Retinex Preprocessing for Improved Multi-spectral Image Classification

B. Thompson, Z. Rahman, and S. Park

College of William & Mary, Department of Computer Science, Williamsburg, VA 23187

ABSTRACT

The goal of multi-image classification is to identify and label "similar regions" within a scene. The ability to correctly classify a remotely sensed multi-image of a scene is affected by the ability of the classification process to adequately compensate for the effects of atmospheric variations and sensor anomalies. Better classification may be obtained if the multi-image is preprocessed before classification, so as to reduce the adverse effects of image formation. In this paper, we discuss the overall impact on multi-spectral image classification when the retinex image enhancement algorithm is used to preprocess multi-spectral images. The retinex is a multi-purpose image enhancement algorithm that performs dynamic range compression, reduces the dependence on lighting conditions, and generally enhances apparent spatial resolution. The retinex has been successfully applied to the enhancement of many different types of grayscale and color images. We show in this paper that retinex preprocessing improves the spatial structure of multi-spectral images and thus provides better within-class variations than would otherwise be obtained without the preprocessing. For a series of multi-spectral images obtained with diffuse and direct lighting, we show that without retinex preprocessing the class spectral signatures vary substantially with the lighting conditions. Whereas multi-dimensional clustering without preprocessing produced one-class homogeneous regions, the classification on the preprocessed images produced multi-class non-homogeneous regions. This lack of homogeneity is explained by the interaction between different agronomic treatments applied to the regions: the preprocessed images are closer to ground truth. The principle advantage that the retinex offers is that for different lighting conditions classifications derived from the retinex preprocessed images look remarkably "similar", and thus more consistent, whereas classifications derived from the original images, without preprocessing, are much less similar.

Keywords: image classification, image enhancement, multi-spectral, retinex, dynamic range compression

1. INTRODUCTION

The analysis of remote sensed imagery obtained over agricultural regions can be used for the detection and discrimination of "stressed" and "non-stressed" vegetation; this is an issue of considerable importance to the agriculture industry. The terms "stressed" and "non-stressed" are used in a qualitative sense to designate the relative plant growth over different regions of a field. The differences in growth patterns can be due to several factors including different agronomic treatments. Various algorithms for the discrimination and detection of vegetation using remote sensed imagery exist in the literature. One way to characterize these algorithms is by the characteristics of the multi-dimensional space in which they operate. For example, many users of remote sensed imagery use spectral signatures to characterize and identify materials in multi-dimensional "spectral" space. The spectral signature of a material can be defined in the solar-reflective spectral region by its reflectance as a function of wavelength, measured at an appropriate spectral resolution. In other spectral regions, signatures of interest are temperature and emissivity (Thermal Infrared TIR) and surface roughness (radar).[1]

There are fundamental problems with the spectral signature approach that are well documented in the literature. First, all spectral signatures are unique to the sample and the environment in which they are obtained. Second, the ability to distinguish spectral signatures is often complicated by natural variability for a given material, spectral quantization of many remote-sensing systems, and modification of signatures by the atmosphere as a result of the image formation process.[1] Therefore, even though one may wish to apply different "labels" to differentiate vegetation

Further author information: (Send correspondence to B.T.)
B.T.: E-mail: bthompso@cs.wm.edu
Z.R.: E-mail: zrahman@cs.wm.edu
S.P.: E-mail: park@cs.wm.edu

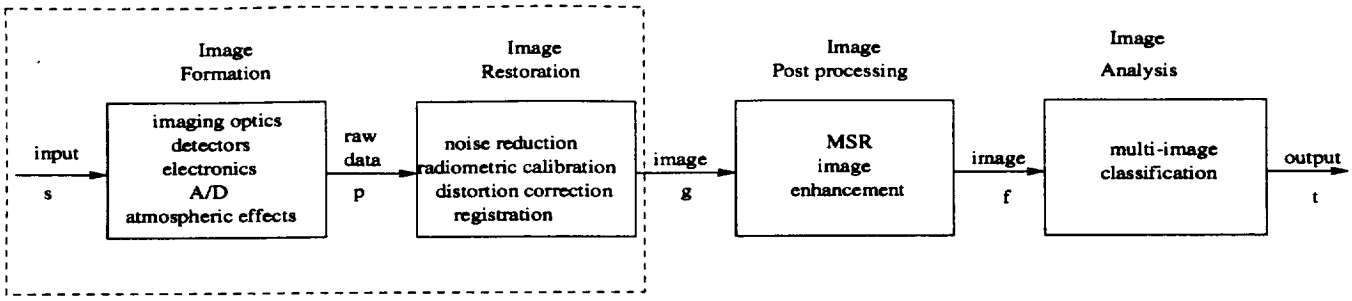


Figure 1. System Model

signatures, there is no guarantee that the signatures obtained by the remote sensing system will exhibit measurably different, or even recognizable, signatures.

In recent years, a considerable amount of ground-based (laboratory) data have been accumulated that describe spectral reflectance characteristics of soils and vegetation, without the problem of atmospheric complications. It is difficult, however, to duplicate natural reflectance measurements under laboratory conditions. The comparison between natural reflectance signatures and laboratory produced signatures, therefore, becomes even more complicated. Furthermore, the spectral signature of vegetation also changes over the seasonal life cycle of many plants, acquiring a "yellow" characteristic in senescence,* with a corresponding increase in the red region reflectance caused by a loss of photosynthetic chlorophyll.[1]

As an alternative to classification based on spectral signatures, multi-dimensional spectral space is transformed into a "feature" space prior to classification. In this way, information in the image is redistributed into a different and, depending on the application, more useful form. For example, transformations such as multi-spectral ratios of Near Infrared (NIR) to visible bands have been used to enhance reflectance differences between soils and vegetation and form "vegetation indices" that aid in classification. In this way, soil and other geological formations will exhibit similar ratios near 1, while vegetation will show a relatively larger ratio of 4 or more. Other common vegetation indices are the Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), Transformed Vegetation Index (TVI), and the Perpendicular Vegetation Index (PVI).[1] The success of using these indices in the past has been affected by relatively few acquisition dates during a growing season, the paucity of ground truth data at the time of acquisition, and the lack of suitable methods to account for atmospheric effects on the radiance measured by the remote sensing device.[2]

Whether a particular classification algorithm uses spectral signatures or multi-spectral ratio indices to facilitate discrimination and detection of vegetation changes in an agricultural region, either technique requires good radiometric calibration of the image before analysis can be performed. Figure 1 illustrates the major steps in the image classification process. Radiometric calibration, a fundamental stage in this process, generally involves (1) sensor calibration: at-sensor radiance values obtained from quantized data during A/D conversion, (2) atmospheric correction: surface radiance values obtained from at-sensor radiance, and (3) solar and topographic correction: surface reflectance values obtained from surface radiance. Usually, detailed information about atmospheric conditions is not available for a given data set. However, parametric atmospheric correction methods can generally be used to compensate for atmospheric conditions. The success of multi-image classification in the analysis stage is based on the quality of these methods.

In this paper, we approach multi-image classification differently. Instead of using band ratios or absolute spectral signatures, we use "relative" signatures in the image to discriminate and detect vegetation changes. We compensate for the atmospheric effects on the multi-spectral images by applying the multi-scale retinex[3] (MSR) image enhancement algorithm to the multi-image, prior to image classification.

2. AGRONOMIC DATA

For our analysis we used remote sensed images of a cotton field in Texas acquired in the summer of 1997. We chose two multi-spectral images of the field taken on consecutive dates. The first image (acquired 08/14/97) has overcast

*Senescense is the physiological death of plants.

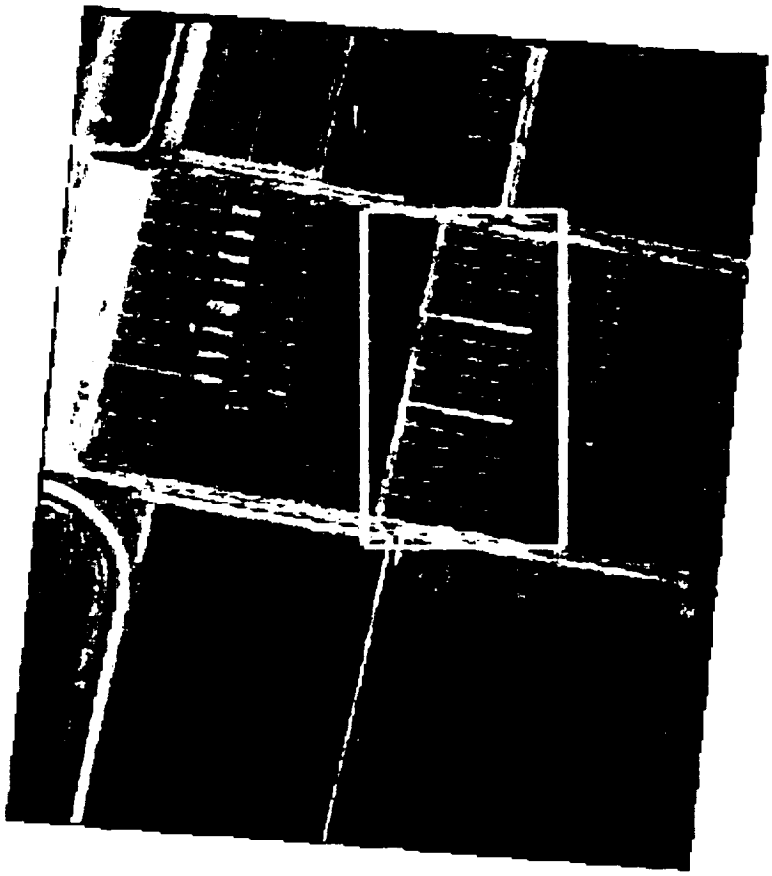


Figure 2. Test site captured on 08/14/97. The area of interest is highlighted within the rectangle.

sky, i.e. diffuse light, and the second (acquired 08/15/97) has almost clear sky, i.e. direct sunlight. The highlighted sub-image in Figure 2 is the area of interest for this experiment. We label the images of this area TXoa for date 08/14/97 and TXob for date 08/15/97. The cotton field was the site of a controlled experiment to study the effects on vegetation growth of different combinations of water and nitrogen treatment levels for a particular soil tillage type.[†] In all, 4 water treatment levels, 5 nitrogen treatment levels, and 2 soil tillage types were used. Figure 3 shows a schematic of the treatment experiments applied to the TXoa and TXob regions.

