

Classification of Cloud Types Using Partial Least Square Regression

Nandhini K¹, Dr.G.Tamilpavai²

Government college of Engineering, Tirunelveli, Tamilnadu

Abstract- An accurate identification of clouds from the captured sky/cloud image using ground-based sky cameras is difficult, due to the fuzzy boundaries of clouds, uncertainties in the cloud etc. Several techniques have been proposed that use color as the discriminatory feature for cloud detection. The objective of this paper is to classify cloud image types using various features of sky/cloud images captured from ground-based Whole Sky Imagers (WSIs). Identification of cloud types, finds its scope in climate modeling, solar energy generation, air-to-ground communications etc. The feature set of sky/cloud images are uncertain, incomplete, highly spatial and temporal. Sky/Cloud image classification is done in three steps. (i) Principal Component Analysis (PCA) is used for optimizing the feature set of sky/cloud images. (ii) Partial Least Regression analysis is used for classification of types of clouds (iii) A comparative study using Support Vector Machine (SVM) classifier is performed with the same dataset. The proposed work identifies (i) Pattered cloud (ii) Thick dark cloud (iii) Thick white cloud and (iv) Veil cloud. The experimental results are verified with Singapore Whole sky IMaging CATegories (SWIMCAT) database and achieved the classification accuracy with PLS regression is 97.14% and with SVM classifier 93.33%.

Keywords- Principal Component Analysis, Whole Sky Imagers, Partial Least Square Regression, Support Vector Machine Classifier, Pattered cloud, Thick dark cloud, Thick white cloud, Veil cloud, SWIMCAT database.

I. INTRODUCTION

Clouds appear in variety of forms over the sky and it is continuously changing. Such cloud images are difficult to capture and monitor. Sky imaging devices such as Whole Sky Imagers (WSIs) is a ground-based electronic imaging system that can be used to capture upper hemisphere of the sky images. The observed sky images are in hyper-spectral band containing high spatio-temporal features. From the captured sky images, it is possible to find cloud fractional coverage, sky polarization, the computation of cloud base height, as well as wind speed at cloud heights, as well as the presence, distribution, shape, and radiance of clouds over the entire sky. The data associated with such sky images are incomplete, vague, and unclear. Hence, feature set need to be optimized for

representation of such uncertain multi-dimensional sky images.

Feature set help the image processing to work with great computational efficiency and enhanced performance. As sky imaging contains high dimensionality of the data, learning from such data is highly inefficient. Moreover, the presence of noise degrades the performance of the algorithms. Therefore, it is necessary to select discriminative features from the input imagery. The captured sky images have issues in extracting features such as altitude related information of cloud due to limited spatial resolution and unknown surface influences on measured radiances. Additionally there exist different views in cloud features associated with satellite and sensors. And due to downward-pointing satellite sensor, there is a solar obstruction in capturing the sky images. In order to overcome these issues, the dataset need to be reduced using rough set reduction principle.

WSI produces images with dynamically changed data of cloud formations at regular intervals of time. There are many algorithms for computing the fraction of the sky covered by clouds, identifying cloud types, finding cloud base height etc. It is easy to identify the type of cloud based on blue/red ratio. Statistical features obtained from red and blue channels are subsequently used for classification of cloud types. Hence, a structured review of color channels is essential to deal with the hyper spectral data of sky images and for recognizing the cloud type using feature matching in cloud base height estimation. This can efficiently represent the images in a lower dimensional subspace. Principal component analysis (PCA) is used to determine favorable channels. PCA provides maximum reparability among the interclasses. And such dimensionality reduction procedure enhances the performance of classification.

Principal Component Analysis is used to determine the structure of an image represented by the 16 color channels and analyze the inherent correlations amongst these

-
1. Nandhini.K is a Post Graduate Scholar in Computer Science and Engineering department at Government College of Engineering, Tirunelveli (:nandhukk28@gmail.com)
 2. Dr.G.Tamilpavai is an Associate Professor(Senior Gr.) & Head in Computer Science and Engineering Department in Government College of Engineering, Tirunelveli (tamilpavai@gcetly.ac.in)

components. The individual color channels c1–16 are extracted for the image and reshaped into column vectors. Normalize with mean and standard deviation of individual color channels of the image. Subsequently, compute the covariance matrix, that provides an insight of how different color channels are correlated. Then compute the Eigen values and Eigen vectors that supports in notifying the contribution of each color channel to orthogonal axes. These Eigen vectors corresponds to the principal component and it refers to the loading factor of the corresponding color channel.

