

ATS 670 Class Project

Supervised Classification and Unsupervised Classification

Xiong Liu

Abstract: This project use migrating means clustering unsupervised classification (MMC), maximum likelihood classification (MLC) trained by picked training samples and trained by the results of unsupervised classification (Hybrid Classification) to classify a 512 pixels by 512 lines NOAA-14 AVHRR Local Area Coverage (LAC) image. All the channels including ch3 and ch3t are used in this project. The image is classified to six classes including water, vegetation, thin partial clouds over ground, thin clouds, low/middle thick clouds and high thick clouds plus unknown class for supervised classification. In total, the results using these three methods are very consistent with the original three-band overlay color composite image and the statistical mean vectors for each class are consistent using different methods and are reasonable. We also note that the ch3t temperature is usually much larger than the thermal channel-measured temperature for clouds, the colder the thermal temperature, the larger their difference. The ch3 reflectance is anti-correlated with the ch1 and ch2 reflectance, which is due to that high reflectance ice clouds can absorb most of the energy in this channel. Look carefully, the results of MMC and MLC trained by the results of MMC are better than that of the MMC trained by picked samples. The MLC trained by picked samples produces more unknown classes than that trained by MMC, which is probably due to that the standard deviation (multivariate spreads) for each class generated by MMC is usually larger than that of picked training samples. It takes more computation time to run MMC (5 iterations) than MLC if the classes are the same, but take more time to pick samples over and over to get comparable results. The results of MLC trained by picking samples is worse than the other two methods due to the difficulty of picking representative training samples. The hybrid supervised/unsupervised classification combines the advantages of both supervised classification and unsupervised classification. It doesn't require the user have the foreknowledge of each classes, and can still consider the multivariate spreads and obtain accurate mean vectors and covariance matrixes` for each spectral class by using all the pixels image as training samples.

1. Introduction

One of the main purposes of satellite remote sensing is to interpret the observed data and classify features. In addition to the approach of photointerpretation, quantitative analysis, which uses computer to label each pixel to particular spectral classes (called classification), is commonly used. Quantitative analysis can perform true multispectral analysis, make use of all the available brightness levels and obtain high quantitative accuracy.

There are two broads of classification procedures: supervised classification unsupervised classification. The supervised classification is the essential tool used for extracting quantitative information from remotely sensed image data [Richards, 1993, p85]. Using this method, the analyst has available sufficient known pixels to generate representative parameters for each class of interest. This step is called training. Once trained, the classifier is then used to attach labels to all the image pixels according to the trained parameters. The most commonly used supervised classification is maximum likelihood classification (MLC), which assumes that each spectral class can be described by a multivariate normal distribution. Therefore, MCL takes advantage of both the mean vectors and the multivariate spreads of each class, and can identify those elongated classes. However, the effectiveness of maximum likelihood classification depends on reasonably accurate estimation of the mean vector m and the covariance matrix for each spectral class data [Richards, 1993, p189]. What's more, it assumes that the classes are distributed unimodal in multivariate space. When the classes are multimodal distributed, we cannot get accurate results. Another broad of classification is unsupervised classification. It doesn't require human to have the foreknowledge of the classes, and mainly using some clustering algorithm to classify an image data [Richards, 1993, p85]. These procedures can be used to determine the number and location of the unimodal spectral classes. One of the most commonly used unsupervised classifications is the migrating means clustering classifier (MMC). This method is based on labeling each pixel to unknown cluster centers and then moving from one cluster center to another in a way that the SSE measure of the preceding section is reduced data [Richards, 1993, p231].

This project performs maximum likelihood supervised classification and migrating means clustering unsupervised classification to an AVHRR Local Area Coverage (LAC) Data image, and compares the results of these two methods. In addition, using the results of MMC to train the MLC classifier is also shown and will be compared together.

2. Data

The NOAA AVHRR series are designed to provide information for hydrologic, oceanographic, meteorological and earth studies data [Richards, 1993, p8]. There are five channels in AVHRR data, including visible (0.58-0.68 μm), near infrared (0.76-0.90 μm), mid-infrared (3.53-3.93 μm), two thermal infrared (10.3-11.3 μm , 11.5-12.5 μm) channels. The visible channel detects the solar reflected radiance and

measures the reflectance; the two thermal-infrared channels measure the earth-emitted radiance and therefore indicate the surface temperature. The mid-infrared channel measures both the reflected radiance and the earth-emitted radiance. AVHRR orbits 14 orbits a day, with a swath of 2700 km, and ground resolution at nadir of 1.1 km. It can monitor the whole globe in one day. The spectral channels and high spatial resolution make it able to detect vegetation, soil, water, smoke, forest fire, clouds, fog, and some meteorological phenomena.