The field was divided into 120 blocks representing $4 \cdot 5 \cdot 2 = 40$ unique combinations of water, nitrogen, and tillage type. Each combination was repeated 3 times over the whole field. The 4 irrigation levels used were: 0.00, 0.25, 0.50 and 0.75 (fraction) of potential evapotranspiration (PET)[‡]. The 5 fertilizer nitrogen application levels were: 0, 20, 40, 60 and 80 lbs/acre Best Practice (BP)[§].

In theory, each of these 40 unique blocks represents a different "spectral class" and there are three samples of each class. However, classification results show that the number of actual spectral classes is fewer than 40. Moreover, the blocks are generally not homogeneous. That is, within each block there are *mixed* areas where the levels of water and/or nitrogen treatment are not uniform. In addition, the ground truth data was reliable for water treatment, but

[†] Tillage prepares the soil for growing crops. This preparation is traditionally accomplished by using a plow to cut and mix the soil.

[‡] Evapotranspiration (ET) is a measurement of the total amount of water needed to grow plants and crops. Since there are thousands of cultivated plants, the potential ET (PET) is a standard ET rate for general reference and use. The water requirements of specific crops and turf grasses can be calculated as a fraction of the PET. This "fraction" is called the crop coefficient (Kc) or turf coefficient (Tc)[4]

[§] Best Management Practices are farming practices that are designed to reduce nutrient contamination of surface and ground water. These practices are based on research results and field experiments and maybe as simple as following fertilizer recommendations and irrigation scheduling.[5]

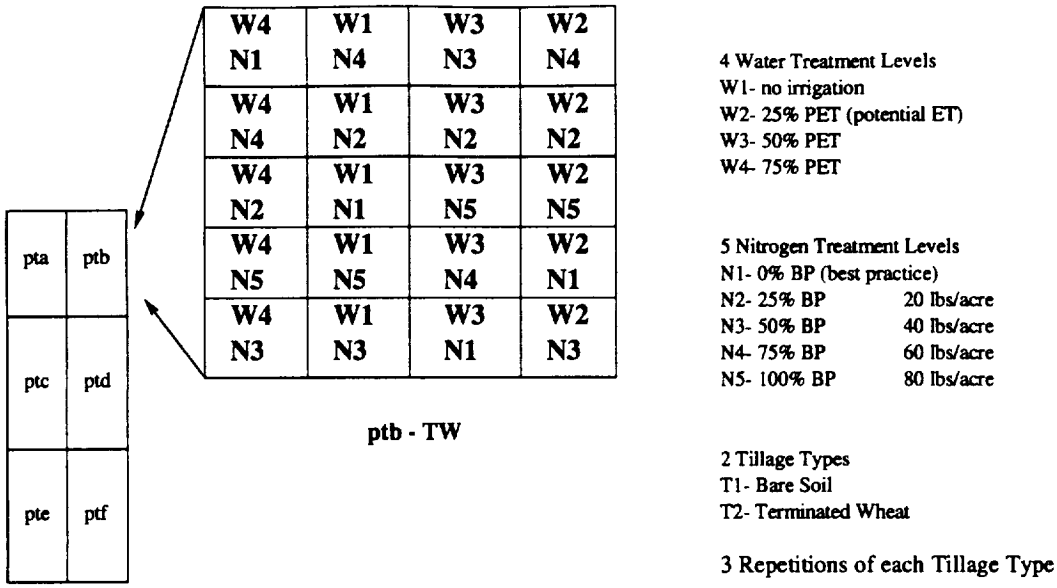


Figure 3. Agronomic treatments applied to the area of interest in Figure 2.

not for nitrogen treatment. The four-band multi-spectral images were acquired from an aircraft platform with an approximate nadir view, and calibrated to reflectance. The aircraft multi-spectral sensor band centers were at 486, 560, 685, and 840 nm.

3. IMAGE PREPROCESSING

3.1. The Multi-scale Retinex

For all (x, y) pixels in the multi-spectral image G, the multi-scale retinex (MSR)[6,7] can be compactly written as

$$F_j(x,y) = \sum_{n=1}^{N} W_n \cdot \{\log[G_j(x,y)] - \log[G_j(x,y) * H_n(x,y)]\} \tag{1}$$

where the subscript j represents the spectral bands, N is the number of spatial scales being used, and W_n are the weighting factors for the scales.[8,9,3] The $H_n(x,y)$ are the surround functions (convolution kernels) given by

$$H_n(x,y) = I_n \exp[-(x^2 + y^2)/\sigma_n^2], \tag{2}$$

where the σ_n are spatial scale parameters that control the extent of the surround function. Smaller values of σ_n provide more dynamic range compression, and larger values provide more lightness/color constancy. The I_n are selected so that $\sum\sum H_n(x,y) = 1$. Each of the expressions within the summation represents a single-scale retinex (SSR).

The MSR combines the dynamic range compression of the small scale retinex with the tonal rendition of the large scale retinex to produce an output which encompasses both. Two fundamental issues in the application of the MSR are the following:

1. The MSR reduces dependency on lighting conditions/geometry caused by such conditions as obscured foregrounds, and poor lighting caused by defects in illumination due to atmospheric conditions or artificial illuminants.

2. As atmospheric conditions change, the MSR will produce results such that the restored image f in Figure 1 will look as much like the original scene s as possible, before image acquisition/digitization.

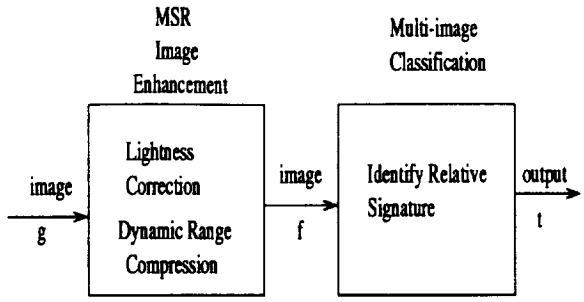


Figure 4. Post processing system model.

4. DISCUSSION

There were two primary motivations for this study: (a) to what extent can a conventional unsupervised classification algorithm yield "good" results when applied to the original images "as is" (i.e., with no preprocessing); and (b) if the multi-spectral images are preprocessed with the retinex algorithm and then the same conventional unsupervised classification algorithm applied, to what extent does that retinex processing influence the "goodness" of the result? This section summarizes the results of our initial experiments.

4.1. MSR Pre-classification Processing

The MSR was used to preprocess the multi-spectral image before it was used for classification. Figure 5 shows the results of the MSR algorithm applied to band 3 (685 nm) of the image in Figure 2. The left column shows the original images: TXoa and TXob, for the cloudy and clear day. The right column shows the MSR processed images: TXra and TXrb. In TXoa the band reflectance is uniformly very dark, because of the cloud cover on that day. The low contrast in this image creates a problem in obtaining spectral signatures that adequately discriminate agronomic variables. After the image is processed with the MSR, subtle patterns emerged that were not visible in the TXoa image. Specifically, the patterns represent the boundaries between the 20 different nitrogen and water treatment regions. The MSR also improved the TXob image, obtained on the clear day. One of the primary results from the application of the MSR is that the processed images, TXra and TXrb, are more "similar" to each other in brightness, contrast and detail than the original images, TXoa and TXob.

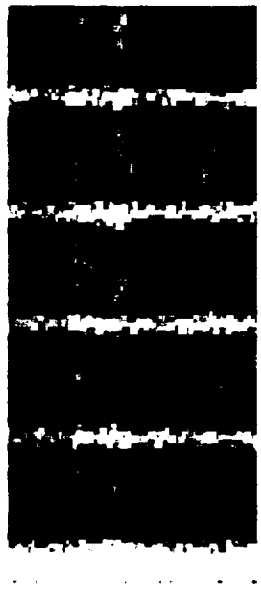
The MSR processed images display far more visual information than is evident in the unprocessed images. Even though radiometric calibration is not preserved by the MSR, we conclude that it can be used as an auxiliary tool for the visualization of spatial patterns in dark regions, as is demonstrated herein. Visual information in darker regions that may not be detected with linear representations which preserve radiometry will "pop out" with a clarity limited only to the dynamic range of the sensor and any intervening digitalization scheme used prior to the MSR. For this experiment, we have not yet conducted extensive performance comparison of the MSR with other image enhancement algorithms such as histogram equalization, gamma correction, and point logarithmic nonlinearity. However, we expect to find that those image enhancement algorithms are not appropriate for use in preprocessing multi-spectral images for remote sensing applications where atmospheric conditions are the major contributor to data inaccuracy.

