Issues in segmenting hyperspectral imagery from histograms

J. Silverman*[a], S.R. Rotman**[b], K.L. Duseau***[c] and C.E. Caefer[c]
[a]Solid State Scientific Corp.
[b]Ben-Gurion Univ. of the Negev
[c]Air Force Research Laboratory, Hanscom AFB;

ABSTRACT

In earlier work, we have shown that starting with the first two or three principal component images, one could form a two or three-dimensional histogram and cluster all pixels on the basis of the proximity to the peaks of the histogram. Here, we discuss two major issues which arise in all classification/segmentation algorithms. The first issue concerns the desired range of segmentation levels. We explore this issue by means of plots of histogram peaks versus the scaling parameter used to map into integer bins. By taking into account the role of $P_{min,}$ the minimum definition of a peak in the histogram, we demonstrate the viability of this approach. The second issue is that of devising a merit function for assessing segmentation quality. Our approach is based on statistical tests used in the Automatic Classification of Time Series (ACTS) algorithm and is shown to support and be consistent with the histogram plots.

Keywords: Hyperspectral, segmentation levels, histograms, principal components, merit function, ACTS algorithm, K-means algorithm

1. INTRODUCTION

At earlier conferences in this series, we introduced a new technique for segmenting hyperspectral imagery.[1,2] Segmentation was based on peaks in 2 and 3 dimensional histograms formed from principal components mapped into integer bins. The algorithm evolved to a semi-automatic form in that the remaining user choice was the desired number of digital levels in the segmentation and the minimum number of pixels present at a maximum value to declare a histogram peak.

Two important issues remain unresolved, which form the focus of the present paper. The first asks whether a fixed number (or narrow range) of segmentation levels is "recommended" by and extractable from the data cube itself. A widely-used algorithm for segmentation/classification, the K-means algorithm[3], requires the user to specify K pixels as initial seed centers. A more recent algorithm, Automatic Classification of Time Series (ACTS) [4], features as its main advantage the automatic extraction of unique clusters, i.e. segmentation levels, as inherent to the algorithm. A second and closely related issue is critical in comparing algorithms designed for the same purpose, in this case segmentation. Namely, in the absence of ground truth and wishing to avoid a subjective user evaluation, can one design a merit function to assess the quality of a segmentation? More specifically, given two segmentations from different algorithms at similar number of levels, or segmentations at different levels from the same algorithm, we seek a useful metric to favor one segmentation over another. A metric which is useful in the second case (different levels from the same technique) would provide a link between our two issues.

The main thrust of the present paper is to describe our initial efforts to address these two issues. Our presentation is organized as follows: Section 2 reviews our segmentation algorithm; Section 3 describes our approach to the two issues; Section 4 presents results on a MWIR and SWIR data cube; and Section 5 is a summary and pointer to future work.

*jerry.silverman@hanscom.af.mil; phone 781 377-3295; fax 781 377-4814; Solid State Scientific Corp., 27-2 Wright Rd., Hollis NH 03049, **srotman@ee.bgu.ac.il; phone +972-8-6413539; fax +972-8-6472949; Ben-Gurion University, Dept. of Elec. Eng., P.O. Box 653 Beer-Sheva, 84105 ISRAEL. ***karen.duseau@hanscom.af.mil; phone 781 377-4047; fax 781 377-4814; Sensors Directorate, Air Force Research Laboratory, AFRL/SNHI, 80 Scott Rd., Hanscom AFB, MA USA 01731;

2. SEGMENTATION ALGORITHM

The process of spectrally clustering pixels, which forms the basis for classifying hyperspectral data[5], is closely linked to segmenting the image because of the spectral/spatial correlation of natural scenes. On the other hand, a powerful classical approach to segmenting single gray-scale images is by means of the extrema of their histograms[6]. In effect, our segmentations techniques[1,2] combine the above approaches. We form the histograms from the principal components (PC) of the hyperspectral datacube. (We will describe the method for two components, although the technique is easily extended to 3 or more dimensions.) A two-dimensional histogram is produced from typically the first two components. Peaks (local maxima) in the histogram are located and labeled from 1 to N. The histogram space is then templated to N levels by calculating the distance of any point in the space from each peak: that point is then labeled with the label of the "closest" peak. Each pixel in the data cube is then assigned the digital label of its corresponding point in the two-dimensional histogram space and segmentation is accomplished.