A frame of NOAA-14 AVHRR LAC image was ordered from <http://ww.saa.noaa.gov>, which is taken in Indonesia regions on September 8, 1998. The ordered ten-bit data is then calibrated to get 8-bit gray levels and view zenith angles and location. The mid-infrared channel is divided into two parts, the reflected part and the emitted part. Therefore, there are six channels (ch1, ch2, ch3, ch3t, ch4, and ch5) after calibration. All the six channels will be used in this project. The ch3t, ch4 and ch5 channels are first gray-flipped before classification. In order to save computation time, only a subset image of 512 pixels by 512 lines was used to perform MLC and MMC.

3. Methodology

Both MLC and MMC are performed on Pentium III 450MHZ PC.

3.1. Migrating means clustering algorithm

MMC is performed by according to the following basic steps [Richards, 1993, p231-233]:

(1). Determine the number of cluster centers and initialize the cluster centers. The cluster centers are chosen uniformly spaced along the multidimensional diagonal of the multispectral pixel space or the results from supervised classification.

(2). Examine each pixel in the image and assign it to the nearest candidate cluster based on the Euclidean distance. In order to reduce computation time, the Euclidean distance is simplified as follows to get the following discriminant function data [Richards, 1993, p190]:

$$d_i(x) = 2m_i x - m_i \cdot m_i \tag{1}$$

And the assignment is performed based on:

$$x \in i \text{ cluster, if } d_i(x) > d_j(x) \quad \text{for all } j \neq i \tag{2}$$

If there are some clusters with less than 60 pixels, then just delete these cluster centers. The pixels belongs to these clusters with be classified in next iteration.

(3) The new set of cluster centers that result from step (2) is computed. The sum of square of errors (SSE) is also computed as an indicator to terminate the iteration.

$$SSE = \sum_{C_i} \sum_{x \in C_i} \|x - m_i\|^2 \tag{3}$$

(4) If SSE is very close to the previous SSE, the procedure is terminated. Otherwise, the cluster centers are redefined as the current cluster means and iterate step 2-4.

(5) Label each cluster as a certain class and color-encode and display the classified image. Estimate the number of pixels and area for each class. The mean gray level values, physical temperature or reflectance, and corresponding standard deviations are provided by using all the available classified pixels.

3.2. Maximum likelihood supervised classifier

MLC is performed according to the following steps [Richards, 1993, p181-184]:

- (1). Display the three-band overlay composite image. The visible channel, near-infrared channel and the 10.3-11.3 um channel are associated with red, green and blue, respectively so that the clouds look white, vegetation looks green, water looks dark and lands without vegetation looks different shades of brown. Take a careful look at the available features and determine the set of classes into which the image is to be segmented.
- (2). Using 'box-cursor' to choose representative training samples for each of the desired classes from the color composite image. These pixels are said to form training data. By comparison, the results of MMC are also used as training samples.
- (3). Use the training samples to estimate the mean vectors and covariance matrixes for MLC classifier. These two parameters determine the properties of the multivariate normal models.
- (4). Using the trained classifier to classify every pixel in the image into one of the desired classes. Since we have no useful information about the priori probability for each class, in which case a situation of equal prior probabilities is assumed. The final discriminant function $g(x)$ is taken as:

$$g(x) = -\ln(|\Sigma_i|) - (x - m_i)' \Sigma_i^{-1} (x - m_i) \quad (4)$$

Where m_i and Σ_i are the mean vector and covariance matrix of the data in class ω_i . N is the number of bands. In order to reduce poor classification due to small probabilities, threshold values T_i are determined for each class based on that 95% of the pixels would be classified. According to χ^2 tables, the threshold values T_i can be obtained by:

$$T_i = -12.6 - \ln(|\Sigma_i|) \quad (5)$$

Finally, we can get the decision rule for maximum likelihood supervised algorithm:

$$x \in \omega_i, \text{ if } g_i(x) > g_j(x) \text{ and } g_i(x) > T_i \text{ for all } j \neq i \quad (6)$$

Classes don't meet the above decision rule will be classified as unknown class.

- (5). Color-encode and show the classified image. Estimate the number of pixels and area for each class and show the statistics for each class.

4. Result and Discussion

4.1. Three-band overlay color composite image

Figure 1 shows the three-band overlay composite image. Ch1, Ch2 and Ch4 are associated with red, green and blues, respectively. The color scheme looks very attractive, we can clearly identify green vegetation, dark water, large lumps of bright-white thick clouds (high clouds) with light yellow clouds around them, small lumps of yellow thick clouds (low/middle clouds), and fluffy bright thin clouds.

