MULTI-LEVEL SEMANTIC LABELING OF SKY/CLOUD IMAGES

Soumyabrata Dev, Yee Hui Lee

School of Electrical and Electronic Engineering
Nanyang Technological University (NTU)
Singapore 639798

Stefan Winkler

Advanced Digital Sciences Center (ADSC)
University of Illinois at Urbana-Champaign
Singapore 138632

ABSTRACT

Sky/cloud images captured by ground-based Whole Sky Imagers (WSIs) are extensively used now-a-days for various applications. In this paper, we learn the semantics of sky/cloud images, which allows an automatic annotation of pixels with different class labels. We model the various labels/classes with a continuous-valued multi-variate distribution. Using a set of training images, the distributions for different labels are learnt, and subsequently used for labeling test images. We also present a method to determine the number of clusters. Our proposed approach is the first for multi-class sky-cloud image annotation and achieves very good results.

Index Terms— Clustering, likelihood estimation, ground-based sky imaging

1. INTRODUCTION

The study of clouds and analysis of their features are important in several applications such as climate modeling, weather prediction, solar energy generation, or air-to-ground communications [1, 2]. Nowadays, cloud analysis is extensively performed using images obtained from ground-based sky cameras [3]. These are able to provide sky/cloud images with high spatial and temporal resolution.

Semantic segmentation is an important part of the analysis of sky/cloud images. It is challenging because clouds do not possess any definite structure, shape, size or color. Traditionally, images are manually annotated with different labels. However, manual annotation is cumbersome and imprecise, which is why automatic semantic labeling approaches are needed.

In this paper, we propose a probabilistic framework for annotating image regions as *thin clouds*, *thick clouds*, or *clear sky*. Using a set of training images with manual labels, we compute the joint probability distribution of the different class labels. We assume that the images can be segmented into non-overlapping regions, and different pixels belonging to the same class can be clustered together. Moreover, in most clustering methods, the number of clusters is either defined manually or known a-priori. We propose a method to automatically determine whether a sky/cloud image is single- or multi-cluster. Finally, we evaluate our proposed labeling approach on a sky/cloud image database with two- and three-level annotations.

This research is funded by the Defence Science and Technology Agency (DSTA), Singapore.

2. RELATED WORKS

Performing automatic annotation of images is a common problem in computer vision. A lot of related works have been carried out to address it for various applications.

The co-occurrence model is a popular approach for image annotation. Each image is divided into regions with keywords, and the correlation between each region and keyword are analyzed [4, 5]. Several other approaches [5–7] have been proposed which assume the entire data distribution to be dependent on a few latent variables. These latent variables can be clusters, features, subspaces of high-dimensional data. Once the latent variables are learnt from a set of training images, the labeling framework is implemented.

In generating the different subspace transformations or creating the various discriminating features of an image, the number of clusters is either assumed a-priori or estimated using a set of inherent parameters. It is very important to determine the correct number of clusters for the specific application. Amongst the various techniques introduced in the literature, gap techniques appear promising and have been used in a variety of applications. Zhang and Zhu [8] have proposed the entropy gap statistic that considers the change in entropy of the dataset. Yang et al. [9] calculated the gap value per pixel in extracting edge information from an image.

Previous works on cloud segmentation proposed several techniques to obtain *binary* segmented images for various sky/cloud images [10–13]. We proposed a *probabilistic* annotation of cloud/sky pixels in our previous work [14]. It allows us to assign a value of *belongingness* to the cloud or sky category for each pixel. We extend this idea to multi-level labeling in this paper, where the different labels can be learnt in a natural way, and new images can be annotated automatically.

3. MULTI-LEVEL SEMANTIC ANNOTATION

We model the pixels in the sky/cloud image as a mixture of continuous multi-variate distributions. In a formal setting, we automatically estimate the set of model parameters that fits the distribution of different labels in the image – clear sky, thin cloud and thick cloud. These labels are learnt from the different non-overlapping regions of the image, as described in the following.