Partial least squares regression (PLS regression) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of maximum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. PLS regression is particularly suited when the matrix of predictors has more variables than observations, and when there is multicollinearity among X values.

II. RELATED WORK

Many distance metrics are available in order to estimate closeness of feature set that can be further used for classification of clouds in sky images. Among the available metrics, Kullback-Leibler (KL) distance measure plays a high scope in band selection in hyperspectral imaging. For sky/cloud segmentation using color channels of hyperspectral images. Bimodality [2] and Principal component analysis [3] are used to select the favorable channel. Yet, there are certain issues like correlation among the color channels during cloud segment results in poor accuracy. Recently, hybrid thresholding algorithm [4] is used for cloud detection from ground-based sky images that uses the normalized red/blue ratio cloud detection.

Color feature plays a vital role in cloud segmentation of hyperspectral sky images; Color features are discriminating features that identifies the cloud in the sky. A. Heinle, et.al., uses the ratio of red and blue channels from RGB color space for segmentation and to create binary masks [5]. Another method which uses atmospheric pressure [6] to find the difference of red and blue channels for successful detection and subsequent labeling of pixels. Super pixel-based segmentation [7] uses the difference of red and blue channels of the color channel of sky images. Saturation(S) [8] channel is used for calculating cloud coverage and thereafter segmentation is applied over the color channel for extracting the discriminating color channel of sky images. Q. Li et. al., generate a model that used the locus [9] of cloud pixels in

RGB color channel extracted from the several color channels of sky images used to detect the existence of cloud.

An adaptive threshold technique [10] is used in normalized red / blue channel in order to discriminate the presence of cloud in sky images. From the extracted feature set of sky images, superpixel based classification [11] techniques provided good accuracy of classification of types of clouds. For the process of cloud segmentation, several color models are constructed. Using the color models, several color channels are observed. Rayleigh scattering of light [12][13][14] at shorter wavelengths are applied in order to select the user choice of color channel. But in this method, manually-defined parameters and thereafter decision making process based on the parameters shows the efficiency of cloud segmentation well. But these methods are error-prone. After segmentation, labels are assigned for each classification and these algorithm shows as the best, flexible and robust.

S. Dev, et al., proposes a method that encodes the textural features of the input images, and using Gabor- and edge-based texture filters are used, e.g. for aerial imagery [15] or landscape image segmentation [16] and such segmentation is more accurate. The same method is modified by a set of Schmid filters for the task of cloud classification [17].

Statistical tools and techniques [18][19][20] are used for best cloud segmentation. The color channels used are the red channel R of RGB color space, the Saturation channel S of HSV, and the red-blue ratio R/B. The image database that is publically available for sky/cloud segmentation is the HYTA database [21]. It consists of 32 images capturing several sky/cloud scenarios under varying illumination conditions, and the corresponding binary segmentation ground-truth images.

In the process of cloud detection of all-sky image, a pixel-based scheme [23] [24][25][26] is used. Using this method, red-to-blue ratio (RBR) of each pixel is used. Next a threshold is applied to RBR which in turn determines cloud pixels. The pixels whose RBRs are lower than the threshold are classified as clear sky and the pixels whose RBRs are higher the threshold are labelled as clouds. Selecting a good threshold is very important for RBR method. In addition relative position of the pixel is also considered along with the positions of sun and horizon. Then neural network [27] is used as the best classifier to classify the types of clouds in sky images.

Prototype selection chooses a representative patterns from the data that guides in reducing the training data. This is NP-hard problem [28], because there is no polynomial

algorithm allowing for the solution. The existing algorithms produces acceptable solutions.

A Support Vector Machine (SVM) [29] based cloud classification model using the high temporal and spatial resolution sky images captured. Firstly, the influence on irradiance under clouds of different shapes and distributions in a sky image is analyzed and four different classes of clouds are distinguished taking into account the meteorology standard as well as the preceding analysis. Secondly, the spectral and textural features are extracted by the statistical tonal analysis and gray level co-occurrence matrix (GLCM) of the sky image. Finally, a support vector classification model with radial basis function (RBF) kernel function is built to classify the different clouds in the sky images. However, because of the lack of the feature of cloud cover ratio, the model faces some difficulties to distinguish clouds. it fails to capture the high variability of the cloud system dynamics.

