# Automated cloud cover assessment for Landsat TM images

Ben Hollingsworth

Liqiang Chen

Stephen E. Reichenbach

Computer Science and Engineering Department
University of Nebraska—Lincoln, Lincoln, NE 68588-0115
(402)472-5007 (voice), (402)472-7767 (fax)
obiwan@jedi.com

Richard Irish

SSAI - Goddard Space Flight Center, Greenbelt, MD 20771
(301)286-0951 (voice), (301)286-1774 (fax)
rich_irish@ccmail.gsfc.nasa.gov

**Abstract**

This paper describes a method for cloud cover assessment using computer-based analysis of multi-band Landsat images. The objective is to accurately determine the percentage of cloud cover in an efficient manner. The "correct" value is determined by an expert's visual assessment. Acceptable error rates are ±10% from the visually-determined coverage.

This research improves upon an existing algorithm developed for use by the EROS Data Center several years ago. The existing algorithm uses threshold values in bands 3, 5, and 6 (red, middle infrared, and thermal, respectively) based on the expected frequency response for clouds in each band. While this algorithm is reasonably fast, the accuracy is often unsatisfactory.

The dataset used in developing the new method contained 329 subsampled, 7-band Landsat browse images with wide geographic coverage and a variety of cloud types. This dataset, provided by the EROS Data Center, also specifies the visual cloud cover assessment and the cloud cover assessment using the current automated algorithm. Mask images, separating cloud and non-cloud pixels, were developed for a subset of these images.

The new approach is statistically based, developed from a multi-dimensional histogram analysis of a training subset. Images from a disjoint test set were then classified. Initial results are significantly more accurate than the existing automated algorithm.

**Key words:** remote sensing, image analysis, automated cloud cover assessment, multi-spectral analysis, environmental sensing.

# 1 INTRODUCTION

Since 1972, NASA has been acquiring image data from its Landsat satellites. In order to determine the usefulness of the images, each one must be assessed to determine its percentage of cloud cover. Since cloud cover affects the accuracy of global weather analysis, and since an image with a high percentage of cloud cover is often useless to organizations which purchase these images, an accurate assessment is crucial. Since the current acquisition rate is over 3600 images per day (248 images per orbit $\times$ 233 orbits to cover the entire earth / 16 days to cover the entire earth), numerous attempts have been made to automate this process by training computers to accurately and consistently determine the cloud coverage of a scene.

The automated algorithm currently used, developed in 1987, uses a threshold-based approach, which has some inherent flaws. This research proposes a statistical pattern matching approach known as supervised classification. Supervised classification uses a training set with known correct answers to determine typical features that describe the desired pattern (in this case, clouds), and then uses this information to classify a disjoint test set for which correct answers are not known a priori.

# 2 THE DATASET

The dataset used in developing the new method contained 329 subsampled, 7-band Landsat browse images provided by the EROS Data Center (EDC). The subsampled browse images were approximately 380 pixels square with a pixel spacing of approximately 480 meters. The images were acquired primarily over Eastern Europe, the Middle East, Africa, South America, and North America, and spanned latitudes from 72 degrees north (northern Alaska and Russia) to 32 degrees south (South Africa). Virtually every major surface and cloud type was represented. All images in the dataset were daytime scenes. This dataset also specifies the visual cloud cover assessment and the cloud cover assessment using the current automated algorithm. Mask images, designating cloud and non-cloud pixels, were developed for 210 of the 329 images.

The image pixels measure the radiance of the scene, and are scaled into the 0–255 range. It was initially thought that reflectance, which can be computed given the radiance value and the sun elevation angle at each point, might be a better measurement to work with than radiance. However, it was determined that the errors between various algorithmic assessments and the “correct” visual

| Value | | Percentage | Value | | Percentage |
|---|---|---|---|---|---|
| 0 | = | 0–4% | 5 | = | 45–54% |
| 1 | = | 5–14% | 6 | = | 55–64% |
| 2 | = | 15–24% | 7 | = | 65–74% |
| 3 | = | 25–34% | 8 | = | 75–84% |
| 4 | = | 35–44% | 9 | = | 85–100% |

Table 1: Mapping of cloud cover assessment values to coverage percentages

assessment were uncorrelated with the sun elevation angle, indicating that this was not the case. Just to be sure, the images were converted from radiance to reflectance prior to processing and the cloud cover assessment results were compared to those using unconverted radiance values. The results using reflectance were worse than those using radiance, so the original radiance values were used in all subsequent research.