Two refinements on the above process were described at last years' conference[2] :

1. The "closest peak" in the above scheme is no longer taken as Euclidian but the peak influences in each dimension are modeled by Gaussian parameters estimated from the co-occurrence histograms of the individual PC's. In effect, this weights the influence of each component separately.
2. In order to generate our histograms, we need to map the PC floating-point values into integer bins. The number of segmented levels (up to a point) increases with the chosen number of integer bins leading from coarse to finer segmentations. We developed an interactive form of the algorithm where the user specifies a desired number of segmentation levels and an entropy measure is used to guide a non-linear histogram mapping[7] into that final set of integer bins that achieves roughly the desired number of levels.

Representative results are shown in Figure 1 for a MWIR hyperspectral image taken with an in-house imaging spectrometer. The imager used is the next generation of the prototype described in earlier work[8]. We show segmentations to 7, 10, 15 and 19 levels respectively, all based on the first two principal components. A key specified parameter in the histogram technique is the minimum of pixels at a maximum location accepted as a peak, P_{min}. In the results of Fig. 1, P_{min} is set at 2. The increasing number of segmentation levels is largely due to a division of building pixels into smaller segments. Can we deem one of these similar-looking segmentations as better or more representative than the others?

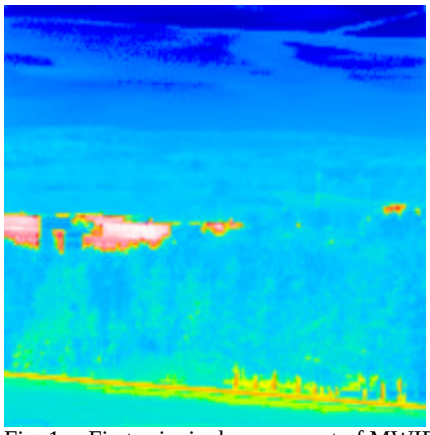

Fig. 1a. First principal component of MWIR data cube.

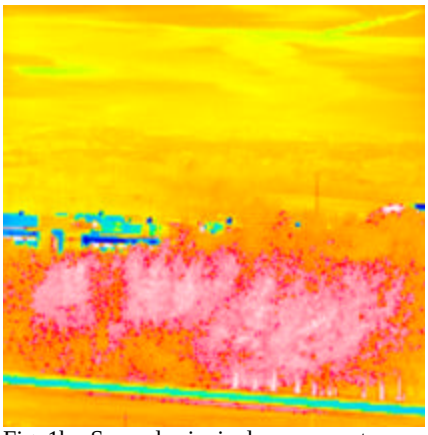

Fig. 1b. Second principal component.

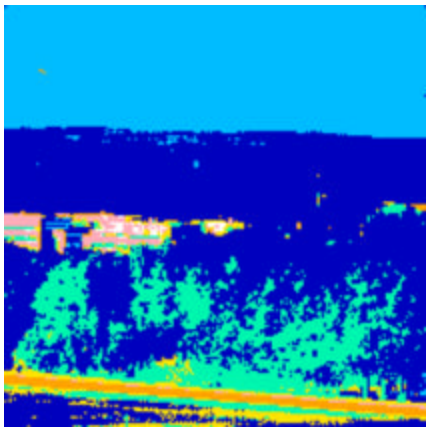

Fig. 1c. Segmentation to 7 levels.