4.2. Multi-image Classification

The literature is rich with both supervised and unsupervised methods for classifying remote sensed multi-spectral images. These methods include spatial/spectral discriminant functions, e.g., the Maximum Likelihood, and spectral specific methods, e.g., linear mixing models that require some a priori knowledge such as ground truth maps or ground samples. Whereas, supervised classification requires training sets to teach the classifier to recognize certain specific features in the image, unsupervised methods require little or no training data and attempt to discover the underlying patterns in multi-dimensional space by using techniques such as gradient descent. In this experiment, we have limited a priori information available about the field being analyzed. The a priori information was used to verify the accuracy of the classes obtained using unsupervised classification.

We used vector quantization (VQ) to perform unsupervised classification on the multi-spectral image. The only user specified parameter is the number of classes K. The primary goal was to study and compare the effects of MSR

08/14/97 TXoa



08/14/97 TXra

08/15/97 TXob

08/15/97 TXrb

Figure 5. TX Images– Original and retinex processed images.

Table 1. List of constants used to process the TXoa and TXob images with the MSR

Constant	σ_1	σ_2	σ_3	σ_4	Gain	Offset
Value	2	5	20	200	180	0.57

preprocessing on the classification results. The same classification algorithm was applied to both of the *original* four-band images TXoa, TXob and both of the *MSR* four-band images TXra, TXrb. For all four images, we systematically experimented with K to see how the number of classes affects the overall classification results.

To cluster the images we used VQ along with a splitting method to define the spectral signatures. The method starts with a one-level quantizer (i.e., the centroid of the entire training set). Next, the one-level quantizer vector is split into two vectors obtained by perturbing the one-level quantizer. The 2-level quantizer is then applied to the training set. The two 2-level quantizer vectors are then split into four vectors and a 4-level quantizer is applied to the training set. The splitting is continued in this manner until K code vectors are generated. This method assumes that K is a power of two. If K is not a power of two, then in the last step, instead of generating two new code vectors from each of the code vectors of the quantizer designed previously, we can perturb as many code vectors as necessary to obtain the desired number of code vectors.[10] As in most classification methods, the performance depends on the quality of the set of spectral means used to discriminate classes in the image. For this analysis, we did not focus on methods to obtain spectral means, but compared the relative accuracy of the spectral means obtained by the VQ to signatures derived from the training areas defined by the schematic map of Figure 3.

5. CLASSIFICATION RESULTS

Figures 6 and 7 show the classification results for the original and MSR preprocessed images. To examine the accuracy of our results we compare these results to the schematic plot of the proposed treatment of water and nitrogen for the cotton field shown in Figure 3. To facilitate analysis, the figures are annotated with a grid that provides an approximate separation boundary between each treatment level block.

Our classification results were very encouraging for a number of classes, from as few as $K = 4$—(i.e., 4 water treatment levels), to as many as $K = 40$—(i.e., 4 water treatment levels $\times$ 5 nitrogen treatment levels $\times$ 2 tillage types). For the case $K = 4$, we were interested in determining how well the four water treatment levels could be discriminated in the image. In the case $K = 8$, we were interesting in discriminating the four water treatment levels for each tillage type. From our results we did not see any major differences between the different tillage types in terms of classification results. That is to say, we could not resolve two different classes of each water treatment type. The primary effect of the case $K = 8$ was that we were able to see more clearly the water-nitrogen iterations.

Generally, for all four images we were able to see very clearly the "block" treatment structure that is present. There are differences in the results, however, depending on whether the lighting was diffuse or direct and depending on whether or not retinex preprocessing was used. The left column of Figures 6 and 7 show that without MSR preprocessing the blocks tend to be classified as homogeneous (one class); the right column of Figures 6 and 7 show that with MSR preprocessing the blocks tend to be classified as more non-homogeneous (multi-class). Comparison of the classification results for TXoa, TXob (original images) and TXra, TXrb (MSR processed images), show considerable variation in classes for the unprocessed images as the atmospheric and lighting conditions vary, but slight or no variations for the MSR processed images.

Because consistent classification results are achieved regardless of the atmospheric conditions, we can argue empirically that MSR preprocessing tends to produce "spectral signature images". Note that classification consistency in this experiment is really a measure of the resiliency of the classification process to changes in the process that affect the formation of the multi-spectral image. In other words, classification consistency is really a measure of how well we can classify the multi-spectral image given that the atmospheric conditions have changed substantially from day to day. To illustrate classification consistency, in Table 2 we show the mean spectral reflectance measurements for the original and MSR preprocessed images for each class. From this table we see that the spectral signatures for the MSR preprocessed images for each day are more similar than the signatures for the original images.

Initially, we expected to see one different class for each column in the field representing a different water treatment level and not the multi-class variations that are shown in Figures 6 and 7. However, because the nitrogen and water