The most important feature that is used for the segmentation purpose deals with Color extraction [32]. This work includes a macro physical observation which classifies the cloud type and cloud cover. The prototype Hemispheric Sky Imager (HSI), which was the precursor to the commercial TSI, was developed at the National Oceanic and Atmospheric Administration Surface Radiation Research Branch (SRRB). The development was a joint effort between SRRB and the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) Program. This instrument was developed into a commercial product named the TSI in a cooperative effort with YES under a Small Business Innovative Research grant from the U.S. Department of Agriculture. The sky images are 24-bit color JPEG format at 352 × 288 pixel resolution captured from a camera based on a commercially available digital camera.

III. DATA AND METHODS





Dataset – SWIMCAT (Singapore Whole sky IMaging CATegories Database)

Public benchmark database related to sky/cloud images are rare. This proposed work uses SWIMCAT database for classification of cloud types in sky images.

The SWIMCAT database contains images captured in Singapore using ground-based whole sky imager over a period of 17 months from January 2013 to May 2014. This database includes 784 images of sky/cloud patches with five categories namely clear sky, patterned clouds, thick dark clouds, thick white clouds, and veil clouds. Each patch has a

dimension of 125x125 pixels. The categories of cloud as per SWIMCAT dataset are listed in the table, Table 3.1.

Table 3.1 Four categories of clouds in SWIMCAT database

Patterned clouds	Thick dark clouds	Thick white clouds	Veil clouds
			

The entire work in this paper is depicted in Figure 3.1. The proposed work accepts the sky/cloud images as input. The input images are highly spectral and multi-dimensional. Feature set is extracted using color space and component model. The analysis of extracted features shows that the data are time variant with many hidden patterns; hence the data are claimed for further processing. Using Principal Component Analysis (PCA), the dataset is optimized to an optimal representation. Optimized feature set are fed to the Partial Least Square Regression model classifier. Available data is divided as training and testing purpose. A strong set of features provides high discriminatory power, reducing the need for complex classification methods. The PLS regression improves the speed and accuracy of classification.

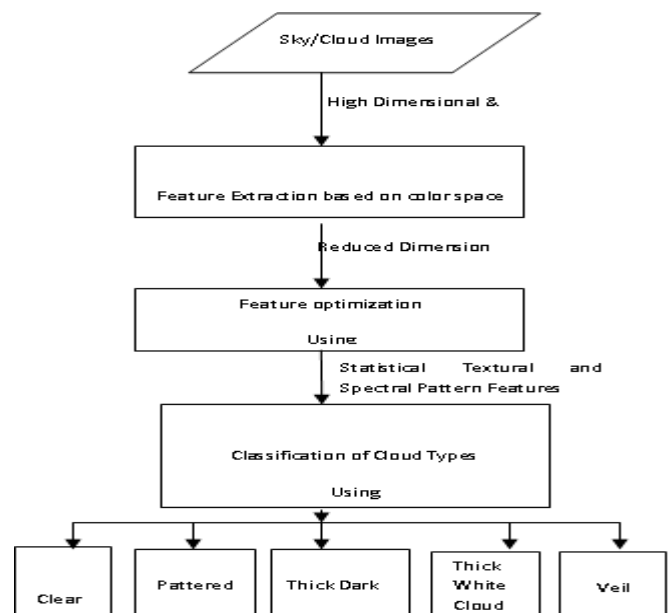


Figure 3.1 Overall block diagram of the proposed work

Feature optimization using Principal Component Analysis

The input sky/cloud images are hyperspectral and multi-dimensional. The data are time-variant, with missing

patterns, partially incomplete, and highly dynamic in nature. Hence, for further processing of such input images, the entire feature set need to be optimized. Among the various Dimensionality reduction methods, Principal Component Analysis [11] produces accurate result in eliminating the redundancy of dataset and improve performance of classifiers. Being a supervised learning algorithm, Principal component analysis (PCA) accepts hyperspectral images using optimized feature set.

Linear transformation is applied to obtain the probability of neighborhood using PCA.

$$P_{ij} = \frac{[\exp(-\|A_{xi} - A_{xj}\|_2)]}{\sum_{k \neq i} \exp(-\|A_{xi} - A_{xk}\|_2)}$$

$$P_{ii} = 0$$

where P_{ij} represents the probability of pixels ‘i’ & ‘j’ and A represents the linear transformation.