The cloud cover assessment values used by EDC are single-digit values ranging from 0–9. Table 1 gives the corresponding coverage percentages. The visual assessments, which were used as a baseline for judging errors in the automated algorithms, were only available as the single-digit values, so all automated results were also converted from percentages into these values before the error calculations. All results and statistics given in this paper are in terms of these values, not percentages.

# 3 THE CURRENT ACCA ALGORITHM

This research improves upon an existing algorithm developed for use by EDC several years ago. The existing Automated Cloud Cover Assessment (ACCA) algorithm uses threshold values in bands 3, 5, and 6 (red, middle infrared, and thermal, respectively) based on the known frequency response for clouds in each band. While this algorithm is reasonably fast, the accuracy is often unsatisfactory.

The current ACCA algorithm processes each pixel sequentially as follows:

1. Each pixel is first thresholded in TM bands three and six to determine whether it is bright enough and cold enough to be either a cloud or snow/ice.

2. If the pixel passes the first thresholds, it is next thresholded in band five to determine if it is bright enough to be cloud, since snow/ice has a lower reflectance than most clouds at the 1700nm wavelength.

3. If it fails that test, it is again thresholded in bands five and three to distinguish between snow/ice and ice cloud.

Figure 1 demonstrates two of the problems with the current threshold-based ACCA algorithm. The coverage assessment for each image is given in the cap-

tion. The images on the left are of a scene taken over the Amazon basin. Most of the clouds here are too warm to pass the thermal (band 6) threshold, and are therefore not marked as clouds. The images on the right are of a scene near Mt. McKinley in Alaska. While there are no clouds in this image, much of the snow on the higher peaks has a similar frequency response, and is therefore classified as clouds. Fortunately, in this case, few enough pixels were miss-classified that the overall coverage assessment was not adversely affected.

For the images in the available dataset, the current ACCA algorithm gave results which were, on average, 1.184 units less than the visual assessment. However, the difference between the two assessments had a root-mean-squared (RMS) error of 1.875. At the extremes, the ACCA assessment was as much as 9 units below the visual assessment or 3 units above. This indicates a large number of images with errors well outside the acceptable RMS error range of $\pm 1.0$.

The distribution of assessment values for ACCA and the visual assessment are graphed in Figure 2. Of the graphed images, 4% had a higher automated than visual assessment, 40% had the same automated and visual assessment, and 56% had a lower automated than visual assessment. Of that 56%, the mean error was -1.813. On a Sun SPARCserver 670MP, this algorithm ran in about 0.1 seconds per image.

# 4 THE NEW STATISTICAL APPROACH

## 4.1 The basic algorithm

The algorithm proposed in this paper uses a supervised classification technique. First, a training set of 61 images was selected from the 329 available images. The training set was built by visually inspecting the images and making sure that all major surface and cloud types were well-represented and that the mask's assessment was similar to the visual assessment.

An $n$-dimensional histogram was then computed from the training set as follows: For each pixel, if the mask image indicated that the pixel was a cloud, the histogram value for that pixel's $n$-dimensional value was incremented. If the mask image indicated that the pixel was not a cloud, the appropriate histogram value was decremented. In the resulting histogram, regions of $n$-tuples with positive values indicated $n$-tuples that usually represented clouds, while ranges of $n$-tuples with non-positive values indicated $n$-tuples that usually did not correspond to clouds.

Using this histogram as the definition of what a cloud "looks like," the remaining 268 images (the "test set") were then classified. The classification algorithm simply takes each $n$-dimensional pixel value from an image and does a lookup into the histogram. If the lookup yields a positive value, the pixel is marked as a cloud; otherwise it is left unmarked. Histogram lookups yielding a value of zero were arbitrarily chosen to indicate a non-cloud pixel.

