging system for short-term solar power forecasting, *Atmos. Meas. Tech.*, 8, 875–890, doi:10.5194/amt-8-875-2015, 2015.
- Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer Science and Business Media, New York, USA, 123 pp., 2000.
- Xia, M., Lu, W., Yang, J., Ma, Y., Yao, W., and Zheng, Z.: A hybrid method based on extreme learning machine and k-nearest neighbor for cloud classification of ground-based visible cloud image, *Neurocomputing*, 160, 238–249, doi:10.1016/j.neucom.2015.02.022, 2015.
- Yang, J., Lu, W., Ma, Y., and Yao, W.: An automated cirrus cloud detection method for a ground-based cloud image, *J. Atmos. Ocean. Techn.*, 29, 527–537, doi:10.1175/JTECH-D-11-00002.1, 2012.
- Yang, J., Min, Q., Lu, W., Yao, W., Ma, Y., Du, J., Lu, T., and Liu, G.: An automated cloud detection method based on the green channel of total-sky visible images, *Atmos. Meas. Tech.*, 8, 4671–4679, doi:10.5194/amt-8-4671-2015, 2015.
- Zhang, J., Marszałek, M., Lazebnik, S., and Schmid, C.: Local features and kernels for classification of texture and object categories: a comprehensive study, *Int. J. Comput. Vision*, 73, 213–238, 2007.