Totally seventeen feature sets which contains statistical, texture, shape and pattern features are used for processing. Hence, it creates 17-dimensional representation. Dataset after dimensionality reduction using PCA is optimized with 12 feature set, that is more appropriate for the process of feature extraction and classification

.Feature Extraction

A detection window, d_i , is used into overlapping blocks and extract a set of features for each block to construct the feature vector v_i . Using cooccurrence matrix, perform the texture analysis. Co-occurrence matrices represent second order texture information which is the joint probability distribution of gray-level pairs of neighboring pixels in a block.

Twelve more descriptors such as angular second-moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, and directionality are extracted. Co-occurrence features are useful in cloud detection as they provide homogeneity and directionality of patches.

Edge information is extracted using histograms of oriented gradients. HOG captures edge or gradient structures that are characteristic of local shape. Since the histograms are computed for regions of a given size within the detection window, HOG is robust to some location variability of several regions. HOG is also invariant to rotations smaller than the orientation bin size.

Finally, color is extracted. Although colors may not be consistent due to variability, certain dominant colors are more significant for certain regions of cloud detection. In order to incorporate color, extract color frequency from the original HOG. In HOG, the orientation of the gradient for a pixel is chosen from the color band corresponding to the highest gradient magnitude. Some color information is captured by the number of times each color band is chosen. Therefore, construct a three bin histograms that tabulates the number of times each color band is chosen.

Table 3.2 Original image, RGB image, Grayscale, Intensity image and Spectral power

Whole Image	RGB Image	Grayscale image	Intensity Image	Spectral Power

Partial Least Square Regression

A strong set of features provides high discriminatory power, reducing the need for complex classification methods. Cloud images have distinguishing characteristics. First, strong vertical edges are present along the boundaries of the image. Second, natural textures are difficult to observe. Third, the ground is composed mostly of uniform textures. Finally, discriminatory color information is found in the many regions. Thus, edges, colors and textures capture are considered as important cues for discriminating cloud from the background. To extract these features, the low-level features the original HOG descriptors with additional color information are obtained, and are referred to as color frequency, and texture features computed from co-occurrence matrices.

To handle the high dimensionality resulting from the combination of features, PLS is employed as a dimensionality reduction technique. PLS is a powerful technique that provides dimensionality reduction for even hundreds of thousands of variables, accounting for class labels in the process. The steps performed in our detection method are the following. For each detection window in the image, features extracted using original HOG, color frequency, and co-occurrence matrices are concatenated and analyzed by the PLS model to reduce dimensionality, resulting in a low dimensional vector. Then, a simple and efficient classifier is used to classify this vector.

Partial Least Square Regression for Dimension Reduction

Partial least square regression is a method for modeling relations between sets of observed variables by means of latent variables. The basic idea of PLS is to construct new predictor variables, latent variables, as linear combinations of the original variables summarized in a matrix X of descriptor variables (features) and a vector y of response variables (class labels). While additional details regarding PLS methods, a brief mathematical description of the procedure is provided below.

Let $X \subset \mathbb{R}^m$ denote an m-dimensional space of feature vectors and similarly let $Y \subset \mathbb{R}$ be a 1-dimensional space representing the class labels. Let the number of samples be n. PLS decomposes the zero-mean matrix X ($n \times m$) and zero-mean vector y ($n \times 1$) into

$$X = TP^T + E$$

$$y = Uq^T + f$$

where T and U are $n \times p$ matrices containing p extracted latent vectors, the $(m \times p)$ matrix P and the $(1 \times p)$ vector q represent the loadings and the $n \times m$ matrix E and the $n \times 1$ vector f are the residuals. The PLS method, using the nonlinear iterative partial least squares (NIPALS) algorithm, constructs a set of weight vectors (or projection vectors) $W = \{w_1, w_2, \dots, w_p\}$ such that

$$[\text{cov}(t_i, u_i)]^2 = \max |w_i|=1 [\text{cov}(Xw_i, y)]^2$$

where t_i is the i-th column of matrix T, u_i the i-th column of matrix U and $\text{cov}(t_i, u_i)$ is the sample covariance between latent vectors t_i and u_i . After the extraction of the latent vectors t_i and u_i , the matrix X and vector y are deflated by subtracting their rank-one approximations based on t_i and u_i . This process is repeated until the desired number of latent vectors had been extracted. The dimensionality reduction is performed by projecting the feature vector v_i , extracted from a detection window d_i , onto the weight vectors $W = \{w_1, w_2, \dots, w_p\}$, obtaining the latent vector z_i ($1 \times p$) as a result. This vector is used in classification.