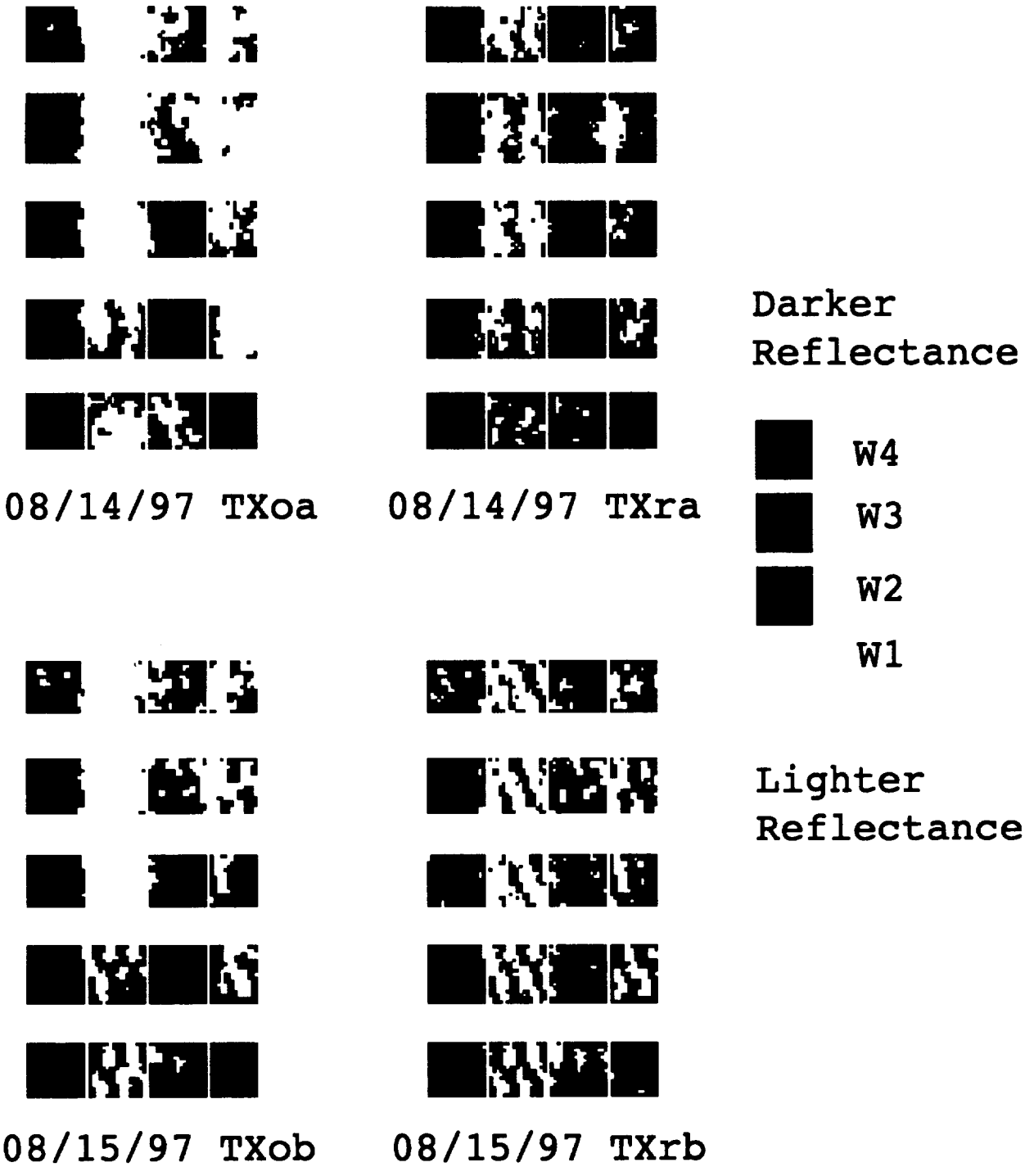


Figure 6. Classification Results: 4 Classes

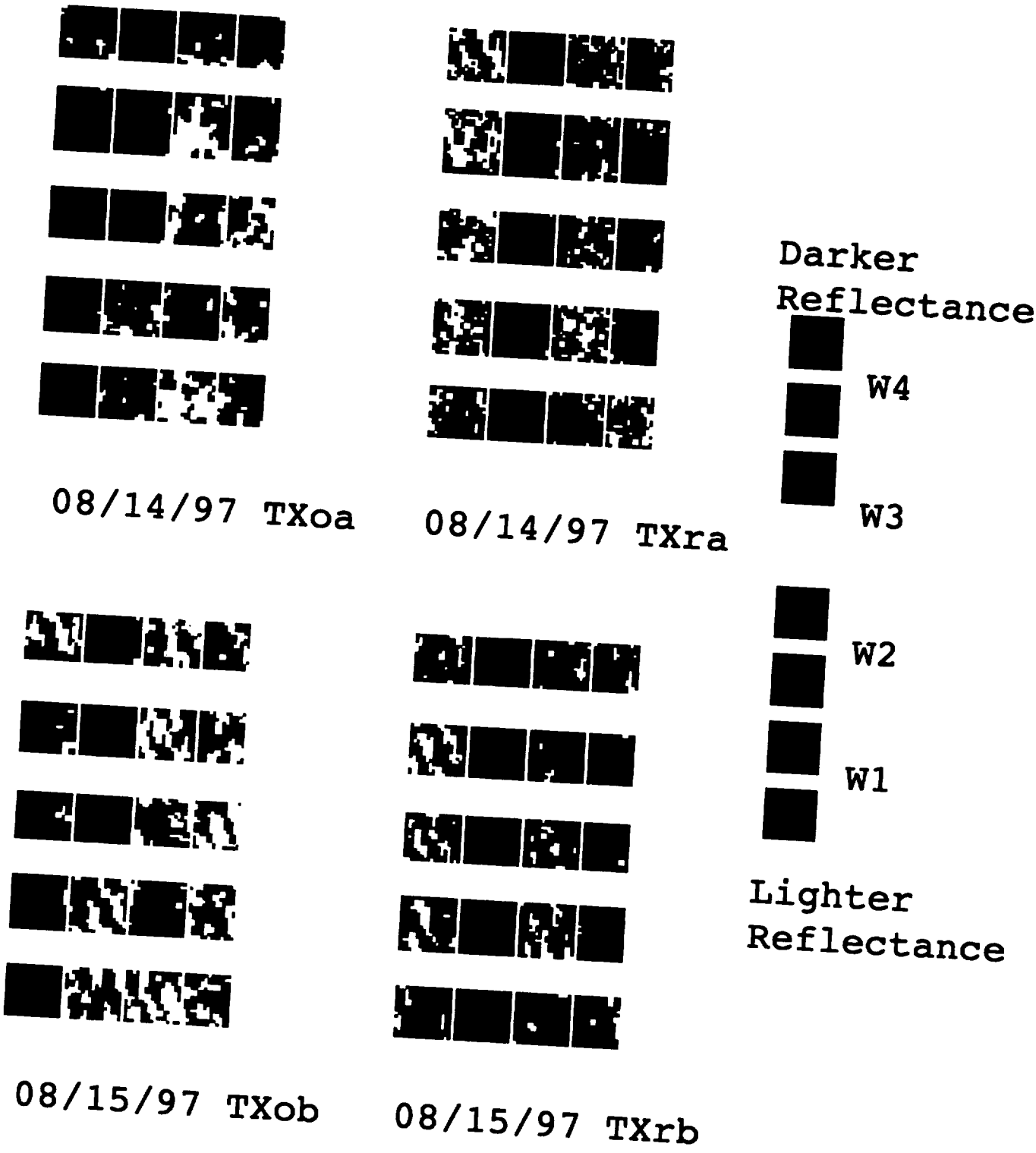


Figure 7. Classification Results: 8 Classes

Table 2. Mean spectral reflectance measurements for each class.

Class	Wavelength (nm)							
	486	560	685	840	486	560	685	840
	08/14/97 diffuse				08/14/97 MSR			
W4	55	49	33	102	140	138	123	161
W3	58	52	38	93	146	145	139	155
W2	60	54	41	104	151	148	151	146
W1	63	58	45	90	158	157	167	144
	08/15/97 direct				08/15/97 MSR			
W4	115	113	66	201	124	119	92	156
W3	127	128	83	195	143	140	126	152
W2	136	138	97	190	146	148	159	149
W1	148	149	106	190	169	167	173	147

Table 3. Mean spectral reflectance measurements for each column in Plotb of Figure 3.

Column	Wavelength (nm)								Class
	486	560	685	840	486	560	685	840	
	08/14/97 diffuse				08/14/97 MSR				
1	54.83	50.68	33.49	103.65	140.46	141.93	125.12	162.60	W4
2	60.37	54.88	41.54	91.22	154.25	153.09	156.12	145.28	W1
3	56.77	52.15	36.85	94.85	143.66	143.80	134.99	153.09	W3
4	58.29	53.42	39.12	92.49	148.19	147.92	144.54	150.00	W2
	08/15/97 direct				08/15/97 MSR				
1	102.07	104.53	61.38	176.28	128.47	127.800	101.82	156.43	W4
2	124.47	126.46	91.42	162.21	157.97	157.38	164.08	146.14	W1
3	112.83	115.05	76.03	165.88	141.24	140.29	130.38	151.37	W3
4	117.06	120.30	83.67	162.03	147.88	147.31	145.53	149.41	W2

treatment affect the vegetation growth jointly, it would be a mistake to consider the two treatments independently. For example, the effect of applying water treatment level 4 and nitrogen level 3 may be the same as applying water treatment level 3 and nitrogen level 5. Therefore the original assumption that there are 40 distinct classes is, of course, not valid. In further study, we realized that what we are actually discriminating is the change in reflectance due to different nitrogen effects *and* water treatment, i.e., joint effects. In Table 2 we see that reflectance generally increased with water stress for all water levels in the 486, 560, and 685 nm wavelength, but decreased in the (NIR) 840 nm range. This increase in reflectance effect has been reported elsewhere as the effect water stress has on diffusive resistance and plant metabolism in general.[11] However, the increase was affected by the nitrogen treatments applied. These same measurements all showed a clear nitrogen and water stress interaction.[¶] Thus water treatment labels for each class were identified as W4 for the darkest spectral reflectance in the 486, 560, and 685 nm range and W1 for the lightest reflectance in that same range.

To identify the water treatment levels, we matched the mean spectral reflectance measurements of each column in Table 3 for the original and MSR processed images shown in Figure 3 with the mean spectral reflectance for each class shown in Table 2. From this analysis we were able to conclude that the mean spectral reflectance for each column matched the correct mean spectral reflectance for each class. Although the magnitude of the results may be different for other tillage types, the results presented here should prove useful for determining the amount of information that can be expected from particular agronomic variable interactions for given atmospheric conditions.

[¶]This merits a re-examination of the data, which is the subject for another paper.

6. CONCLUSIONS

Spectral signatures alone do not provide adequate classification of a scene, especially if the atmospheric or lighting effects have severely affected the multi-image components.[1] This is evident if we compare the classification results before and after MSR preprocessing. Without additional ground truth, or results from other classification studies, it is difficult to state with any confidence whether the classifications obtained with the preprocessed images are "better" in some absolute sense than the classifications obtained from the original images. We can state, however, that the classifications from the MSR preprocessed images for the two different lighting and atmospheric conditions are remarkably "similar" both visually, and in terms of the mean spectral reflectance of a class. We speculate that this occurs because the MSR preprocessing is minimizing the effects of the atmospheric conditions on the multi-spectral image, leading to consistent classifications from consistent data.

To summarize, we conclude that conventional unsupervised classification can be applied to this significant problem of detection and discrimination of stressed and unstressed vegetation. Although classification results from both the original and the MSR preprocessed images are encouraging, the MSR preprocessed images are more robust to changes in atmospheric and lighting conditions. We need to conduct additional experiments to test the validity of our speculation that MSR preprocessed multi-image classification is more robust in the presence of atmospheric and lighting changes. In addition, we need to substantiate our conjecture that other image enhancement algorithms do not have the same "beneficial" effect on the classifications. A color version of the figures in this paper is available at http://dragon.larc.nasa.gov/viplab/retinex/background/pubabs/spie3716-1999.html.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Pamela Blake, Dr. Joan Hayashi, and Jim Sweet of GDE Systems, Inc for the use of the multi-spectral images and ground truth data. This research was supported by the Virginia Space Grant Consortium (VSGC), NASA Graduate Student Researchers Program (GSRP), and NASA Langley Research Center Cooperative Agreement NCC-1-258.

REFERENCES

1. R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, Academic Press, 1997.
2. R. Jackson, P. Slater, and P. J. Pinter, "Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres," *Remote Sensing of Environment* , 1983.
3. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing* , 1997.
4. T. A. E. Service, *PET Documentation*, Texas Evapotranspiration Web Site: http://www.agen.tamu.edu/pet.
5. N. E. Derby, R. E. Knighton, and D. D. Steele, *Temporal and Spatial Distribution of Nitrate Nitrogen Under Best Management Practices*, North Dakota State University, Department of Soil Science, 1994.
6. E. H. Land, "The retinex theory of color vision," *Scientific American* , pp. 108–129, December 1977.
7. M. Olshaker, *The Instant Image: Edwin Land and the Polaroid Experience*, Stein and Day, Scarborough House, Briarcliff Manor, N.Y., 1978.
8. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multiscale retinex for color image enhancement," *International Conference on Image Processing ICIP'96* , 1996.
9. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing* , 1997.
10. K. Sayood, *Introduction to Data Compression*, Morgan-Kaufmann, 1995.
11. J. Schepers, T. Blackmer, W. Wilhelm, and M. Resende, "Transmittance and reflectance measurements of corn leaves from plants with different nitrogen and water supply," *J. Plant Physiol* (148), pp. 523–529, 1996.