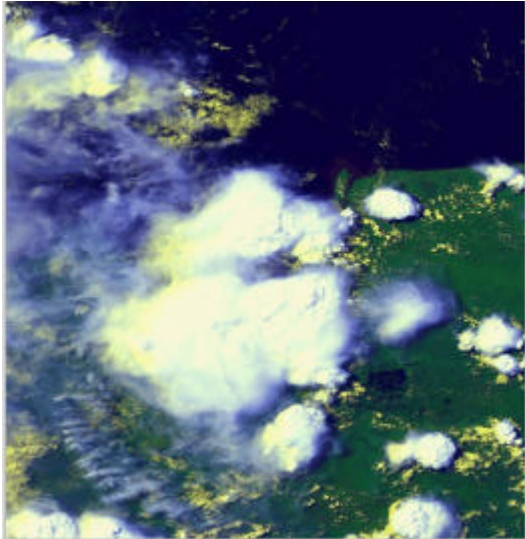


Figure 1. Three bands overlay color composite image. (Ch1, ch2, and ch4 are associated with red, green and blue respectively.)

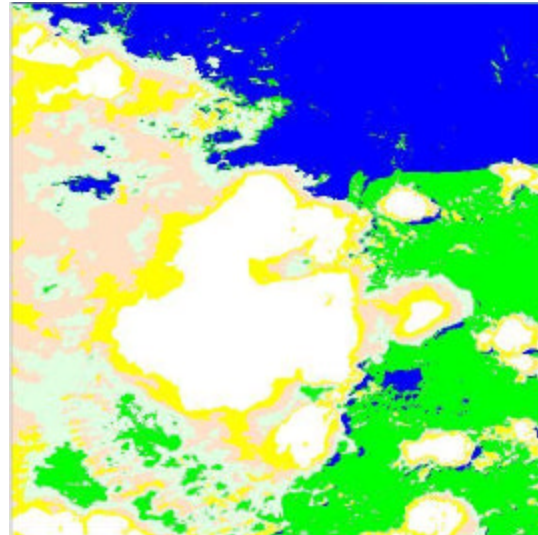


Figure 2. Thematic map of produced by the migrating means clustering classification. Blue represent water and cloud shade, green is vegetation, gray green is thin cloud over ground, pink is thin cloud, yellow is low and middle thick clouds, white is high thick clouds.

4.2. Migrating means clustering classification

Ten initial cluster centers are selected uniformly distributed along the multivariate diagonal line. In the first iteration, only six cluster centers have number of pixels greater than 60. Other four cluster centers are deleted in next iteration. After 20 iterations, the SSE is decreased slowly from initial 3.8787×10^8 to 3.7728×10^8 , and then the process is terminated. The color-encoded thematic map is shown in Figure 2. The results are quite consistent with the original color composite image. Compared with Figure 1, we can identify that the six classes are corresponding to water (blue), green vegetation (green), very thin partial clouds over ground (gray green), thin clouds (pink), low/middle thick clouds (yellow) and high thick clouds (white). The cloud shades are also classified to the same class as water. It is probably due to that ten initial clusters are not enough to differentiate water and cloud shades. The light yellow borders of those bright white clouds and those small lumps of yellow clouds are classified together. The only exception is that the clouds labeled as 'A' in Figure 1 which appears yellow, are mostly classified as thin partial clouds. The statistics after classification including gray level and physical (Reflectance for ch1, ch2 and ch3, Temperature for ch3t, ch4 and ch5) mean vectors, standard deviations and covariance matrix are shown in Table 1, and the tabular summary of the thematic map is shown in Table 4.