3.1. Generation of Feature Vectors

An important part for image annotation is the generation of discriminative features which facilitate classification. It has been observed that color is the dominant feature for detecting cloud pixels [10, 13]. In our previous work [14], we have provided an extensive analysis of 16 different color channels for the task of sky/cloud image segmentation. Using statistical tools and techniques, we identified the color channels that work best for cloud segmentation. They are the red channel R of RGB color space, the Saturation channel S of HSV, and the red-blue ratio R/B.

Suppose that a sample *RGB* image in the dataset is represented by $\mathbf{I}_i \in \mathbb{R}^{a\times b\times 3}$, having a dimension $a \times b$ for each of *R*, *G* and *B* channels. Using the appropriate color channels for segmentation [14], we generate the probability map $\mathbf{H}_i$, $(i = 1, 2, .., N)$ for all *N* images of database $\mathcal{S} = \{\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_N\}$. Our objective is to convert these probability maps into three-level semantic regions. The steps for generating this probability map and subsequently the input feature vector in the model are the following:

- For each image $\mathbf{I}_i$, the color channels R, S and R/B are extracted to form the concatenated matrix $\mathbf{F}_i = [R, S, R/B]$.
- Fuzzy c-means clustering is applied on $\mathbf{F}_i$. The output generated by this step is a probability map $\mathbf{H}_i \in \mathbb{R}^{a\times b}$, which denotes the probability of a pixel belonging to a given cloud category.
- $\mathbf{H}_i \in \mathbb{R}^{a\times b}$ is reshaped into a column vector $\mathbf{x}_i \in \mathbb{R}^{ab\times 1}$; and $\mathbf{x}_i$ serves as the feature vector for our semantic segmentation approach.

After the generation of these input feature vectors $\mathbf{x}_i$ for all training images, the model parameters of the multi-variate Gaussian distribution for three labels are computed.

3.2. Model for Sky/Cloud Images

The task of multi-level semantic annotation can be modelled as a discrete labeling problem, wherein each image region $\boldsymbol{\xi} = \{\xi_1, \xi_2, ...\}$ of the image $\mathbf{I}_i$ is annotated with its corresponding class labels. Assuming non-overlapping regions, a pixel can possess a single class label from the set of labels $\boldsymbol{\Omega} = \{\Omega_1, \Omega_2, \Omega_3\}$. This section introduces a statistical model to map different image regions of an image $\mathbf{I}_i$ to different labels from the set $\boldsymbol{\Omega}$ by modelling the joint probability distribution $P(\boldsymbol{\xi}, \boldsymbol{\Omega})$.

Consider a set $\mathcal{S} = \{\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_N\}$ consisting of *N* sky/cloud images representing varying sky/cloud scenarios. The labels Ω_1, Ω_2 and Ω_3 represent clear sky, thin clouds, and thick clouds respectively, for example:

$$\Omega_1 = \begin{cases} 1 & \text{if pixel is a } \textit{clear sky} \\ 0 & \text{otherwise} \end{cases} \qquad \text{etc.}$$

This essentially generates *three* categories – *clear sky*, *thin cloud*, and *thick cloud* – each with pixels consisting of both positive and negative samples. By this definition, all the pixels of an image are categorically distributed into disjoint labels. Assuming that the labels generated from set $\boldsymbol{\Omega}$ are independent and identically distributed from a multi-variate Gaussian distribution, we can learn this joint probability distribution $P(\boldsymbol{\xi}, \boldsymbol{\Omega})$. For example, from a set of manually annotated sky/cloud images, we can learn how a patch of *clear sky* looks like. In the same manner, we can model the distribution of *thick clouds* and *thin clouds*. Subsequently, the labels for a test image can be predicted accordingly.