A multiscale retinex for improved multispectral image classification

B. Thompson, Z. Rahman, S. K. Park

SPIE International AeroSense Symposium

Visual Information Processing IX

Orlando, FL

(April 2000)

A Multiscale Retinex for Improved Performance in Multispectral Image Classification

B. Thompson[a], Z. Rahman[b], and S. Park[b]

[a]Sandia National Laboratories, P.O. Box 5800, MS 1138, Albuquerque, NM 87185
[b]College of William & Mary, Department of Computer Science, Williamsburg, VA 23187-8795

ABSTRACT

Image preprocessing is useful in helping to identify "spectral response patterns" for certain types of image classification problems. The common artifacts in remotely sensed images are caused by the blurring due to the optics of the image gathering device, illumination variations, and the radiative transfer of the atmosphere. The Multi-Scale Retinex (MSR) image enhancement algorithm that provides dynamic range compression, reduced dependence on lighting conditions, and improved (perceived) spatial resolution has proven to be an effective tool in the correction of image degradations such as those in remote sensing images. In this paper, we measure the improvement in classification accuracy due to the application of the MSR algorithm. We use simulated images generated with different scene irradiance and with known ground truth data. The simulation results show that, despite the degree of image degradation due to changes in atmospheric irradiance, classification error can be substantially reduced by preprocessing the image data with the MSR. Furthermore we show that, similar to the results achieved in previous work, the classification results obtained from the MSR preprocessed images for various scene irradiance are more similar to each other than are the classification results for the original unprocessed images. This is evident in the observed visual quality of the MSR enhanced images even before classification is performed, and in the difference images obtained by comparing image data under different irradiance conditions. We conclude that the application of the MSR algorithm results in improved visual quality and increased spatial variation of multispectral images that is also optimal for certain types of multispectral image classification.

Keywords: image classification, image enhancement, multispectral, retinex, dynamic range compression

1. INTRODUCTION

Various algorithms for image classification using remotely sensed imagery exist in the literature. One way to characterize these algorithms is by the characteristics of the multi-dimensional space in which they operate. For example, many users of remotely sensed imagery utilize spectral signatures to characterize and identify materials in multi-dimensional "spectral" space. The spectral signature of a material may be defined in the solar-reflective region of the electromagnetic spectrum by its reflectance as a function of wavelength, measured at an appropriate spectral resolution. In other spectral regions, signatures of interest are temperature and emissivity (TIR) and surface roughness (radar). The motivation for using remote sensed data for material identification is that different types of materials exhibit different spectral signatures,[1] and so can be distinguished on this basis.

There are fundamental problems with the spectral signature approach that are well documented in the literature. One fundamental problem is that all spectral signatures are unique to the sample *and* to the environment in which they are obtained. Further, the ability to distinguish spectral signatures is often complicated by the natural variability of a material, the spectral quantization of many remote-sensing instruments, and the modulation of the signatures by the atmosphere in the image formation process.[1] Therefore, there is no guarantee that the spectral signatures obtained by the remote sensing system will either be similar to the ones obtained under a different environment, or exhibit measurably different, or even recognizable characteristics.

In recent years, considerable quantities of ground-based (laboratory) data have been accumulated that describe spectral reflectance characteristics of several types of soils and vegetation. However, it is virtually impossible to duplicate natural reflectance variations under laboratory conditions. In addition, the spectral signature of vegetation changes over the seasonal life cycle of plants. Thus the comparison between natural reflectance signatures and laboratory produced signatures becomes even more complicated.

Send correspondence to B.T.: bdthomp@sandia.gov. Co-authors: Z.R.: zrahman@cs.wm.edu; S.P.: park@cs.wm.edu

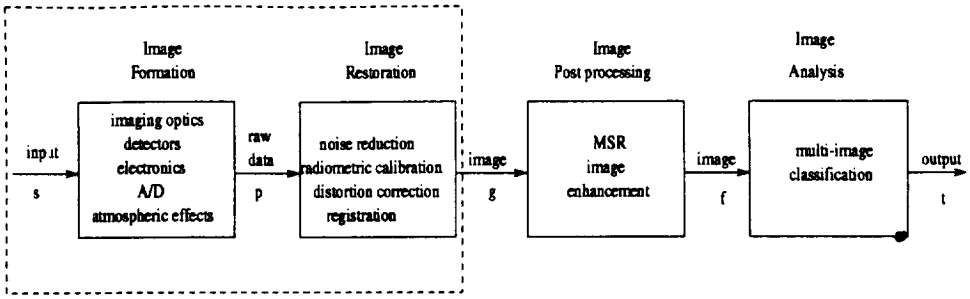


Figure 1. System Model.

As an alternative to classification based on spectral signatures, the multi-dimensional spectral space can be transformed into an application dependent "feature" space that may prove more useful for classification purposes. For example, transformations such as multi-spectral ratios have been used to enhance reflectance differences between types of soil and vegetation, and are used to form "vegetation indices" that aid in classification. Typically, soil and other geological formations exhibit ratios near 1, while vegetation shows a ratio of 4 or more. Other transformations such as scale space filtering have been applied to identify "fingerprints" of certain types of minerals using local points of inflection to characterize absorption characteristics. The success of using these indices has been greatly affected by the lack of suitable methods to account for atmospheric effects on the radiance measured by the remote sensing device.[2]

Figure 1 illustrates the major steps that relate the image acquisition to the image classification process. Regardless of which feature classification method is to be used, the (raw) image data needs to be radiometrically calibrated before it can be used for analysis. That is, the raw data must be converted from sensor DN values to surface reflectance values. Radiometric calibration generally involves several steps: sensor calibration—calculation of gain and offset coefficients that convert sensor DN values to at-sensor radiance values; atmospheric correction—converion of at-sensor radiance values to surface radiance using atmospheric modeling and estimation correction techniques; and, solar and topographic correction—conversion of surface radiance to surface reflectance by correcting for topographic slope and aspect, solar spectral irradiance, solar path transmittance, and down-scattered "skylight" radiance.[1] In addition, detailed information about atmospheric conditions at the time of data acquisition may be required but is generally not available. Parametric atmospheric correction methods can also be used to compensate for atmospheric conditions, but they also require some information about the atmospheric conditions at the time of data acquisition. The success of multi-image classification in the analysis stage is, thus, dependent on the quality of these calibration and correction methods.

In this paper, we approach multi-image classification differently. Instead of applying (parametric) atmospheric correction methods to remote sensed imagery, we compensate for the atmospheric effects by applying the multi-scale retinex (MSR) image enhancement algorithm to the multispectral data prior to classification. The dynamic range compression and color constancy properties of the MSR aid in minimizing the effects of variations in illumination conditions, and the sharpening compensates for the device blurring and atmospheric amplitude modulation.

2. IMAGE PREPROCESSING

There are many factors that contribute to degrade the acquired image. For example, the device signal-to-noise ratio (SNR) and the blurring due to the point spread function (PSF) of its optics, and the quantization artifacts due to the analog-to-digital converters are produced by the image acquisition device. In addition, platform perturbations, atmospheric modulations, and sampling artifacts also degrade the acquired image. Most researchers agree that geometric and radiometric artifacts are the most common cause of image degradations in remotely sensed imagery.[3] With reference to Figure 1 image restoration is an attempt to make the restored image g be geometrically and radiometrically as "close" as possible to the radiant energy characteristics of the original scene s. The *closeness* is measured in some metric space, such as the minimum mean square error (MSRE) space, and the goal of the restoration process is to minimize the MSRE between the restored image and the original scene. Although, generally termed an image enhancement technique, the MSR has proven to be an effective technique for correcting image

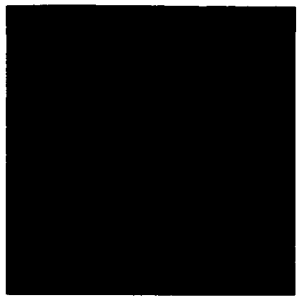


Figure 2. 2-d Mondrian.

degradations due to the optical blurring of the image acquisition device, illumination variations, and atmospheric modulation. Thus, the MSR can be used to "restore" the acquired data, without any prior knowledge about the atmospheric conditions at the time of acquisition.

2.1. The Multi-scale Retinex

For all (x, y) pixels in the multi-spectral image G, the multi-scale retinex (MSR)[4,5] can be compactly written as

$$F_j(x,y) = \sum_{n=1}^{N} W_n \cdot \{\log[G_j(x,y)] - \log[G_j(x,y) * H_n(x,y)]\}, \quad j = 1, \cdots, J \tag{1}$$

where J represents the number of spectral bands, N is the number of spatial scales being used, and W_n are the weighting factors for the scales.[6–8] The $H_n(x, y)$ are the surround functions (convolution kernels) given by

$$H_n(x,y) = I_n \exp[-(x^2 + y^2)/\sigma_n^2], \tag{2}$$

where σ_n are the spatial scale parameters that control the extent of the surround function and the I_n are selected so that $\sum \sum H_n(x, y) = 1$. Smaller values of σ_n provide more dynamic range compression, and larger values provide more lightness/color rendition. Each of the expressions within the summation represents a single-scale retinex (SSR).