Table 1. Class signatures for the results of unsupervised classification

Classes	Mean Vector		Standard Deviation		Covariance Matrix					
	Gray Level	R% or T (K)	Gray Level	R% or T (K)						
Water/ Cloud shade	13.15	5.16	6.13	2.40	37.59	37.31	19.86	5.50	31.85	34.59
	8.57	3.36	7.30	2.86	37.31	53.23	23.95	8.56	40.61	44.72
	10.45	4.10	4.42	1.73	19.86	23.95	19.55	2.94	27.48	30.74
	203.03	299.43	3.25	1.91	5.50	8.56	2.94	10.57	20.07	21.89
	193.67	293.93	8.10	4.77	31.85	40.61	27.48	20.07	65.66	72.29
Vegetation	185.90	289.35	8.96	5.27	34.59	44.72	30.74	21.89	72.29	80.21
	32.34	12.68	15.33	6.01	234.87	71.67	74.29	-1.26	75.30	70.32
	73.44	28.80	14.32	5.62	71.67	205.13	31.88	-44.05	-41.94	-46.03
	19.65	7.71	7.39	2.90	74.29	31.88	54.58	-14.40	32.78	33.65
	206.66	301.56	6.48	3.81	-1.26	-44.05	-14.40	42.02	56.97	60.02
Thin Cloud Over Ground	189.93	291.72	11.66	6.86	75.30	-41.94	32.78	56.97	136.01	142.66
	181.59	286.82	12.30	7.23	70.32	-46.03	33.65	60.02	142.66	151.24
	77.03	30.21	29.17	11.44	850.83	760.85	45.74	-112.80	-226.19	-280.62
	92.83	36.40	29.20	11.45	760.85	852.63	70.65	-142.72	-237.45	-283.74
	22.33	8.75	7.02	2.75	45.74	70.65	49.32	-72.32	-66.79	-66.54
Thin Cloud	186.08	289.46	11.79	6.94	-112.80	-142.72	-72.32	139.05	184.69	188.85
	141.76	263.39	19.25	11.32	-226.19	-237.45	-66.79	184.69	370.43	392.21
	133.12	258.31	20.55	12.09	-280.62	-283.74	-66.54	188.85	392.21	422.41
	106.09	41.60	17.47	6.85	305.31	298.25	-31.37	76.25	44.64	17.02
	120.89	47.41	19.04	7.47	298.25	362.61	-22.63	57.29	46.02	20.02
Low/Middle Thick Cloud	12.83	5.03	4.47	1.75	-31.37	-22.63	19.98	-49.83	-27.15	-16.59
	159.12	273.60	11.60	6.83	76.25	57.29	-49.83	134.67	104.66	77.17
	90.28	233.10	17.90	10.53	44.64	46.02	-27.15	104.66	320.27	299.39
	84.65	229.79	16.97	9.98	17.02	20.02	-16.59	77.17	299.39	288.07
	167.64	65.74	23.61	9.26	557.64	510.68	52.58	-125.76	-342.30	-341.38
High Thick Cloud	184.71	72.43	22.66	8.89	510.68	513.59	35.98	-66.15	-216.27	-222.39
	8.73	3.42	7.16	2.81	52.58	35.98	51.33	-129.48	-169.54	-162.85
	144.79	265.17	19.32	11.36	-125.76	-66.15	-129.48	373.15	527.73	504.62
	78.00	225.88	34.57	20.34	-342.30	-216.27	-169.54	527.73	1195.37	1148.63
	76.84	225.20	33.28	19.58	-341.38	-222.39	-162.85	504.62	1148.63	1107.53
High Thick Cloud	233.23	91.46	19.12	7.50	365.61	204.94	4.98	-24.07	-2.99	-0.63
	247.45	97.04	11.91	4.67	204.94	141.83	2.12	-10.49	-1.12	0.41
	5.94	2.33	1.38	0.54	4.98	2.12	1.90	-6.89	4.45	4.54
	134.16	258.92	5.37	3.16	-24.07	-10.49	-6.89	28.81	-5.13	-5.77
	42.21	204.83	15.97	9.39	-2.99	-1.12	4.45	-5.13	255.08	255.57
	42.12	204.78	16.07	9.45	-0.63	0.41	4.54	-5.77	255.57	258.17

According to the classified results, water has very small reflectance of 5% in ch1, and 3% in ch2, higher temperature in thermal channels and small standard deviations, which are consistent with its spectral signature. But the reflectance of 4.1% in ch3 seems too high. The reason for this is not clear. For vegetation, it has reflectance of about 13% in ch1 and about 29% in ch2. The reflectance of 29% in ch2 is comparatively low compared to that of vivid vegetation with reflectance of 30-50%. For thin partial cloud over ground, it has mean reflectance of 30% in ch1 and ch2, and temperature of about 260 K. This is because the sensor see both land and clouds, the measured temperature and the reflectance depends on the cloud fraction. The thin clouds have reflectance of about 40% in ch1 and ch2, and have

temperature of about 230 K, which are probably cirrus ice clouds. The mean reflectance for low/middle clouds is understandably as about 70%, but the temperature seems too low. And there are very large standard deviations as large as 20 K in the thermal channels. I think that yellow borders around high thick clouds are still high and thick even thinner and lower than center clouds, and have lower temperature. So when they are classified together with the low thick clouds (with less pixels), the average temperature is still low. The mean vectors and standard deviation for high thick clouds sounds very good. It is also very interesting to note the ch3 and ch3t information. The ch3t temperature is usually much larger than the thermal channel-measured temperature for clouds, the colder the thermal temperature, the larger their difference. The ch3 reflectance is anti-correlated with the ch1 and ch2 reflectance. This is due to that ice clouds can absorb most of the energy in this channel, which has been used to discriminate ice clouds and water clouds [Scorer, 1989; Menzel and Purdom et al., 1994].