3.3. Estimation of Model Parameters

From the set of training images, we estimate the input discriminating feature $\mathbf{x}$. Assuming $\mathbf{x}$ to be a continuous random variable, we use Gaussian Discriminant Analysis to model the different class labels using multivariate normal distributions. For example, the *clear sky* label can be modeled as follows:

$$\Omega_1 \sim \text{Bernoulli}(\alpha)$$
$$(\mathbf{x}|\Omega_1 = \psi) \sim \mathcal{N}(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}),$$

where Ω_1 follows a Bernoulli distribution with the parameter α, and $\psi = 0$ and $\psi = 1$ denote negative and positive samples for label Ω_1, respectively. The negative samples for label Ω_1 follow a normal distribution with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}$. Similarly, the distribution for positive samples have mean vector $\boldsymbol{\mu}_1$ and covariance matrix $\boldsymbol{\Sigma}$. The distributions can be written as:

$$p(\Omega_1) = \alpha^{\Omega_1}(1-\alpha)^{1-\Omega_1}$$

$$p(\mathbf{x}|\Omega_1 = \psi) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_\psi)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_\psi)\right)}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}$$

describing the parameters for both positive ($\psi = 1$) and negative ($\psi = 0$) samples of clear sky label Ω_1.

The likelihood for a given feature vector to belong to this

label is given by the log-likelihood of the data:

$$l(\alpha, \boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}) = \log \prod_{i=1}^{m} p(\mathbf{x}^{(i)}, \Omega_1^{(i)}; \alpha, \boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\Sigma})$$
$$= \log \prod_{i=1}^{m} p(\mathbf{x}^{(i)} | \Omega_1^{(i)}; \boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}) p(\Omega_1^{(i)}; \alpha)$$

The maximum likelihood parameters α, $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}$ are estimated by maximizing the log-likelihood l of the data for the set of annotated training images.

Models for the other two labels *thick cloud* (Ω_2) and *thin cloud* (Ω_3) are computed in exactly the same fashion. And thus, in entirety, we generate three different multi-variate Gaussian distributions consisting of both positive and negative samples for each of the categories.

4. DETERMINING THE NUMBER OF CLUSTERS

The number of clusters in the observed data space is an important pre-requisite for image annotation and clustering. This is because some sky/cloud images do not need any segmentation in the first place (e.g. because they show only sky). We distinguish between two cases, single- and multi-cluster. If an image is single-cluster, there is no need for segmentation. Otherwise, the probability map is converted into a 3-level labeled image.

In our proposed approach, we use the S color channel for determining the type of an image. From our previous work [14], we observe that the Saturation channel (S) of the HSV color model has the highest bimodality, making it a good candidate for determining cluster number. Supposing the inter-sample distance between two samples i and j in the saturation channel is denoted by $d_{ij} = \sum(S_i - S_j)^2$. By calculating these distances for all possible pairs in the observation space, we can represent the sum of all pairwise distances as $D = \sum_{i,j}^{N} d_{ij}$. For the observation data space, we calculate the within-class dispersion $W = \frac{D}{2N}$ (assuming all data points belong to a *single* cluster). A higher dispersion indicates the likely presence of multiple clusters.

From a set of single-cluster training images in the database, the average value $\overline{W}$ is computed. If $W < \overline{W}$ for a given the test image, it is considered single-cluster (*clear sky* only) and does not need any further segmentation; otherwise it is considered a multi-cluster image.

5. EXPERIMENTAL EVALUATION & RESULTS

Currently, the only public image database for sky/cloud segmentation is the HYTA database [10]. It consists of 32 images capturing several sky/cloud scenarios under varying illumination conditions, and the corresponding binary segmentation ground-truth images. However, in this paper, we evaluate the performance for a higher number of semantic labels. Therefore, instead of binary labels, we manually annotate all the images of the HYTA database using three category labels – *clear sky*, *thin cloud* and *thick cloud* – in consultation with cloud experts from Singapore's Meteorological Services.