The algorithm for Partial Least Square Regression is given below

```
function PLS1(X, y, l)
    X(0) ← X
    w(0) ← XTy/||XTy||, an initial estimate of w.
    for k = 0 to l - 1
        t(k) ← X(k)w(k)
        tk ← t(k)Tt(k) (note this is a scalar)
        t(k) ← t(k)/tk
        p(k) ← X(k)Tt(k)
        qk ← yTt(k) (note this is a scalar)
        if qk = 0
            l ← k, break the for loop
        if k < (l - 1)
            X(k+1) ← X(k) - tkt(k)p(k)T
            w(k+1) ← X(k+1)Ty
    end for
    define W to be the matrix with columns w(0), w(1), ..., w(l-1).
    Do the same to form the P matrix and q vector.
    B ← W(PTW)-1q
    B0 ← q0 - P(0)TB
    return B, B0
```

The extracted feature sets from SWIMCAT dataset are shown in the Table 3.3

Table 3.3 Feature vectors of SWIMCAT database.

Features	Clear Sky	Pattern Cloud	Thick Dark Cloud	Thick white cloud	Veil Cloud
Sum Average	96.1	86.3	92.6	91.5	81.7
	93.3	83.6	89.8	88.7	79
Sum Variance	96.0	92.0	92.5	91.4	87.4
	94.5	90.8	91	89.9	86.2
Variance	95.5	87.0	92	90.9	82.4
	93.9	83.1	90.4	89.3	78.5
Inverse difference moment	95.5	85.8	92	90.9	81.2
	94.8	81.2	91.3	90.2	76.6
Entropy	96.1	88.3	92.6	91.5	83.7
	94.6	84.2	91.1	90	79.6
Sum Entropy	95.9	86.7	92.4	91.3	82.1
	93.3	83.3	89.8	88.7	78.7
Correlation	1.0	1.0	0.9	0.92	0.88
	0.9	0.92	0.88	0.87	0.82
Angular second-moment	0.006	0.013	0.002	0.003	0.001
	0.017	0.012	0.015	0.002	0.001
Contrast	0.00	0.00	0.33	0.58	0.74
	0.00	0.17	0.24	0.52	0.71
Difference Variance	0.05	0.13	0.04	0.06	0.05
	0.06	0.15	0.02	0.01	0.07
Difference Variance	96.1	86.1	92.6	91.5	81.5
	92.3	87.7	88.8	87.7	83.1
Difference Entropy	96.2	80.4	92.7	91.6	75.8
	94.9	85.4	91.4	90.3	80.8
Directionality	74.5	81.2	90.3	91.2	82.6
	76.2	80.9	91.2	90.7	81.3

IV. EXPERIMENTAL EVALUATION AND RESULTS

This work extracted spectral, textural and shape features of the sky/cloud images from the SWIMCAT database. The images of entire database is categorized as training, testing and validation images in each class and is shown in Table 4.1. The images are used for training, testing and validation in the ratio of 50:25:25 percentage from the dataset.

The number of training images are fixed after optimization of input dataset. Optimization determines the relevant feature set that are used for classification. The test images can be chosen in a random way.

Table 4.1 Training and testing images of SWIMCAT database

Sl.No.	Type of Cloud	No. of		
		Training Images	Testing Images	Validation Images
1.	Clear Sky	150	75	75
2.	Pattern Cloud	49	25	25
3.	Thick dark cloud	176	88	88
4.	Thick white cloud	181	91	91
5.	Veil Cloud	50	25	25
Total Images		606	304	304

The above feature values are adequate to perform the classification. However, to validate the correctness of classification, a contingency matrix is formed for the SWIMCAT dataset as shown in Table 4.2

Table 4.2 Confusion matrix of classification with SWIMCAT dataset

	Clear Sky	Pattern Cloud	Thick Dark Cloud	Thick white cloud	Veil Cloud	Total	Precision
Clear Sky	127	0	5	6	2	150	90.7
Pattern Cloud	3	35	0	6	5	49	71.4
Thick Dark Cloud	2	2	165	3	4	176	93.7
Thick white cloud	0	3	3	170	5	181	93.9
Veil Cloud	1	5	4	2	38	50	75
Recall	95.4	77.7	93.2	90.9	70.3	606	

Finally, the testing phase included 30 images which are trained and classified using PLS regression and SVM. They are categorized according to the classification of the 5 classes based on the SWIMCAT database.