The MSR combines the dynamic range compression of the small scale retinex with the tonal rendition of the large scale retinex to produce an output which encompasses both. The MSR reduces dependency on lighting conditions/geometry caused by such conditions as obscured foregrounds, and poor lighting caused by atmospheric conditions or defects in artificial illuminants.

3. SCENE GENERATION

In order to exactly measure the effectiveness of MSR preprocessing on multi-image classification, we use simulated images with exactly known ground truth. The simulated images are created by combining known atmospheric transmittance profiles with scenes with known mean spatial detail and surface reflectance. We use a simple model[9,1] of the radiance field which has the following characteristics:

1. The scene is a two-dimensional Mondrian flat surface divided into patches of uniform reflectance.
2. The effective irradiance (or the atmospheric transmittance) $I(x, y)$ varies slowly and smoothly across the entire scene, and
3. The reflected radiance field $L(x, y)$ is everywhere independent of the viewer's position.

These assumptions permit us to express the radiance field $L(x, y)$ by the simple relationship

$$L(x,y) = \frac{1}{\pi}\rho(x,y)I(x,y), \tag{3}$$

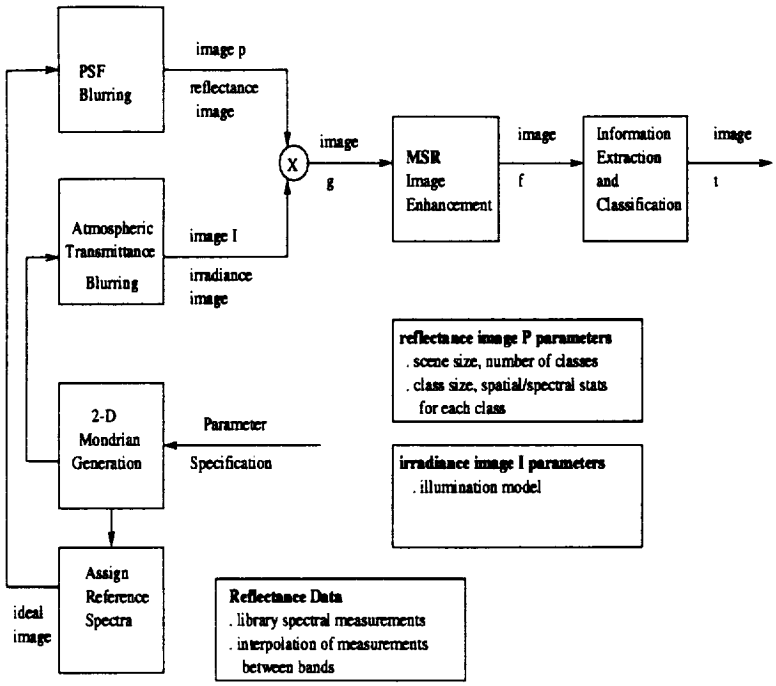


Figure 3. Scene generation system model.

where $\rho(x,y)$ is the Lambertian surface reflectance. The target scene with reflectance $\rho(x,y)$ is represented by the two-dimensional (2-d) Mondrian illustrated in Figure 2. This scene consists of random polygons whose boundaries are Poisson distributed and whose reflectances are distributed according to independent zero-mean Gaussian statistics.

The scene used in the simulation has 256 x 256 pixels. The reference spectra (See Section 3.1) were subsampled at every $0.05\mu m$ to produce ideal spectra of 42 points. This results in a multi-image with of 256 x 256 x 42 values. Figure 3 illustrates the process by which the 42 band real multi-image is generated. The multi-image is generated by assigning each spatial location in the Mondrian a number corresponding to a specific reference spectra that corresponds to the identification number for a vegetation type.

The atmospheric transmittance function is simulated by significantly blurring a generated Mondrian image. In addition to providing regions of different transmission characteristics, this also simulates the umbra and penumbra profiles of shadows. Two instances of the atmospheric transmittance profile are shown in Figure 4: the dark regions represent absorption in the atmosphere, or clouds; and the light regions represent transmittance, or absence of clouds. The first profile, atrans1, has two moderately sized regions of low transmittance (< 30%) at the lower left and right corners of the image. For the majority of the scene the transmittance is about 45%, with some higher transmittance areas located at the bottom center portion of the image which have a transmittance of about 80%. The second profile, atrans2, has on average a transmittance of about 65%. As shown in Figure 3, after the ideal scene image m and atmospheric transmittance profile m' have been generated, the simulated multi-image g is obtained by doing a pixel by pixel multiplication of m' and m,

$$g(x,y) = m(x,y) \cdot m'(x,y)$$

Finally, the real image g is processed with the MSR algorithm to create the processed image f.

In order to analyze the data, fidelity metrics are computed for the ideal, real, and the MSR processed images. Because we have "ground truth" data, these fidelity metrics can be applied and the results compared to the ground truth. The fidelity analysis is based on two metrics that measure the accuracy and consistency of the results as they are affected by the application of the two atmospheric transmittance profiles. A ***mean squared-error*** metric is used to measure the similarity between the two original images, and the similarity between the two real retinex images for each atmospheric transmittance profile. A ***sensitivity*** metric is used to measure classification consistency and

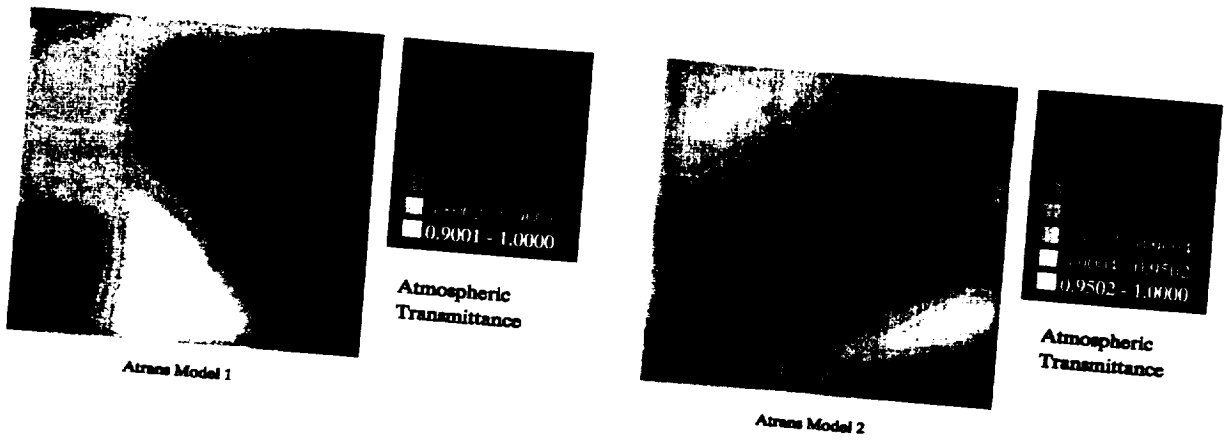


Figure 4. Transmittance Models.

accuracy for the real original and real MSR processed images using the real ground truth and the MSR ground truth as the basis of comparison.

3.1. Agronomic Data

The reference spectra used to create the ideal Mondrian scene were taken from an AVIRIS (Airborne Visual and Infrared Imaging Spectrometer) dataset for the San Luis Valley in Colorado.[10] Reference spectra plots for 9 representative species of vegetation are used in the scene (Figure 5), of which 8 are used in this experiment. The study used farmland reference spectra representing potatoes, alfalfa, barley, oat hay, canola, and open fields containing chico.[10] Based on the analysis provided by the USGS study by Clark et al,[10] the alfalfa, canola, oat hay, and nugget potato spectra showed the plants to be green and healthy. The barley had lost all of its chlorophyll signature. The norkotah potatoes were not being irrigated as they were about to be harvested, and consequently showed weak chlorophyll and cellulose absorptions, with soil (clay) absorptions from exposed soil. These potatoes were also being sprayed with a defoliant, so they showed decreased chlorophyll absorption, and a shift of the red edge of the absorption spectrum to shorter wavelengths. The chico and pasture spectra showed combinations of chlorophyll and cellulose (dry vegetation) absorptions. There was rain in the valley in the few days before the data acquisition flight so the chico and pasture did not show much water deprivation stress—being native plants they are hardy and can also withstand more reduced precipitation compared to the crops. The bare field calibration spectrum is from a sample measured on a laboratory spectrometer; all others are averages of several spectra extracted from the AVIRIS data.