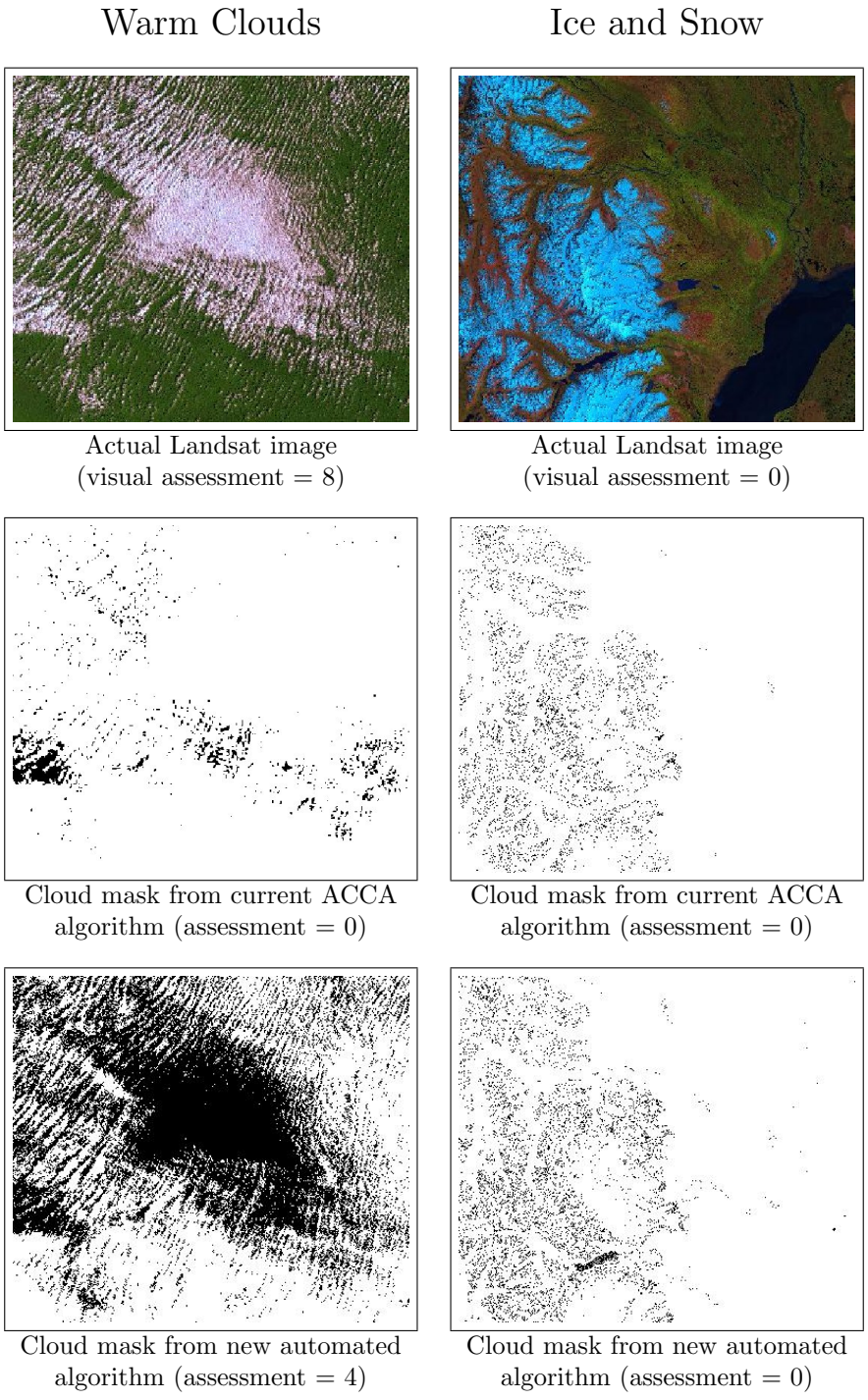



Figure 1: Problems with current ACCA algorithm

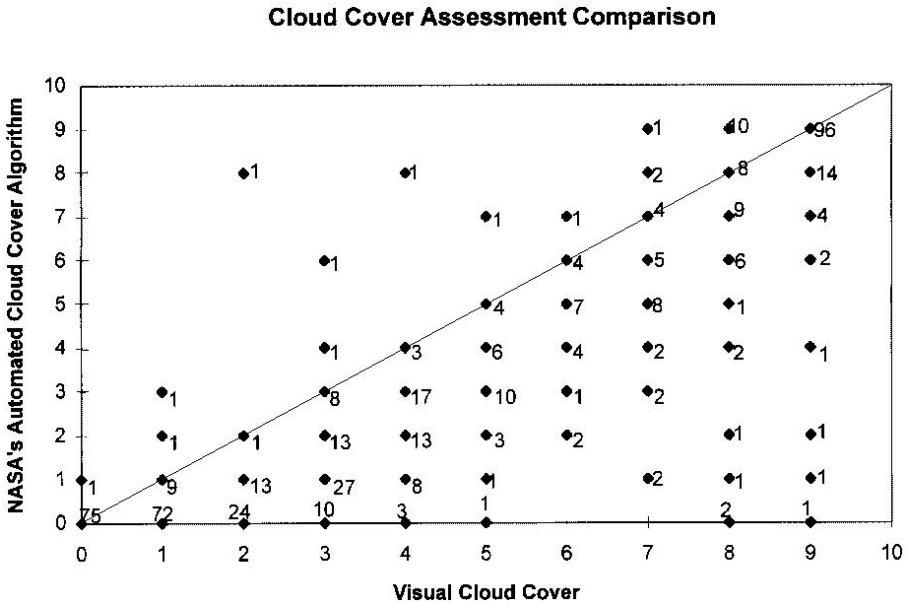



Figure 2: ACCA vs. visual assessment

## 4.2 Internal data storage

While the basic algorithm remained unchanged, numerous implementation details were modified in order to achieve good results. The most important of these related to the size of the histogram. Since each pixel value has a range of 0–255, a 7-dimensional histogram could potentially contain $7.2 \times 10^{16}$ entries. If each entry was implemented as a 16-bit integer, a full histogram would require 144,000 terrabytes of storage space. Clearly, this is impractical.

The first approach used to lessen the storage requirement was to aggregate the 256 possible pixel values for each dimension (band) into larger groups. Each time the number of possible pixel values is cut in half, the storage requirement is divided by $2^n$ ($n$ is the number of bands), or 128 for a 7-band histogram. Still, in order to bring the histogram down to a manageable size (approximately 4MB), each dimension must be aggregated into 8 buckets, each encompassing 32 values in the original 0–255 range.

Quantizing the pixel values this way also has the side effect of increasing the likelihood of a histogram lookup "hit" (finding a non-zero value) when running the algorithm on the test set. This is important when you consider the number of possible $n$-tuples that can fall within $\pm 1$ value from a given $n$-tuple, and hence the probability of a "miss" if the training set did not contain enough unique pixel $n$-tuples to fully populate the cloudy regions of the histogram. As implemented, using a "quantization factor" of $q$ divides the range for each image pixel value by $2^q$, creating $2^{8-q}$ buckets of $2^q$ values each. When using an $n$-

| Band | RMS when missing |
|---|---|
| 6 (thermal) | 1.802 |
| 1 (blue) | 1.600 |
| 5 (middle-IR) | 1.536 |
| 3 (red) | 1.514 |
| 4 (near-IR) | 1.485 |
| 2 (green) | 1.462 |
| 7 (middle-IR) | 1.433 |

Table 2: Resulting RMS error when bands are omitted

bands histogram, a quantization factor of $q$ will reduce the size of the histogram by a factor of $2^{nq}$.