4.3. Maximum Likelihood Classification

In order to compare the results of supervised classification with that unsupervised classification, samples for the above sixes classes in Figure 2 are picked from the color composite image, by using the Figure 2 as a reference. About 500-4000 training samples (20-50 boxes) are picked for each class. The generated parameters for the maximum likelihood classifier are listed in Table 2. The results of unsupervised classification in Table 1 are also used to train the supervised classifier. Compare Table 1 with Table2, the mean vectors for these two cases are within 5 GL. Figure 3a and 3b shows the corresponding thematic maps. The statistics for each class after maximum likelihood classification are shown in Table 3 and the tabular summary of the thematic maps are shown in Table 4.

Compare Figure 3 with Figure 2, we can see that their structure are very similar. It's not surprising to see this similarity, since the Figure 3b use the results of Figure 2 as training samples, and picking samples for producing Figure 3a is referred to Figure 2. Another reason is that there are no elongated classes present in this image, as we can see from Table 1 and Table 2. The difference mainly lies in that supervised classification adopts threshold limitation so that not all the pixels are labeled to those known classes. Those unknown classes include those cloud shades (classified as water in unsupervised classification), vegetation, and different types of clouds in Figure 1. These pixels classified as unknown class are probably not pure, and contaminated by other classes. The results using unsupervised results as training samples sounds better than that using picked training samples in this case, since the former produces 13.8%(This is the best that I have got, usually 20-40%) unknown pixels, and the latter only 6.8%. The mean vectors for these three classifications are close, but the unsupervised classification usually has higher standard deviation than the picked training samples. The larger standard deviation (related to multivariate spreads and covariance matrix, but it's difficult to see something from the listed digits in covariance matrix) might contribute to less number of unknown pixels. The tabular summaries and area estimates of each class listed in Table 4 for these three classifications are also very close, if considering the presence of unknown classes. The number of pixels for vegetation and thin partial

clouds obtained by picked training samples is 10,000-15,000 less than the other two, which are due to that the picked training samples have smaller standard deviation in all the channels, and those pixels contaminated by clouds are classified as unknown class. The number of thin clouds in Figure 3a is about 5,000 larger than the other two, even though the picked training samples have much smaller standard deviation. It's probably because less pixels are classified as thin partial clouds using picked samples. If no threshold limitation is used in supervised classification, the results trained by MMC are almost the same as that of unsupervised classification, that trained by picked training samples are not as good as the unsupervised classification if compared to the original image.

Table 2. Class signatures for the trained samples

Classes	Mean Vector		Standard Deviation				Covariance Matrix				
	Gray Level	R% or T (K)	Gray Level	R% or T (K)							
Water	14.33	5.62	6.70	2.63	44.89	21.49	9.15	3.46	15.42	16.89	
	8.40	3.29	7.29	2.86	21.49	53.18	17.16	6.51	28.08	32.30	
	9.47	3.71	3.63	1.43	9.15	17.16	13.21	1.48	17.88	20.93	
	202.84	299.32	3.12	1.83	3.46	6.51	1.48	9.71	14.94	15.65	
	194.63	294.49	6.55	3.85	15.42	28.08	17.88	14.94	42.86	47.43	
Vegetation	186.89	289.93	7.29	4.29	16.89	32.30	20.93	15.65	47.43	53.20	
	26.41	10.36	11.56	4.53	133.58	58.00	54.19	8.46	75.41	79.13	
	73.16	28.69	11.29	4.43	58.00	127.36	12.02	0.39	19.82	19.05	
	16.50	6.47	5.94	2.33	54.19	12.02	35.33	-3.73	30.60	33.88	
	208.41	302.59	3.45	2.03	8.46	0.39	-3.73	11.87	16.71	16.61	
Thin Cloud Over Ground	195.74	295.14	7.93	4.67	75.41	19.82	30.60	16.71	62.91	66.06	
	187.67	290.39	8.37	4.92	79.13	19.05	33.88	16.61	66.06	70.07	
	70.12	27.50	19.37	7.60	693.99	620.59	37.31	-86.24	-172.92	-214.54	
	83.14	32.60	20.29	7.96	620.59	695.46	57.62	-109.11	-181.54	-216.92	
	22.95	9.00	4.47	1.75	37.31	57.62	40.23	-55.29	-51.06	-50.87	
Thin Cloud	188.14	290.67	11.32	6.66	-92.01	-116.41	-58.99	106.30	141.20	144.38	
	135.46	259.69	24.21	14.24	-184.49	-193.68	-54.48	141.20	283.20	299.85	
	125.92	254.07	22.27	13.10	-228.89	-231.43	-54.27	144.38	299.85	322.94	
	101.37	39.75	9.39	3.68	357.17	348.91	-36.69	62.33	36.49	13.91	
	118.25	46.37	10.48	4.11	348.91	424.21	-26.47	46.83	37.62	16.37	
Low/Middle Thick Cloud	13.76	5.39	4.94	1.94	-36.69	-26.47	23.37	-40.73	-22.19	-13.56	
	160.02	274.13	11.28	6.63	89.21	67.02	-58.29	110.08	85.55	63.08	
	88.30	231.94	9.31	5.48	52.22	53.84	-31.76	85.55	261.79	244.72	
	81.74	228.08	9.75	5.74	19.91	23.42	-19.41	63.08	244.72	235.46	
	176.54	69.23	37.63	14.76	455.96	417.57	43.00	-157.01	-427.35	-426.19	
High Thick Cloud	189.09	74.15	40.83	16.01	417.57	419.94	29.42	-82.58	-270.00	-277.65	
	11.91	4.67	11.08	4.34	43.00	29.42	41.97	-161.65	-211.66	-203.31	
	151.56	269.16	30.09	17.70	-102.83	-54.09	-105.87	465.86	658.84	629.99	
	85.90	230.53	43.86	25.80	-279.89	-176.83	-138.63	658.84	1492.35	1434.00	
	83.99	229.41	42.42	24.95	-279.14	-181.84	-133.16	629.99	1434.00	1382.68	
High Thick Cloud	233.55	91.59	17.01	6.67	484.38	271.52	6.60	-21.42	-2.66	-0.56	
	246.93	96.84	10.72	4.20	271.52	187.91	2.81	-9.34	-1.00	0.36	
	6.03	2.37	1.61	0.63	6.60	2.81	2.52	-6.13	3.96	4.04	
	133.77	258.69	6.42	3.77	-31.89	-13.90	-9.13	25.64	-4.57	-5.14	
	41.38	204.34	1.76	1.03	-3.96	-1.48	5.90	-4.57	227.06	227.50	
	40.75	203.97	2.29	1.35	-0.83	0.54	6.01	-5.14	227.50	229.81	