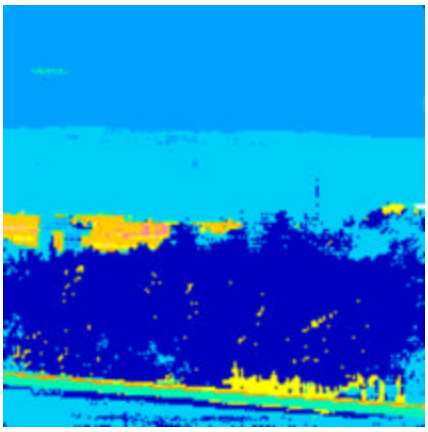

Fig. 1d. Segmentation to 10 levels.

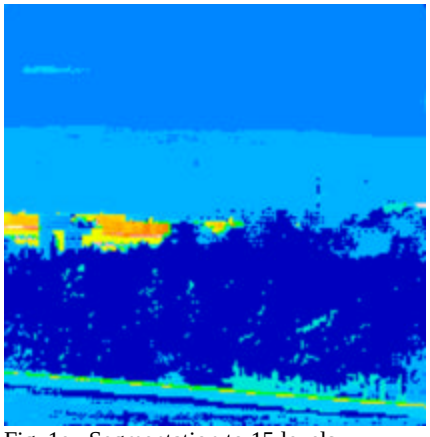

Fig. 1e. Segmentation to 15 levels.

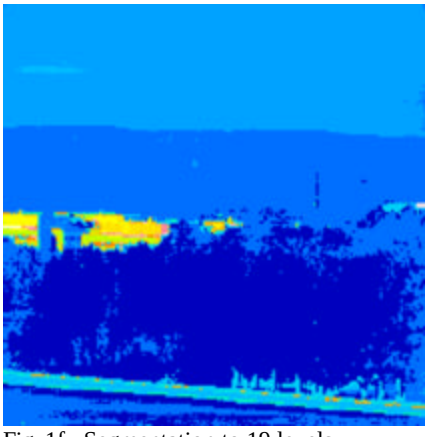

Fig. 1f. Segmentation to 19 levels.

3. APPROACH

While the scaling into integer bins in principle provides a gradation of segmentations from coarse to fine, alternatively one could ask whether there is a fixed number (or narrow range) of distinct clusters (segmentation levels) inherent in the data. Plots of the number of peaks in the histograms that generate our segmentations versus scale (number of integer bins) is our trial approach to explore this issue. As a guide to interpreting these plots on full data cubes with many border regions, it is useful to start with simplified versions of the full data cubes composed of regions selected for their relative homogeneity. Two such simplifications are shown in Figure 2. The first simplified data cube (Fig. 2a) of 80 by 120 pixels consists of four 80 by 30 pixel rectangular regions from the MWIR (74 bands) data cube of Fig. 1 selected respectively from the sky; the distant foliage; the closer tree foliage; and finally (as the bottom rectangle) a region of ground, road and tree. There are five major clusters here and the plots of histogram peak number versus scale levels (Fig. 2c) does indeed level off at 5 starting with P_{min} of 6-10. At lower P_{min}, smaller segments split off from border areas or from the tree foliage which has considerable texture.

We increase the realism of our example in the 120 by 120 image (Fig. 2b) which has appended to the Fig. 2a image a 40 by 120 pixel swath centered on the building including additional sky, distant foliage, and tree areas. We have added one or more major clusters, depending on P_{min}, and the corresponding plot (Fig. 2d) shows level regions between 6-8. We conclude from these semi-simulations that this plotting approach is very promising but that the roles of P_{min} and border pixels needs to be taken into account. The results on the full original data cubes are treated in the next section.

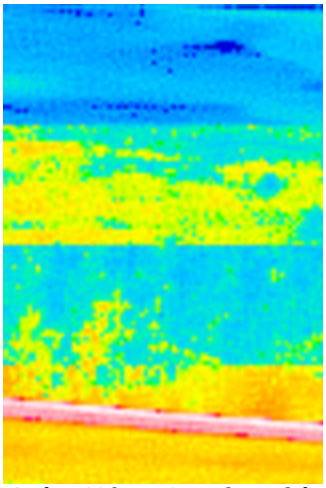

Fig. 2a. Band 16 of an 80 by 120 pixel simplified version of the Figure 1 image.