Statistics are given in this paper for several variations of the new supervised classification method. All variations use a quantization factor of 5, which partitions the 256 possible pixel values into 8 blocks of 32 values each. This not only keeps the storage requirements manageable, but also gives slightly better results than do quantization factors of 4 or 6.
Another optimization that reduces the size of the histogram is based on the fact that a histogram resulting from the training stage is typically extremely sparse. Even when the pixel values are quantized by a factor of 5, the histogram for a 50-image training set is still only about 1% dense. The storage requirement of the histogram can thus be further reduced by storing the histogram as a linked list rather than an array. This reduces the size of most histograms by a factor of 13, and also gives the ability to change the dimensionality (number of included bands) in the histogram with minimal code changes.

### 4.3 Omitting bands

In order to both reduce the amount of storage space required to store the histogram and reduce the processing time of the algorithm, one or more bands may be ignored. To determine which bands had the least impact on the final results, the new algorithm was run seven times, omitting a different band each time. The RMS differences between the new results and the visual assessment are listed in Table 2. It can be seen from the table that bands 7, 2, and 4, in that order, are the best candidates for omission.
By testing various combinations of bands, it was found that omitting both bands 7 and 2 from consideration produced the best results. The initial supervised approach, using all 7 Landsat bands as input, differed from the visual assessment by a mean of -1.029, an RMS of 1.613, and a range of -5 to +2. This 7-band configuration ran in about 1.9 seconds per image. When both bands 2 and 7 were omitted, the mean became -0.943, the RMS dropped to 1.547, and the range increased slightly to -5 to +3. This smaller, 5-band configuration took about 1.3 seconds per image. Since this trend was consistent throughout most

of the test variations, the majority of the tests include only bands 1, 3, 4, 5, and 6, ignoring bands 2 and 7.