Table 3. Class signatures for Supervised Classification

Classes	Supervised Classification Using Picked Training Samples				Supervised Classification Using Results of Unsupervised Classification			
	Mean Vector		Standard Deviation		Mean Vector		Standard Deviation	
	Gray Level	R% or T (K)	Gray Level	R% or T (K)	Gray Level	R% or T (K)	Gray Level	R% or T (K)
Water	11.26	4.42	3.10	1.22	11.56	4.53	3.44	1.35
	6.11	2.40	3.06	1.20	6.50	2.55	3.65	1.43
	9.20	3.61	2.65	1.04	9.61	3.77	3.26	1.28
	203.48	299.69	1.33	0.79	203.46	299.68	1.58	0.93
	195.81	295.19	3.43	2.02	195.34	294.91	4.54	2.67
	188.24	290.73	4.06	2.39	187.73	290.43	5.24	3.08
Vegetation	25.91	10.16	8.58	3.36	29.52	11.57	12.13	4.76
	73.67	28.89	10.68	4.19	73.46	28.81	12.57	4.93
	16.98	6.66	4.63	1.82	18.58	7.28	5.72	2.24
	208.24	302.49	3.44	2.02	206.88	301.69	5.31	3.13
	195.41	294.95	6.50	3.83	191.45	292.62	10.40	6.12
	187.19	290.11	6.84	4.02	183.07	287.69	11.02	6.48
Thin Cloud Over Ground	68.13	26.72	24.46	9.59	72.83	28.56	28.36	11.12
	84.49	33.13	26.29	10.31	88.79	34.82	29.67	11.63
	22.63	8.87	4.18	1.64	22.57	8.85	5.23	2.05
	187.60	290.35	8.82	5.19	187.09	290.05	10.02	5.90
	143.72	264.54	17.12	10.07	143.01	264.12	18.16	10.68
	134.69	259.23	17.94	10.55	134.03	258.84	19.30	11.35
Thin Cloud	103.67	40.66	20.42	8.01	105.29	41.29	18.33	7.19
	119.44	46.84	22.81	8.95	120.37	47.20	20.64	8.09
	13.44	5.27	4.70	1.84	12.83	5.03	4.15	1.63
	160.68	274.52	11.74	6.91	159.32	273.72	10.71	6.30
	92.10	234.18	18.46	10.86	90.24	233.08	17.19	10.11
	85.88	230.52	16.84	9.91	84.45	229.68	16.09	9.46
Low/Middle Thick Cloud	166.35	65.23	20.35	7.98	166.49	65.29	20.01	7.85
	181.85	71.31	21.11	8.28	183.46	71.95	20.24	7.94
	8.14	3.19	5.34	2.09	7.48	2.93	4.39	1.72
	143.79	264.58	18.36	10.80	141.53	263.25	15.21	8.95
	76.85	225.21	35.47	20.87	72.58	222.69	31.10	18.30
	75.60	224.47	33.81	19.89	71.52	222.07	29.59	17.41
High Thick Cloud	232.76	91.28	19.42	7.62	233.73	91.66	18.49	7.25
	247.15	96.92	12.14	4.76	247.97	97.24	10.98	4.31
	5.91	2.32	1.14	0.45	5.92	2.32	1.14	0.45
	134.03	258.84	4.71	2.77	134.07	258.87	4.77	2.80
	41.71	204.53	14.76	8.68	41.55	204.44	14.59	8.58
	41.50	204.41	14.83	8.73	41.40	204.35	14.69	8.64