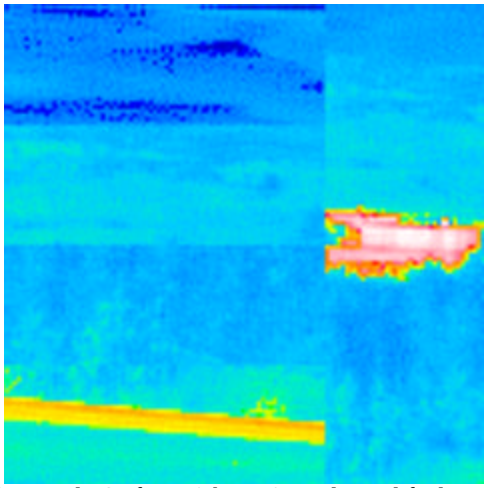

Fig. 2b. Band 16 of a 120 by 120 pixel simplified version of the Figure 1 image.

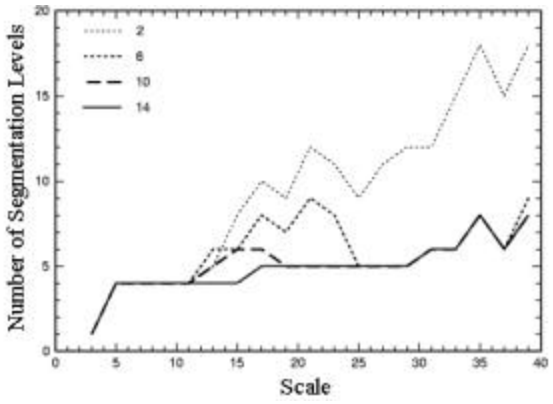

Fig. 2c. Number of histogram peaks or segmentation levels versus scale at 4 values of P_{min} for the image in 2a.

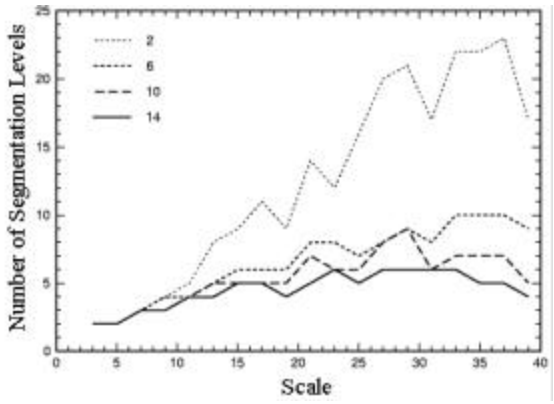

Fig. 2d. Corresponding plots for the image in 2b.

Our approach to the second issue, namely a merit function to assess segmentation quality, is based in part on the ACTS algorithm[4]. The ACTS method works in the context of a descending tree structure in which each parent cluster is potentially spilt into two child clusters based on the scatter in the parent cluster. After iterating this split to a specified degree of stability, one applies two statistical tests to the child clusters to ratify or reverse the potential split. The two tests, which form the basis of our merit function, are as follows.

In test 1, the centers of each child cluster, X_1 and X_2, are computed as average profiles of the pixels assigned to each cluster and X_P, the average profile of the parent cluster, is also computed. One can then compute F, the fraction of the pixels in the total whose Euclidian distance is closer to X_P than to its nominal child cluster X_1 or X_2. Lower values of this fraction favor the split.

In test 2, the standard deviations σ_1 and σ_2 of each child cluster, (summed over all the bands), as well as the parent cluster σ_P, are computed and the ratio σ_r of $(\sigma_1 + \sigma_2)/ \sigma_P$ is computed. Again lower values favor the split. In the implementation of the ACTS algorithm, upper bounds are set for these test values for accepting or negating a potential split. We propose to apply these two tests to all pairs of clusters (segments) in a final segmentation, generated by whatever algorithm, as the basis of a possible merit function. Specifically, for each of the N (N-1)/2 pairs of an N-level segmentation, we compute F and σ_r and then compute the averages and the maximum over the pairs: F_{aver}, F_{max}, σ_{aver} and σ_{max} as possible metrics.