## 4.4 Windowing

Although the results obtained with the new algorithm as described thus far were good, they were still often closer to the results of the original ACCA algorithm than to the visual assessment. Visual analysis of the images indicated that most of the images that produced high errors for both the ACCA algorithm and this new approach contained a large number of "popcorn" clouds, or small puffs of clouds scattered throughout an area. The mean error of the ACCA algorithm for images containing primarily popcorn clouds was -2.41, compared with -0.93 for all other cloud types. An example of popcorn clouds is given in the upper-right image in Figure 3.

It appears as though the visual assessment is classifying the entire popcorn-covered area as "cloudy," when in reality, only 50–60% of the individual pixels may represent clouds. This is probably a reasonable action on the part of the visual assessor, as most people purchasing satellite images require large sections of cloud-free pixels, not tiny strips intertwined with small clouds. However, this means that any purely pixel-based algorithm will never match the visual assessment, even if it can recognize individual cloud pixels 100% of the time. The solution to this is to take into account larger spatial areas, or "windows," when determining cloud cover.

The windowing technique uses an $N \times N$ window centered around the pixel whose value is being computed. The cloud/non-cloud status of each pixel in the window is determined individually. The cloud cover percentage for the entire window is then compared to a threshold value. If the percentage of cloud cover in the window is greater than the threshold, then the final status of the pixel in the center of the window is set to "cloud." If the window's coverage is below the threshold, the center pixel is set to "non-cloud." This windowing algorithm has the effect of converting the staccato, single-pixel resolution of the resulting cloud masks into more blob-shaped areas.

This windowing modification can be applied to both the current ACCA algorithm and the new statistical pattern matching approach, and obtains significant improvements for each.

Both algorithms were tested using window sizes ranging from $1 \times 1$ to $25 \times 25$. Window coverage thresholds ranged from 1% to 50%. The RMS errors for all tested combinations of window size and threshold are listed in Table 3. In each table, the combination with the lowest RMS error is surrounded with a box.

It can be seen in Table 3 that the best results for both algorithms occurred when a $5 \times 5$ window was used. Since the goal of windowing is to smooth out geographical features in the image, it is likely that the appropriate window size is a function of the surface area (about 2.4 km square for a $5 \times 5$ window) rather than the pixel area.

The current ACCA algorithm achieved the best results with a threshold of just 5%. This threshold seems quite low, and may only be required to compensate for

Current ACCA Algorithm:

| Window Size | Threshold | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1% | 2% | 5% | 10% | 20% | 30% | 40% | 50% |
| $1 \times 1$ | 1.875 | 1.875 | 1.875 | 1.875 | 1.875 | 1.875 | 1.875 | 1.875 |
| $3 \times 3$ | 1.517 | 1.517 | 1.517 | 1.517 | 1.619 | 1.708 | 1.844 | 1.921 |
| $5 \times 5$ | 1.558 | 1.558 | 1.487 | 1.551 | 1.604 | 1.700 | 1.836 | 1.955 |
| $7 \times 7$ | 1.611 | 1.611 | 1.544 | 1.551 | 1.602 | 1.698 | 1.878 | 1.983 |
| $11 \times 11$ | 1.719 | 1.672 | 1.602 | 1.571 | 1.595 | 1.713 | 1.882 | 2.052 |
| $17 \times 17$ | 1.860 | 1.782 | 1.658 | 1.577 | 1.579 | 1.709 | 1.893 | 2.068 |
| $25 \times 25$ | 1.978 | 1.874 | 1.748 | 1.604 | 1.606 | 1.710 | 1.906 | 2.108 |

New Statistical Approach:

| Window Size | Threshold | | | | | |
|---|---|---|---|---|---|---|
| | 10% | 15% | 20% | 30% | 40% | 50% |
| $1 \times 1$ | 1.547 | 1.547 | 1.547 | 1.547 | 1.547 | 1.547 |
| $3 \times 3$ | 1.148 | 1.108 | 1.121 | 1.268 | 1.475 | 1.641 |
| $5 \times 5$ | 1.141 | 1.119 | 1.089 | 1.298 | 1.489 | 1.735 |
| $7 \times 7$ | 1.271 | 1.137 | 1.112 | 1.306 | 1.547 | 1.764 |
| $11 \times 11$ | 1.296 | 1.174 | 1.137 | 1.321 | 1.598 | 1.779 |
| $17 \times 17$ | 1.370 | 1.200 | 1.158 | 1.338 | 1.602 | 1.839 |
| $25 \times 25$ | 1.430 | 1.222 | 1.189 | 1.355 | 1.613 | 1.863 |