Table 4. Tabular summary of the thematic maps of Figure 2, 3a and 3b.

Classes	MMC (Figure 2)		MLC(Figure 3a)		MLC(Figure 3b)	
	No. of Pixels	Area(km ²)	No. of Pixels	Area(km ²)	No. of Pixels	Area(km ²)
Water	57051	69031.7	48890	59156.9	50669	61309.5
Vegetation	51631	62473.5	35275	42682.8	46662	56461.0
Thin Cloud over ground	43682	52855.2	33288	40278.5	41421	50119.4
Thin cloud 2	36578	44259.4	41529	50250.1	37992	45970.3
Low thick cloud	26558	32135.2	22393	27095.5	23598	28553.6
High thick cloud	46644	56439.2	44709	54097.9	44406	53731.3
Unknown			36060	43632.6	17396	21049.2

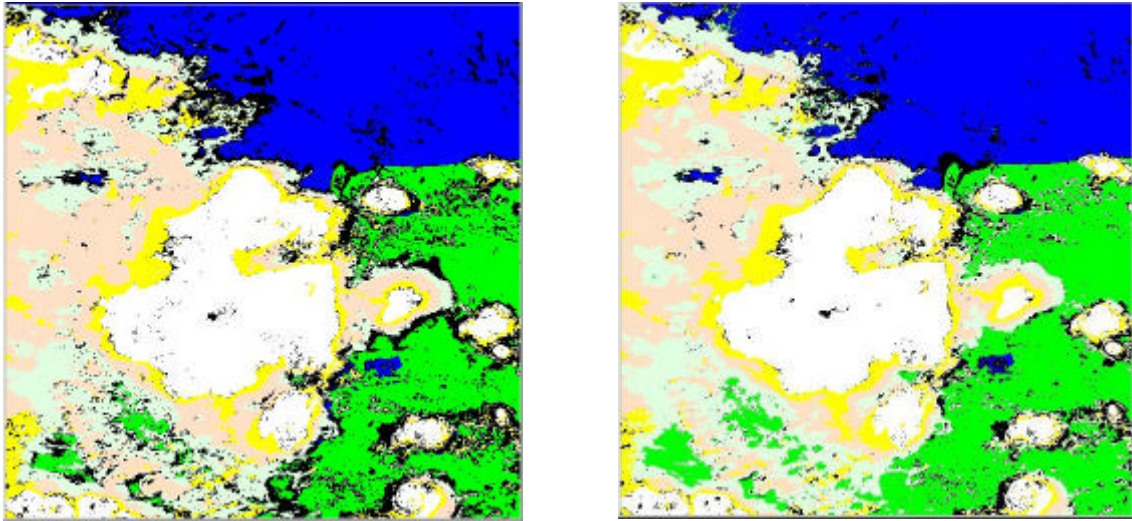


Figure 3. Thematic maps of Maximum Likelihood classification (95% pixels are classified) trained by picked training samples (left) and by unsupervised classification (Right). Blue represent water and cloud shade, green is vegetation, light green is thin cloud over ground, pink is thin cloud, yellow is low and middle cloud, white is high clouds, and black represents unknown classes.

4.4. Comparison of MLC supervised classification and MMC unsupervised classification.

Even though I didn't get significantly different classified image by using supervised classification and unsupervised classification due to the easily identified features no elongated classes present in the selected image, we still can tell some disadvantages and advantages of these two methods.

In the view of computation time, the MLC supervised classification is even faster than the MMC unsupervised classification. In this project, it takes 580s to run the unsupervised classification (20 iterations), with 28s in each iteration. It takes about 80s to perform the supervised classification. If only iterate five times, MMC will take 1.75 time as that of MMC (Actually, 5 iterations is also enough to produce good results in this case).