In order to perform an objective evaluation of our proposed approach, we report Precision, Recall and F-score values. Suppose that the true positive, true negative, false positive and false negative samples of a binary image are represented by *TP*, *TN*, *FP* and *FN* respectively. The precision and recall scores are calculated as:

$$\begin{aligned} \text{Precision} &= TP/(TP + FP), \\ \text{Recall} &= TP/(TP + FN). \end{aligned} \tag{1}$$

In addition, we also report the F-score, which is the harmonic mean of precision and recall:

$$\text{F-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2}$$

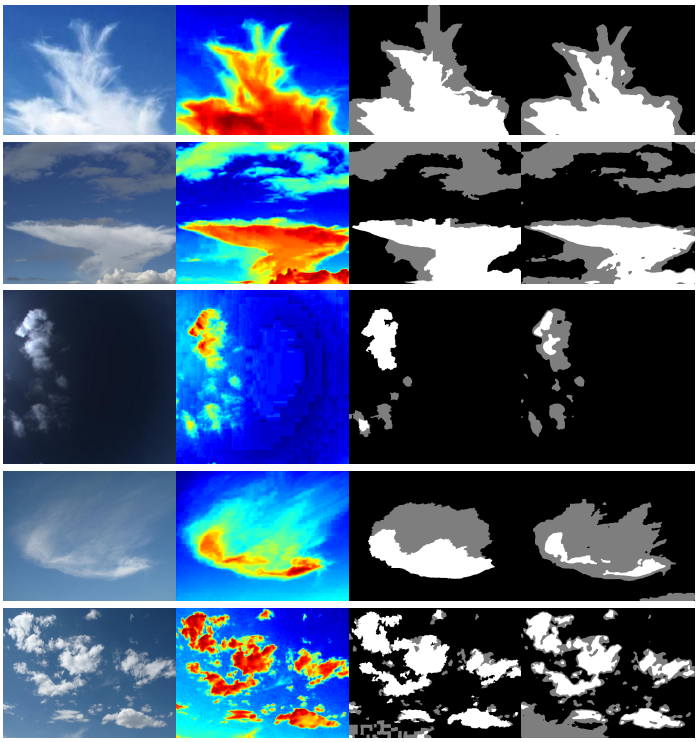

Fig. 1: Five image samples, from left to right: Original image from HYTA database; probabilistic cloud detection using [14]; 3-level ground-truth; 3-level labels generated using our proposed approach.

5.1. Number of Clusters

In HYTA, two images exhibiting a single cluster (*clear sky*) are selected to determine the average value of within-class

dispersion. The value of $\overline{W} = 49.35$ is found as per Section 4. This serves as the threshold in evaluating the number of clusters for images in HYTA database.[1] Using the remaining 30 images in the database as the test set, all are correctly identified as either single- or multi-cluster with this approach.

5.2. Multi-level Semantic Labels

The entire HYTA database is divided equally into training and testing sets. The training set consists of 16 images, from which the different parameters α, $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}$ of the multivariate distributions are learnt individually for each of the three class labels. These 16 training images are chosen randomly from the dataset. We have observed that the classification accuracy is not significantly affected by the specific choice and the number of training images.

During testing stage, the log-likelihood estimate for three different classes for all pixels of the remaining 16 test images are evaluated as described in Section 3.3. The label having the maximum log-likelihood estimate will be chosen as the final label for a particular pixel. In this manner, the output 3-level annotation is generated. Fig. 1 shows the results obtained using our proposed approach, alongside the input image with its manually annotated 3-level ground truth image and probabilistic map. The generated 3-level output image matches well with the corresponding ground truths.

For a more objective evaluation, precision, recall, and F-score for *clear sky*, *thin cloud* and *thick cloud* are computed over all images in the test set. Fig. 2 shows these results. While our method lacks precision especially for thin clouds, it achieves near-perfect results for clear sky and thick clouds. Thin clouds are the most problematic because of their imprecise edges between sky and cloud regions; owing to their low-contrast, they are sometimes misclassified as clear sky or thick clouds.

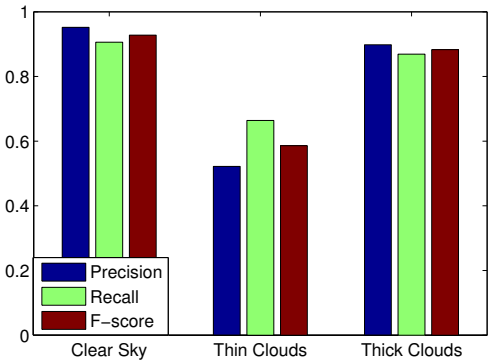

Fig. 2: Evaluation results for 3-level semantic labeling of the images in the HYTA database.