Table 3: RMS errors for ACCA and statistical windowing algorithms

the (sometimes excessive) underestimation of the correct number of individual cloud pixels caused by the threshold-based approach. For example, see "Warm Clouds" in Figure 1. The statistical approach obtained the best results with a threshold of 20%. (Its RMS error of 1.089 is the lowest error produced by this research thus far.) It can be seen that the farther the non-windowed results are below the correct values, the lower the threshold must be in order to correct the results.

The results of these windowing modifications to both algorithms is demonstrated in Figure 3. The actual satellite image is shown in the upper right. This image, taken in June over western Russia, demonstrates a scene almost completely obscured by popcorn clouds. The visual assessment of this image was 8, indicating that approximately 80% of the image was cloudy. The manually-created mask image in the upper left indicates that only 30% of the individual pixels were clouds, yielding an assessment value of 3. The lower four images are the computed cloud masks for both ACCA and the statistical algorithms, with and without windowing. The assessment value for each image is given. It can be seen from the resulting cloud masks that windowing fills in the more dense areas of clouds and clears out the more sparse areas, giving results which are much closer to the visual assessment.

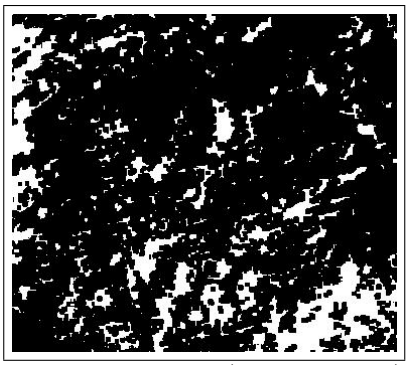
Unfortunately, the windowing modification does have its drawbacks. The sta-

tistical approach required about 23 seconds per image to run with a $25 \times 25$ window, compared to 1.3 seconds without windowing. The $5 \times 5$ window which yielded the best results took about 2.7 seconds per image. The current ACCA algorithm with the windowing modifications took about 2.0 seconds per image with a $25 \times 25$ window and 0.3 seconds for a $5 \times 5$ window, compared to 0.1 seconds without windowing.

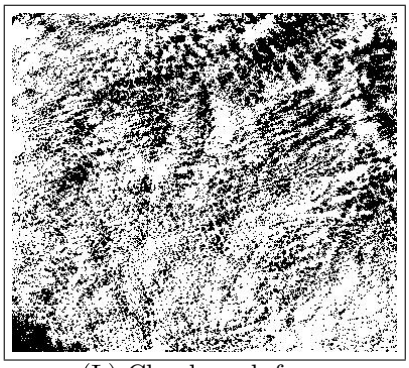
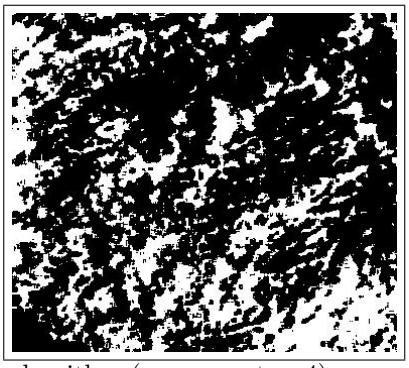
# 5 CONCLUDING REMARKS

If we assume, for the sake of statistical comparison between the current ACCA algorithm and this research, that the experts' visual assessment of each image is correct, then the cloud cover assessment results from the new algorithm are significantly better than those of the ACCA algorithm, achieving an RMS error 40% smaller than ACCA (1.089 vs. 1.875). Still, there is room for improvement. Future work may include reworking the quantization approach to increase the histogram resolution in more dense areas of the histogram. Another approach that may be attempted is the use of unsupervised classification via the fuzzy c-means algorithm or neural networks.


(L) Manually-created cloud mask provided by EDC (assessment = 3)
(R) Actual Landsat image (visual assessment = 8)

(L) Cloud mask from the current ACCA automated algorithm (assessment=4)
(R) With $5 \times 5$ window and 5% threshold (assessment = 8)

(L) Cloud mask from new automated algorithm (assessment = 4)
(R) With $5 \times 5$ window and 20% threshold (assessment = 7)

Figure 3: The advantages of windowing

# References


[1] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, Inc., 1973.