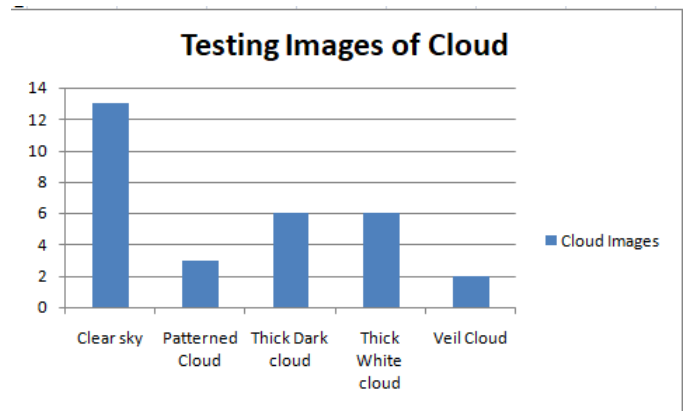


Figure 4.1 Testing Phase of Cloud images

In this work, both PLS regression and SVM were performed. The classification is made using the training and testing phase of the images. In Training phase, 70% of the images are trained and remaining 30% of the images are tested. According to the classification result, PLS regression gives better performance than SVM which is described in Figure 4.2.

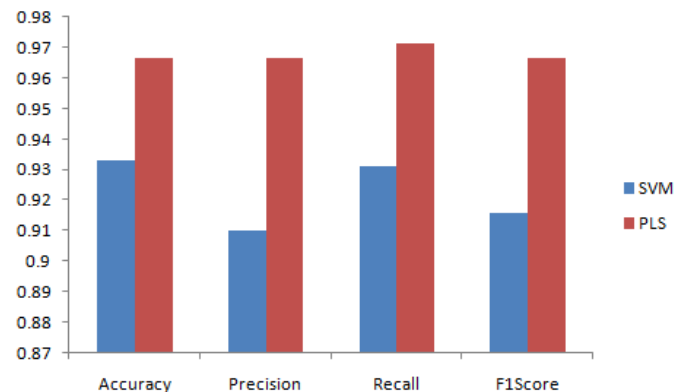


Figure 4.2 Performance Measure

V. CONCLUSION

In this work, a systematic analysis of color spaces and components, and proposed a probabilistic approach using PLS-based regression for the classification of ground-based sky/cloud images. This approach is entirely learning-based and does not require any manually-defined thresholds, conditions, or parameters at any stage of the algorithm. This work includes an extensive sky/cloud image database captured with a calibrated ground-based camera that has been annotated with ground truth mask. The future work will include the annotation of a database with probabilistic ground-truth segmentation maps as well as the extension of this method to High-Dynamic-Range (HDR) images. Going beyond segmentation, it is also important to classify clouds into different types or estimate cloud altitude and movement. This work leads to misinterpretation of pixels near the sun when

there is change in the position of solar disk. And hence in future to rectify this misinterpretation geometrical features have to be included for the classification process. This framework can be extended for evaluation using the benchmark dataset Spinning Enhanced Visible and InfraRed Imager (SEVIRI) and in near future it can be evaluated with real dataset and types of clouds can be extended. Eleven new cloud classifications such as the volutus, or roll cloud, as well as the asperitas cloud, the flumen, "cataractagenitus", "flammagenitus", "homogenitus", "silvagenitus" etc. discovered in March 2017 can also be identified in future using our framework.