To digress slightly, we should mention that the pair of tests used in the ACTS algorithm does have a built-in bias against pairs with a large disparity in size. If say $N_1 >> N_2$, then $X_P \approx X_1$ and high values of F and σ result. One can remove such a bias by weighting each pixel in the smaller N_2 cluster by $N_1 \div N_2$. In effect, this treats each child cluster as a probability distribution stemming from equal numbers of pixels. A merit function based on such a bias removal has been tested and gives much less useful results than the results based on the tests in the form used in the ACTS algorithm. Hence, we present only the latter results.

For simplicity, we show the average metric values, F_{aver} and σ_{aver} hereafter called F and σ; the corresponding maximum metrics generally track these averages. In Figure 3, we plot the average metric values at $P_{min} = 2$ and 14 for the image in Fig. 2a (Figs. 3a and 3b) and for the image in Fig. 2b (Figs. 3c and 3d). The upper dotted curves are the average σ metric and the lower solid curves are the fractional F metric. The dark lines in Figure 3 are the corresponding plots at $P_{min} = 2$ and 14 from Figure 2 of the number of segmentation levels. The results show promise for these simplified data cubes. We find a minimum in the merit values at 5 segmentation levels and minor oscillations thereafter (Fig. 3a) and a flat broad minimum in both metrics corresponding to the segmentation plot (Fig. 3b). The picture is less clear in Figs. 3c and d. The latter ($P_{min} = 14$) has metric minima at five and six clusters and generally mirrors the segmentation number plot but the former ($P_{min} = 2$) has a gradually decreasing σ metric and a very flat F metric. Our proposed merit function looks promising, but again careful interpretation as a function of P_{min} will be needed.

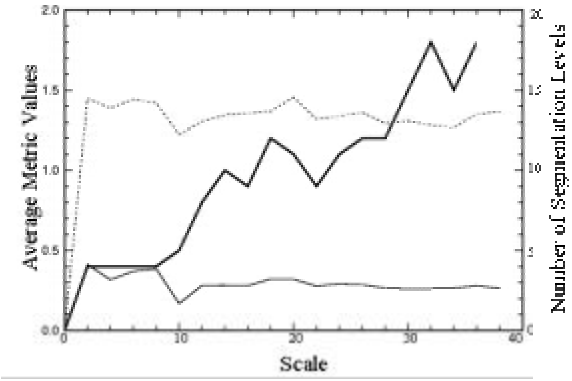

Fig. 3a. Metric plots of Fig. 2a versus increasing scale at P_{min} = 2. Dark curve is number of segmentations.

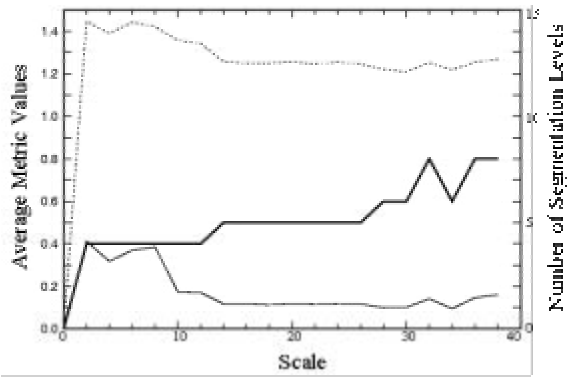

Fig. 3b. Corresponding metric plots of Fig. 2a at P_{min} = 14. Dark curve is number of segmentations.