3.2. MSR Preprocessing

Figures 6 and 7 show RGB and linear contrast stretched (LCS) composites of the real original and MSR processed scene. In Figure 6(l), the original scene has a reflectance of at most 60, thus making it difficult to compare it to the MSR processed image of Figure 6(l). However, in the LCS versions of the images we are able to compare features in both images. The most striking observation between the two images is that the MSR enhances details between class borders and within the class regions so that the sharpness of features in the image distinguishes it over that of the original image. We do note the appearance of edge artifacts within the borders of the regions for the MSR image. These edge artifacts are caused by Mach band undershoots and overshoots displayed as dark boundaries around the border of certain regions.[11]

4. DISCUSSION

Figure 8 shows the real original and MSR images created with the two atmospheric transmittance profiles shown in Figures 4a and 4b. The original images have been linearly stretched so that subtle differences between them and the MSR image can be compared. Recall that atrans1 (Figure 4(l)) has an average transmittance of 65%. Comparing the effects of the transmittance profiles on and atrans2 (Figure 4(r)) an average transmittance of 45%, the original and MSR images, it is evident that the MSR images appear visually consistent for both the models. However, the original images show the effects of the transmittance models as various dark reflectance areas in the images. For instance, the affect of the low transmittance areas is easily seen in the bottom left and right portion of

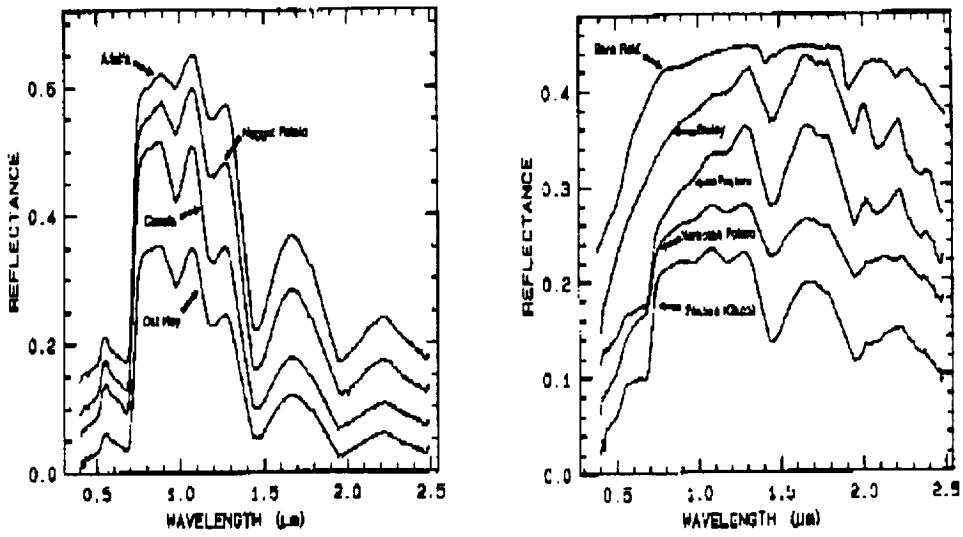


Figure 5. Reference spectra of vegetation and soil types. (Source: USGS Speclab.[10])

Figure 6. real 2-d Mondrian scene: (l) original (r) linear contrast stretched version.

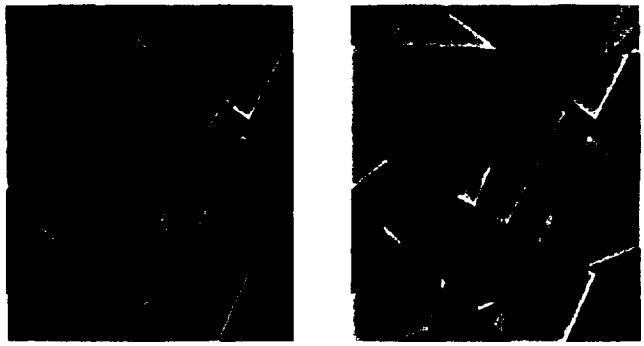


Figure 7. real 2-d Mondrian scene: (l) MSR processed (r) linear contrast stretched version.

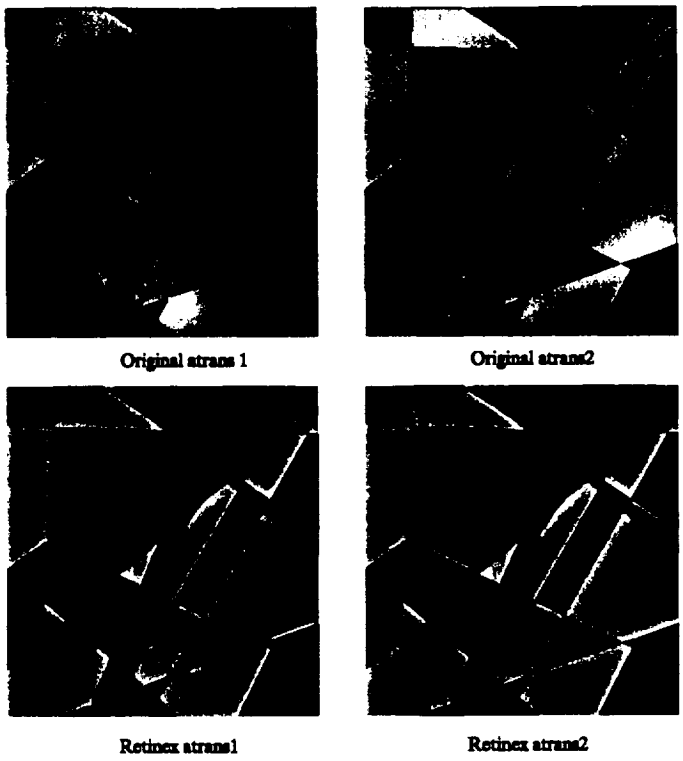


Figure 8. Effect of atmospheric models on original and MSR processed images.

the original atrans1 image. Very bright areas of reflectance in the original atrans2 image are also discernible in areas where the transmittance was the highest. In the MSR atrans1 image there is some slight darkening of reflectance in the lower left corner of the processed image, however, overall the MSR images are more similar to each other than are the unprocessed images. We also observe that the boundaries between regions in the MSR images are more clear and are in greater detail than the original. The previously identified edge artifacts are, however, evident at the transition between dark and bright areas in the processed images.

5. CLASSIFICATION

We used vector quantization (VQ) to perform unsupervised classification on the multi-spectral image. The only user specified parameter is the number of classes K. For classification, a 9 band subset of features were chosen from the original 42 band image.

To cluster the images we used VQ along with a splitting method to define the spectral signatures.[12] The splitting algorithm used to generate the trained codebook, splits each training set codebook vector using the best perturbation factor for that dataset. The preferred perturbation factor is the one that generates the smallest MSE for the input training set. The algorithm is designed to produce cluster means for a specific codebook size. As is the case for most classification methods, the performance depends on the quality of the set of spectral means used to discriminate classes in the image. For this analysis, we did not focus on methods to obtain spectral means, but compared the relative accuracy of the spectral means obtained by the VQ to signatures derived from the training areas

Because we have a ground truth map of our ideal classification, the training set will be selected from regions in the image shown in Figure 2. The training set vectors were input into the splitting algorithm and the trained codebook vectors were generated. In the testing stage, the images were classified with the trained codebook vectors, using a MSE VQ clustering algorithm. The resulting test codebook vectors were used as candidate spectral means to identify each vegetation species.

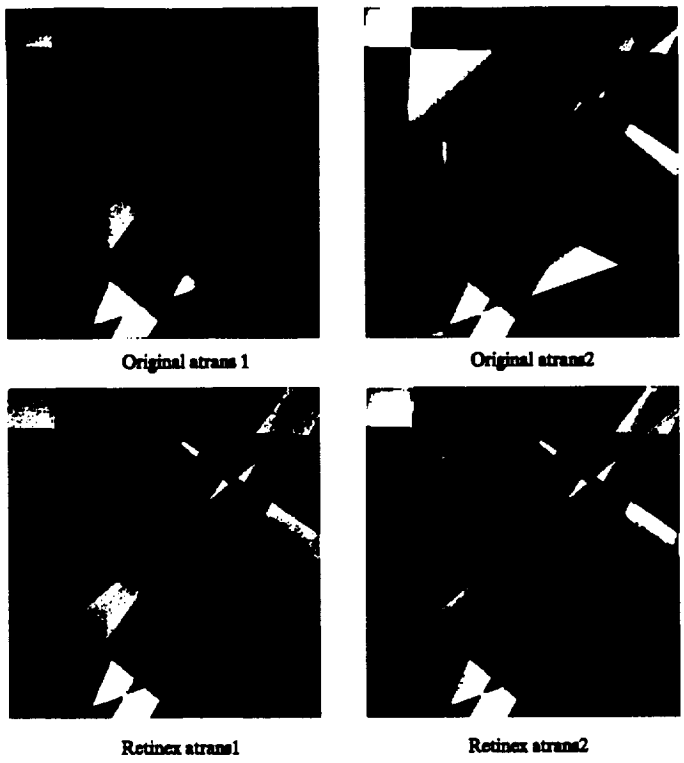


Figure 9. Classification results for original and MSR — Atmospheric Models Applied.

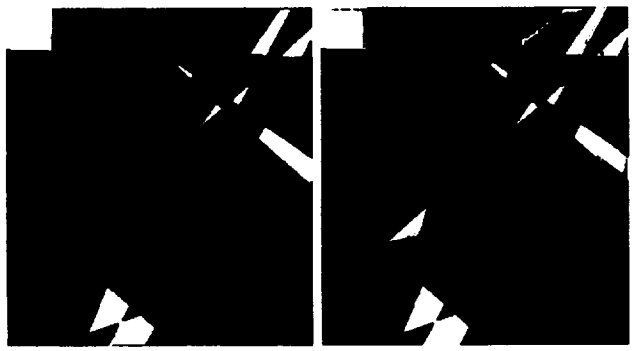


Figure 10. Classification results for original and MSR — No Atmospheric Models Applied.