[2] Elizabeth E. Ebert. Analysis of polar clouds from satellite imagery using pattern recognition and a statistical cloud analysis scheme. *Journal of Applied Meteorology*, 28:382–399, May 1989.

[3] Timothy C. Gallaudet and James J. Simpson. Automated cloud screening of avhrr imagery using split-and-merge clustering. *Remote Sensing Environment*, 38:77–121, 1991.

[4] J. Key and R. G. Barry. Cloud cover analysis with arctic avhrr data. *Journal of Geophysical Research*, 94(D15):18,521–18,535, December 1989.

[5] J.R. Key, J.A. Maslanik, and R.G. Barry. Cloud classification from satellite data using a fuzzy sets algorithm: A polar example. *International Journal of Remote Sensing*, 10(12):1823–1842, 1989.

[6] Morton Nadler and Eric P. Smith. *Pattern Recognition Engineering*. John Wiley and Sons, Inc., 1993.

[7] D. Pairman and J. Kittler. Clustering algorithms for use with images of clouds. *International Journal of Remote Sensing*, 7(7):855–866, 1986.

[8] G. S. Pankiewicz. Pattern recognition techniques for the identification of cloud and cloud systems. *Meteorological Applications*, 33:257–271, 1995.

[9] Robert J. Schalkoff. *Pattern Recognition: Statistical, Structural and Neural Approaches*. John Wiley and Sons, Inc., 1992.

[10] James J. Simpson and Jason I. Gobat. Improved cloud detection in geos scenes over land. *Remote Sensing of Environment*, 52:36–54, 1995.

[11] James J. Simpson and Christopher Humphrey. An automated cloud screening algorithm for daytime advanced very high resolution radiometer imagery. *Journal of Geophysical Research*, 95(C8):13,459–13,481, August 1990.

[12] Bruce A. Wielicki and Lindsay Parker. On the determination of cloud cover from satellite sensors: The effect of sensor spatial resolution. *Journal of Geophysical Research*, 97(D12):12,799–12,823, August 1992.

## Biographies

**Ben Hollingsworth** received his B.S. degree in Computer Science from the University of Nebraska—Lincoln in 1994. He is presently pursuing his M.S. in Computer Science also from the University of Nebraska—Lincoln.

**Liqiang Chen** received the B.S. degree in Meteorology at Nanjing Institute of Meteorology, P. R. China, in 1985, and the M.S. degree in Meteorology at South Dakota School of Mines and Technology, Rapid City, in 1993.

Now he is a graduate student for the M.S. degree in computer science. His present research interests include image processing, neural networks, and their application to remote sensing.

**Stephen E. Reichenbach** is an assistant professor in the Computer Science and Engineering Department of the University of Nebraska—Lincoln.

He earned a PhD at William & Mary (1989), MS at Washington University in St. Louis (1985), and BA at the University of Nebraska (1976). He held a 1995 ASEE Faculty Research Fellowship at the NASA Goddard Space Flight Center and previously held a National Research Council Post-Doctoral Research Fellowship at the NASA Langley Research Center and a George Holmes Faculty Research Fellowship at the University of Nebraska.

Professor Reichenbach is an associate editor for the IEEE Transactions for Image Processing and serves on the Steering Committee and the Technical Program Committee for the 1996 International Geoscience and Remote Sensing Symposium. Professor Reichenbach was awarded the UNL Engineering and Technology College Outstanding Faculty Research Award (Assistant Professor, 1994) and has published more than two dozen papers on image restoration, compression, system design, and analysis.

**Richard R. Irish** received the B.S degree in Range-Forest Management from the Colorado State University in 1977 and the M.S. degree from the Pennsylvania State University, in 1982.

Following his graduate work, he held a research associate position for two years at Hunter College in New York City. He came to NASA's Goddard Space Flight Center in 1984 and developed a calibration package for the Advanced Solidstate Array Spectroradiometer — a 30 channel airborne imager that flew on NASA's C-130. In 1989 he joined Hughes STX Satellite Mapping Technologies to promote the growth of the company's commercial remote sensing business. In 1993, he joined Science Systems Applications, Inc. to work on NASA's Landsat-7 program. He is responsible for improving cloud cover detection performance and designing distribution products for users.