Laurel Garrett

Lewis & Clark College

August 5, 2015

Automated Cloud Classification from Total Sky Images.

**Introduction**

Clouds play an important role in the climate of our atmosphere and earth. This is because clouds reflect incoming solar radiation, and absorb and re-emit terrestrial radiation. The occurrence and composition of clouds depends on local temperature, humidity and meteorological conditions as well as aerosol content. Different cloud types have a different net heating effect on the atmosphere and surface, therefore it is important to be able to differentiate cloud types from observations. However, systematic ground observations of cloud type were largely discontinued in the mid-1990's. Furthermore, there is often disagreement and discontinuity between observations depending on the instrument and location of observation. Since having a human observer is not feasible for a study of this scale, we must rely on cloud imaging devices. We would like to construct an automated cloud classifier to distinguish cloud types for one of these devices; the Total Sky Imager (TSI).

**Instrument and data**

The TSI is a camera pointed down at a hemispheric mirror that reflects the whole sky (Figure 1). Our dataset comes from the atmospheric radiation measurement (ARM) program. While ARM has field stations worldwide, our dataset is from the Southern Great Plains site near Lamont, Oklahoma, where a TSI has been taking photos of the sky every 30 seconds since July, 2000. The data includes original whole-sky images that are slightly compressed into JPG format, as well as a processed image that has been masked and thresholded to show clear sky, thin clouds, and opaque clouds. There is also a data file saved in netcdf format for each day that includes the clear, thin, and opaque pixel counts for different regions of the decision image.

**Preprocessing of TSI images**

Even though we were provided with masked and thresholded images, we noticed major errors that would introduce many artifacts in our data. We decided to replicate and improve their mask and threshold before moving on. The mask went through a number of iterations where the center of the image had to be manually adjusted to account for slight movements in the TSI, the options to include a sun circle and horizon area mask were added, and the sun arm and instrument arm were made fatter. This version was used for image processing, but still introduces some minor artifacts. Another version of the mask was developed that erodes and opens the mask if there are dark neighboring pixels. This mask has the capabilities to mask out dirt on the mirror but slows down the run time.

There are also many discontinuities in the data set. The data switched from 352 by 288 pixels to 640 by 480 pixels in 2011 so we had to process these sets of images differently. The naming convention changed over time, also complicating the image processing. We renamed the latest images, but mostly developed our code to adjust to the different naming conventions in order to preserve the data set as it was downloaded from the ARM site. As discussed before, the center of the images may shift over time due to changes in the camera rig. In the centroid lookup table we also included a list of dates that do not exist in the data or are otherwise compromised.

**Processing image statistics**

My part of this research was to replicate and build on the classification techniques used in other cloud classification papers which have not yet been employed for the TSI. Following Calbo and Sabburg (2008), the images are converted to the red to blue ratio, which makes it easier to capture cloud information. I calculated the Mean (ME) Standard Deviation (SD), Third Moment (TM), Uniformity (UF), Entropy (EY), and Smoothness (SM) from the R/B image histogram. Using Matlab’s built in image gradient function, I also used the median magnitude of the gradient (MG). These statistics can be used to classify sky images in to clear, overcast, and mixed clouds, shown in Figure 2.

First, a training set was created of 120 images: 40 clear sky, 40 mixed sky, and 40 overcast skies. The 7 statistics were computed for each of these 120 training images (Figure 2, panels a, b, c). These images are then used to train the classifiers, to subsequently classify an image as clear, mixed, or overcast.

I ran the statistics from 07/03/2000 to 05/30/2015 selecting one image every 30 minutes. These were separated into two sets to be classified: the low resolution set from 2000 to 2011 and the high resolution set from 2011 to 2015. There were 91,937 images from the low resolution set and 32,833 from the high resolution set. Both sets of statistics were normalized to have values ranging from zero to one using the minimum and maximum value of each statistics vector for each (low resolution or high resolution) data set.



The statistics for this mixed sky image appears as a black star on subplots (a,b,c).

**Classifiers**

I tested two classifiers, a support vector machine (SVM) and a k-nearest neighbor (KNN). The SVM performed best in Zhou et al.’s analysis of classifiers. The KNN classifier performed second best but it seems to be a common and trusted classifier used in Heinle et al. (2010). For each image in the test set, the KNN code loads the statistics from the training set (eg. 120 images) and calculates the Euclidean distance to the point with the closest statistics. It then assigns the test image the classification of the the single nearest neighbor. That is, k=1 in our implementation. The SVM code loads both sets and uses a built in SVM matlab function to find the optimal plane between the classification of the training set.

I made two training sets of images to train the classifiers. The low resolution set has 50 clear sky images, 50 mixed clouds images, and 50 overcast sky images. The high resolution set has 40 clear sky images, 40 mixed clouds images, and 40 overcast sky images.

I then ran both the KNN and the SVM on the test and training images. To check the accuracy of the classifiers, I randomly selected 600 hundred images from 2000 to 2011 and 600 from 2011 to 2015 that three checkers independently marked with 

0=CLR,

1=borderline clear (eg. 0-⅛ cloudy),

2=MIXED,

3=borderline ovc (eg. ⅞ - 100% cloudy),

4=OVC,

8=rain,

9=UNSURE.

I left out any rainy images, compromised images, and any where the checkers didn’t agree on the classification. Which left 398 checked images out of the 2000 to 2011 set and 425 from the 2011 to 2015 set. For our comparison, we combined the borderline clear cases with the clear cases and the borderline overcast cases with the overcast cases.

**Results and discussion**

The results for the SVM classifier and the KNN were almost identical. In the high resolution set, we had a overall 89.65% agreement between the algorithm and the manual classifications. The clear images had the best agreement with 99% of the images that humans classified as clear, the computer also classified them as clear (Figure 3). This was followed closely by overcast with 96.7% and mixed with 73.7%. The error in the mixed clouds is most likely with thin clouds which are largely transparent to the blue sky. The low resolution results looked suspicious and need to be debugged. It is likely that the lower resolution camera picked up more red in the images and the statistics did not accurately describe this. We also left out images that the checkers disagreed on before combining the border cases. We could potentially revise the methodology to have a larger sample of checked images. Now that we can accurately identify clear and overcast images, the next step is to further segregate the mixed images into cloud types.

**Works Cited**

Calbó, Josep, and Jeff Sabburg. “Feature Extraction from Whole-Sky Ground-Based Images for Cloud-Type Recognition.” *Journal of Atmospheric and Oceanic Technology* 25, no. 1 (January 2008): 3–14. doi:10.1175/2007JTECHA959.1.

Heinle, A., A. Macke, and A. Srivastav. “Automatic Cloud Classification of Whole Sky Images.” *Atmos. Meas. Tech.* 3, no. 3 (May 6, 2010): 557–67. doi:10.5194/amt-3-557-2010.

Zhuo, Wen, Zhiguo Cao, and Yang Xiao. “Cloud Classification of Ground-Based Images Using Texture–Structure Features.” *Journal of Atmospheric and Oceanic Technology* 31, no. 1 (November 6, 2013): 79–92. doi:10.1175/JTECH-D-13-00048.1.