Figure 9 shows the classification results for the original and MSR images for the two transmittance profiles. The classification results for the MSR processed images are more consistent with each other than the results obtained with the original images. However, there were problems separating certain class pairs such as pasture (c) and potato (nh), and barley and bare field. The pasture (c) spectra is not identified in the atrans 2 model image for the original. Both the original and the MSR processed atrans 1 images have classification errors resulting from the affect of the low transmittance area in the lower left portion of the images.

Tables 1 and 2 show the classification sensitivity results for the original and MSR image for different ground truth images. The original and MSR ground truth images refer to the images shown in Figure 10. The sensitivity measurements are listed in terms of the percentage of pixels correctly classified and the actual number of pixels correctly classified out of the total 65536 (256×256) pixels that belong to the classified image. As can be seen in Table 1, the MSR processed image for the atrans1 transmittance profile provided better classification sensitivity. When used with the real *original* ground truth image, the MSR produced sensitivity results slightly better, 2.5%,

than those produced by using the original atrans1 data. However original image for the atrans 1 model. However, when the MSR processed original data was used for training, the classification accuracy for the MSR processed image was substantially better, $\sim$ 15%, than the original. An even greater separation in results is seen between the original and MSR images for the atrans2 transmittance profile. When the MSR processed data is used for training, the MSR classified image is approximately 23% better than the original classification. Even when the original unprocessed data is used for training, the MSR classification is about 10% better.

Table 1. Sensitivity (# pixels out of 65536)

Ground Truth	Image	Model 1	
		% correct	# pixels correct
Real Original	Original	57.56	37722
Real Original	MSR	60.99	39970
Real MSR	MSR	70.63	46288

Table 2. Sensitivity (# pixels out of 65536)

Ground Truth	Image	Model 2	
		% correct	# pixels correct
Real Original	Original	54.32	35599
Real Original	MSR	74.34	48719
Real MSR	MSR	86.86	56924

Figure 11 shows an RGB composite of the quantized images. The original images have been linearly stretched so that subtle differences between them and the MSR image can be compared. Comparing the effects of the atmospheric models on the original and MSR images, we see that the MSR images appear visually consistent between both models which is consistent with the results obtained using the 42 band multispectral image results discussed earlier in this experiment. However, the effects of the transmittance models on the original images is more apparent in these quantized images. The low transmittance areas are sharply contrasted with the high transmittance regions. In the MSR atrans1 image we do observe the same darkening of reflectance in the lower left corner of the image which corresponds to the same low transmittance in the atrans1 model in the same area. We observe the same edge artifacts effects around the borders of regions in the MSR quantized images as seen in the 42 band MSR image. For this experiment, the advantage of generating classification and quantized images simultaneously is that the classified image may provide an indication of how closely the quantized image will match the original data.

In Figure 13 we illustrate the results of using the squared-error difference metric to compare the original and MSR images for each atmospheric transmittance model. We conclude from the difference images that the consistency observation between the MSR images for the difference transmittance models is again confirmed. These results are similar to the results shown in Figure 12. However, we observe more high difference regions in Figure 13 for the original image, than in Figure 12.

6. CONCLUSIONS

Although image enhancement is typically applied to improve the visual quality of multispectral images, in this experiment we have given quantitative evidence that the application of the MSR algorithm restores images that are degraded by atmospheric transmittance effects, and improves the results of multispectral image classification. Because the MSR algorithm was applied before clustering, the classification algorithm generated candidate spectra that were better separated in reflectance for the MSR images than the spectra generated for the original images. Furthermore, the MSR candidate spectra maintained separability and high reflectance values regardless of the atmospheric transmittance models applied. This leads us to conclude that the application of the MSR algorithm produces (approximately) illuminant invariant spectral signature images. Except for class regions in which edge artifacts produced incorrect classifications around region boundaries, the classification results and the difference measurement results show a consistency between MSR images that is not evident in the classifications based on the original images.

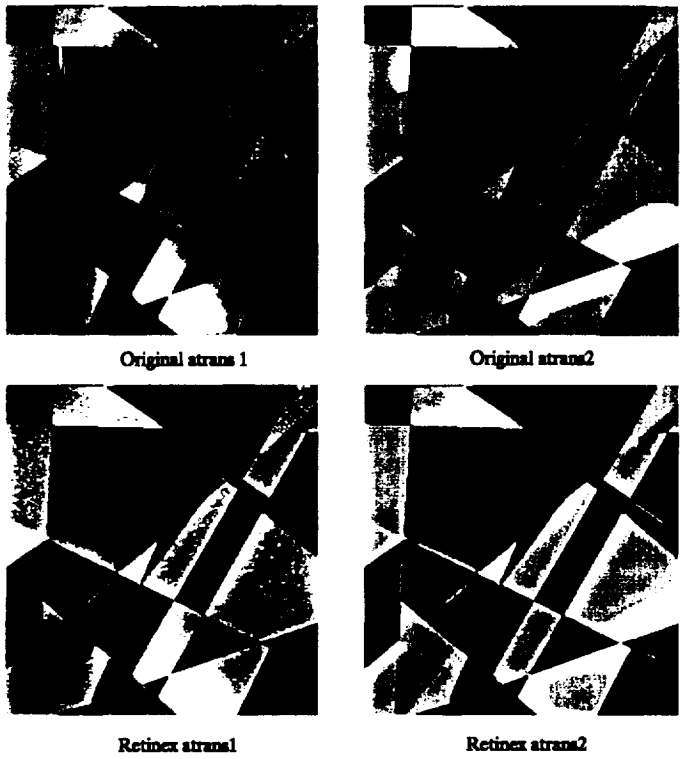


Figure 11. Effect of atmospheric models on VQ original and MSR images.

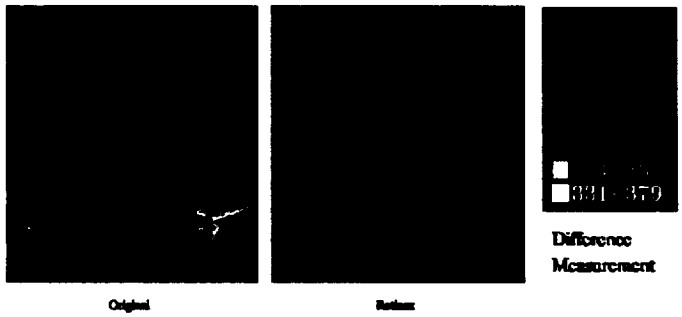


Figure 12. Difference images for the (l) original and (r) MSR images.

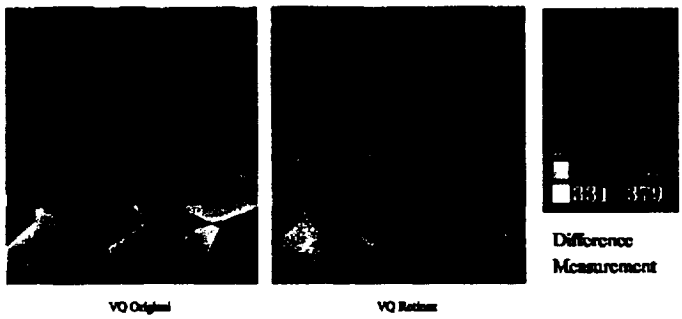


Figure 13. VQ Difference images for the (l) original and (r) MSR images.

ACKNOWLEDGMENTS

B. Thompson's research was done as part of her Ph.D. dissertation at the Department of Computer Science, College of William & Mary. Z. Rahman was supported by the NASA Langley Research Center Cooperative Agreement NCC-1-258. A version of this paper which includes color figures is available at:
`http://dragon.larc.nasa.gov/retinex/background/pubabs/spie4041-2000.html`.

REFERENCES

1. R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, Academic Press, 1997.
2. R. Jackson, P. Slater, and P. J. Pinter, "Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres," *Remote Sensing of Environment* **13**, pp. 187–208, 1983.
3. J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice-Hall, Inc., 1996.
4. E. H. Land, "The retinex theory of color vision," *Scientific American* **237**, pp. 108–129, December 1977.
5. M. Olshaker, *The Instant Image: Edwin Land and the Polaroid Experience*, Stein and Day, Scarborough House, Briarcliff Manor, N.Y., 1978.
6. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multiscale retinex for color image enhancement," in *IEEE International Conference on Image Processing*, vol. 3, pp. 1003–1006, 1996.
7. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing* **6**, pp. 451–462, March 1997.
8. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing* **6**, pp. 965–976, July 1997.
9. F. O. Huck, C. L. Fales, R. E. Davis, and R. Alter-Gartenberg, "Visual communication with retinex coding," *Applied Optics* **39**, April 2000.
10. R. N. Clark, V. King, and G. A. Swayze, "Initial vegetation species and senescence/stress indicator mapping in the San Luis Valley, Colorado using imaging spectrometer data," 1995. `http://speclab.cr.usgs.gov/PAPERS.veg1/vegispc2.html`.
11. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, 3rd ed., April 1992.
12. K. Sayood, *Introduction to Data Compression*, Morgan-Kaufmann, 1995.