Even though the MLC supervised classification usually works more effectively since it takes advantage of the information of multivariate spreads of each classes, but the MMC classification works more effectively in this case. The key and the most difficult thing for MLC is to pick high qualitative training samples, which makes it more difficulty for us none-experts. The trained samples are required not only to represent the mean vectors but also the spreads for each class. Otherwise, one might get too many unknown classes. In this project, I did supervised classification using picked samples as training samples for many times, if set that 95% of pixels will be classified, then I usually get 20-40% unknown classes. And the repeatability is very bad, doing it more cannot guarantee that you obtain better results, it can

even be worse. While using MMC classification, we don't need to have foreknowledge of each class, just need to set the not-so-important initial cluster centers. The repeatability is very good even the initial cluster centers are different. I also try to use some of the previous parameters of trained data set to serve as initial cluster centers, it makes little difference to the final products. (But the initial cluster centers do affect the results and the computation time. For example, if one select the initial cluster centers uniformly distributed along the diagonal line with the infrared gray level not flipped, we can not get good results, and it will take long computation time.). Even though MMC takes more computation time, but save us a lot to picking training samples again and again.

The idea of using the results of MMC as training samples, called hybrid supervised/unsupervised classification [Richards, 1993, p270] combines the advantages of both supervised classification and unsupervised classification. We don't have foreknowledge the classes to pick out the unimodal clusters. Since the results of unsupervised classification are usually not so bad compared to the supervised classification even for those experts. Using all the pixels as training samples is expected to obtain probably more accurate parameters including the mean vectors and the covariance matrix, and I expect it to obtain good parameters for those elongated classes as well. As we can see from Figure 3b, this hybrid classification can separate cloud shades from water, can label those vegetation contaminated by clouds as unknown class. So this hybrid classification can consider the multivariate spreads of each spectral classes without the user having foreknowledge about the classes. That's really cool. If we are required to do global classification, picking representative training samples is very difficult due to too many classes and temporal and spatial variability, the hybrid supervised/unsupervised classification is a very good method. However, this method will take more computation time. So if we need to classify a very large image, we might only use a subset of the image to perform unsupervised classification to get the trained supervised classifier. In addition, the obtained statistics (including mean and covariance matrix) are weighted by the number of available pixels within a certain range gray levels of a class. For example, If in a subset image, if a class consists of 95% of green vegetation and 5% yellow vegetation, then the generated parameters will mainly like that of green vegetation, so the yellow vegetation in a large image might not be well classified as vegetation.

5. Conclusion

From the results and discussions above, the results of maximum likelihood classification, migrating means clustering classification the hybrid classification are satisfactory, and are close and consistent with the original three bands overlay image. Even the generated thematic maps are not beautiful as that of three bands overlay image due to the solid color effects and natural color scheme in the three-band overlay image. The advantage of classification is obvious, we can get the physical meaningful reflectance or temperature and their multivariate spreads, we can know the estimate the area coverage for each class, which is important for quantitative analysis.

The statistical mean vectors for each class are quite reasonable to my knowledge. What's new to me is the ch3 and ch3t information. The ch3t temperature is usually much larger than the thermal channel-measured temperature for clouds, the colder the thermal temperature, the larger their difference. The ch3 reflectance is anti-correlated with the ch1 and ch2 reflectance, which is due to that high reflectance ice clouds can absorb most of the energy in this channel. Even the difference among them are small, the MMC and the MLC trained by MMC works better than the MLC trained by selected training samples. That's because it is difficult especially for us non-experts to get representative training samples. And in this project, the standard deviation (spreads) for each class generated by picked samples are small than that of unsupervised results, so we get more pixels labeled as unknown class using the MLC trained by picked samples. MMC usually takes more computation time, if the number classes are consistent and MMC run five iterations. But if consider the consumed time needed to pick training samples over and over to get better results, the MCC and MLC trained by unsupervised results are less time-consuming compared to MLC trained by picked samples.

The hybrid supervised/unsupervised classification combines the advantages of both supervised classification and unsupervised classification. It doesn't require the user have the foreknowledge about the classes about each classed looks like in multispectral space, but still can obtain accurate mean vectors and covariance matrix, and consider the multivariate spreads of each spectral classes by using all the pixels image as training samples.

6. Reference

- Menzel, W. P., and J. F. W. Purdom, Introducing GOES-I: The first of a new generation of geostationary operational environmental satellites. *Bulletin of American Meteorological Society*, 75(5), 757-781, 1994.
- Richards, J. A., *Remote sensing digital image analysis: an introduction* (second edition), 1993.
- Scorer, R. S., Cloud reflectance variations in channel-3. *Int. J. Remote Sens.*, 10,675-686, 1989.