REFERENCES

- [1] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clusteringbased hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp.4158–4171, Dec. 2007.
- [2] S. Dev, Y. H. Lee, and S. Winkler, "Color-based segmentation of sky/cloud images from ground-based cameras," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. PP, no. 99, pp. 1–12, 2016.
- [3] 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, pp. 5102–5106.
- [4] 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, May 2006.
- [5] A. Heinle, A. Macke, and A. Srivastav, "Automatic cloud classification of whole sky images," *Atmospheric Measurement Techniques*, vol. 3, no. 3, pp. 557–567, May 2010.
- [6] 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, Feb. 2015.
- [7] 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. 3, no. 5, pp. 437–447, March 2006.
- [8] S. L. Mantelli-Neto, A. von Wangenheim, E. B. Pereira, and E. Comunello, "The use of Euclidean Geometric Distance on RGB color space for the classification of sky and cloud patterns," *Journal of Atmospheric and Oceanic Technology*, vol. 27, no. 9, pp. 1504–1517, Sept. 2010.
- [9] 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, Oct. 2011.
- [10] Y. Yuan and X. Hu, "Bag-of-words and object-based classification for cloud extraction from satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 8, pp. 4197–4205, Aug. 2015.
- [11] J. Calbó and J. Sabburg, "Feature extraction from whole-sky groundbased images for cloud-type recognition," *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 1, pp. 3–14, Jan. 2008.
- [12] 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.
- [13] J. Shao and W. Foerstner, "Gabor wavelets for texture edge extraction," in *Proc. SPIE ISPRS Commission III Symposium: Spatial Information from Digital Photogrammetry and Computer Vision*, 1994, vol. 2357.
- [14] 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, pp. 716–723.
- [15] S. Dev, Y. H. Lee, and S. Winkler, "Categorization of cloud image patches using an improved texton-based approach," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2015.
- [16] 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, 2011.
- [17] 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.
- [18] 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.
- [19] 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, 2011.
- [20] Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., and Washom, B.: Intra-hour forecasting with a totalsky imager at the UC San Diego solar energy testbed, *Sol. Energy*, 85, 2881–2893, 2011.

- [21] Johnson, R., Hering W., and Shields, J.: Automated visibility and cloud cover measurements with a solid-state imaging system. Tech. Rep., Scripps Institution of Oceanography, Marine Physical Laboratory, University of California, San Diego, USA, SIO Ref. 89-7, GL-TR-89-0061, 128, 1989.
- [22] Long, C. N., Sabburg, J., Calbó, J., and Pagès, D.: Retrieving cloud characteristics from ground-based daytime color all-sky images, *J. Atmos. Ocean. Tech.*, 23, 633–652, 2006.
- [23] Shields, J., Karr, M., Burden, A., Johnson, R., and Hodgkiss, W.: Continuing support of cloud free line of sight determination including whole sky imaging of clouds, Final Report for ONR Contract N00014-01-D-0043 DO #13, Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, USA, Technical Note 273, 2007.
- [24] Roy, G., Hayman, S., and Julian, W.: Sky analysis from CCD images: cloud cover, *Lighting Res. Technol.*, 33, 211–222, 2001.
- [25] West, S. R., Rowe, D., Sayeef, S., and Berry, A.: Short-term irradiance forecasting using skycams: Motivation and development, *Sol. Energy*, 110, 188–207, 2014.
- [26] Huo, J. and Lu, D.: Cloud determination of all-sky images under low visibility conditions, *J. Atmos. Ocean. Tech.*, 26, 2172–2180, 2009.
- [27] J. E. Shields, T. L. Koehler, and R. W. Johnson, “Whole sky imager,” in *Proceedings of the Cloud Impacts on DOD Operations and Systems, 1989/90 Conference (Science and Technology Corporation, Meetings Division, 101 Research Drive, 1990)*, pp. 123–128.
- [28] A. V. Zukhba, “Np-completeness of the problem of prototype selection in the nearest neighbor method”, *Pattern Recognition and Image Analysis*, vol. 20, pp. 484-494, 2010.
- [29] Zhao Zhen, Fei Wang, Yujing Sun, Chun Liu, Bo Wang, Jing Lu, “SVM Based Cloud Classification Model Using Total Sky Images for PV Power Forecasting”, *IEEE Power and Energy Society*, 2015
- [30] Zhengping Wu, Xian Xu, Min Xia, Meifang Ma, Lin Li, “Ground-based Vision Cloud Image Classification based on Extreme Learning Machine”, *The Open Cybernetics and Systemics Journal*, 2015
- [31] Yu Liu, Jun Xia, Chun-Xiang Shi, Yang Hong, “An Improved Cloud Classification Algorithm for China’s FY-2C Multi-channel Images Using Artificial Neural Network”, *Sensors*, 2009.
- [32] C. N. Long, J. M. Sabburg, J. Calbó, and D. Pages, “Retrieving cloud characteristics from ground-based daytime color all-sky images”, *J. Atmospheric Ocean. Technol.*, vol. 23, no. 5, pp. 633–652, May 2006.