[1] As the HYTA dataset contains no images with completely overcast sky, we were not able to determine appropriate thresholds for that case.

5.3. Binary Semantic Labels

As there are no other methods for multi-level classification of sky/cloud images, we apply our proposed approach to check its efficiency for binary labels, in order to provide a comparative analysis with other state-of-art cloud detection algorithms. The model parameters as described in Section 3.3 are thus learnt for binary labels (viz. sky and cloud). Finally, the pixels of the test images are labeled using maximum log-likelihood estimation.

We compare our proposed approach of semantic segmentation with the current state-of-the-art cloud segmentation algorithms [10–13] on the HYTA database. Fig. 3 shows the results of the different algorithms based on their corresponding precision, recall and F-scores. Based on F-score values, our proposed approach surpasses the performance of other algorithms. The numbers also put in perspective the results shown in Fig. 2, which compare favorably in terms of absolute F-scores.

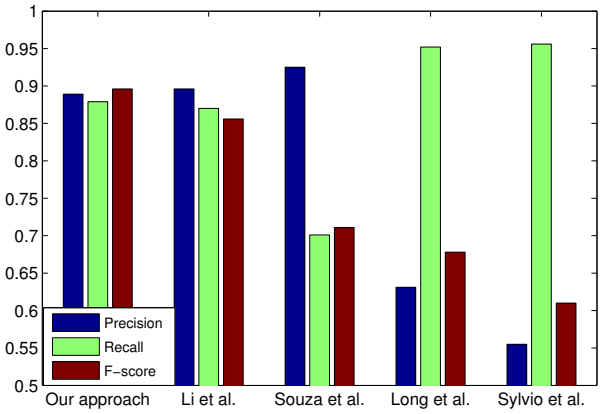

Fig. 3: Evaluation results for binary semantic labeling of the images in the HYTA database.

6. CONCLUSION

In this paper, we have introduced a systematic framework whereby sky/cloud images captured by a ground-based camera are automatically annotated into various labels. Unlike conventional binary labeling in the task of sky/cloud image segmentation, we presented an approach for multi-level labeling. Benchmarking on a publicly available database confirms the effectiveness of our proposed approach. Future work will focus on the rectification of remaining misclassifications of thin clouds and the creation of a larger database for evaluation. We also plan to provide a more fine-grained categorization of clouds for better image interpretation, as done in [15] for small image patches.

7. REFERENCES

[1] J. X. Yeo, Y. H. Lee, and J. T. Ong, "Performance of site diversity investigated through RADAR derived results," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 10, pp. 3890–3898, 2011.

[2] 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.

[3] 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.

[4] Y. Mori, H. Takahashi, and R. Oka, "Image-to-word transformation based on dividing and vector quantizing images with words," in *Proc. International Workshop on Multimedia Intelligent Storage and Retrieval Management (MISRM)*, 1999.

[5] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," in *Proc. European Conference on Computer Vision (ECCV)*, 2002.

[6] K. Barnard and D. Forsyth, "Learning the semantics of words and pictures," in *Proc. International Conference on Computer Vision (ICCV)*, 2001, vol. 2, pp. 408–415.

[7] P. Carbonetto, N. de Freitas, and K. Barnard, "A statistical model for general contextual object recognition," in *Proc. European Conference on Computer Vision (ECCV)*, 2004.

[8] Z.-J. Zhang and Y.-Q. Zhu, "Estimating the image segmentation number via the entropy gap statistic," in *Proc. 2nd International Conference on Information and Computing Science*, 2009.

[9] Q. Yang, L. Tang, W. Dong, and Y. Sun, "Image edge detecting based on gap statistic model and relative entropy," in *Proc. 4th International Conference on Fuzzy Systems and Knowledge Discovery*, 2009.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.