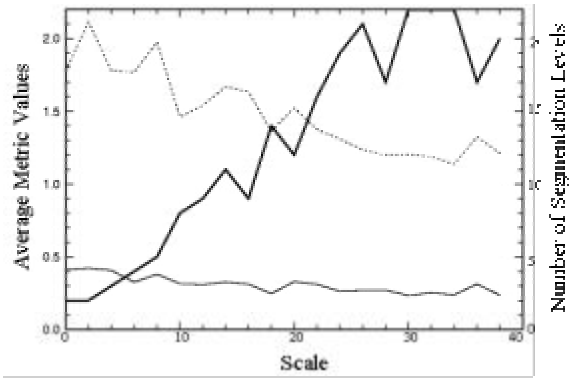

Fig. 3c. Metric plots of Fig. 2b versus increasing scale at P_{min} = 2. Dark curve is number of segmentations

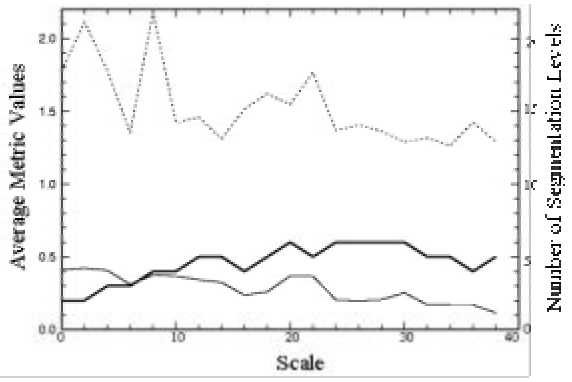

Fig. 3d. Corresponding metric plots of Fig. 2b at P_{min} = 14. Dark curve is number of segmentations.

4. RESULTS

The two kinds of plots which address the two main issues of this paper, histogram peaks versus scale (segmentation number) and metric function versus scale, are plotted in Figure 4 for the full data cube of the Figure 1 MWIR image. If one compares Figs. 4a with 2d, one notes that the second simplified image is quite similar to the full cube in the leveling off at high P_{min} indicating 6-8 major clusters. Fig. 4b is the modified plot when a spectral edge detector is employed to identify the top 9.5 % edge or border pixels and eliminate them from the histogram; one sees that the segments in the P_{min} = 2-6 segmentations contain border pixels but the large scale segments show little influence from their presence. The general appearance of the metric plots of the full MWIR data cube (Figs. 4c and 4d) is similar to but smoother than those of Figs. 3c and 3d. For P_{min} = 2, the metrics reach low values at about 7-8 clusters and thereafter oscillate for the F metric and decrease gradually for the σ metric. The plots in Fig. 4d mimic the saturation at 5-6 segmentation levels of the P_{min} = 14 plot.

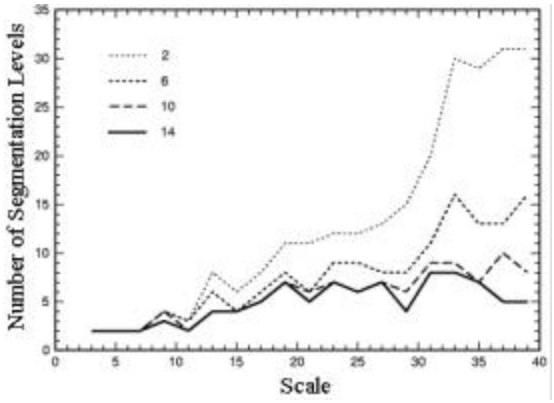

Fig. 4a. Number of histogram peaks or segmentation levels versus scale at 4 values of P_{min} of Fig. 1.

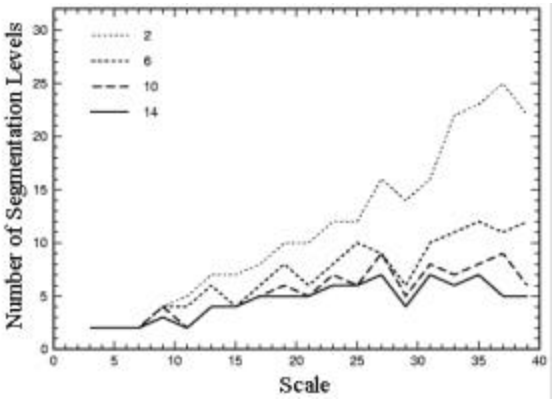

Fig. 4b. Plot of Fig. 4a with 9.5% border pixels removed.

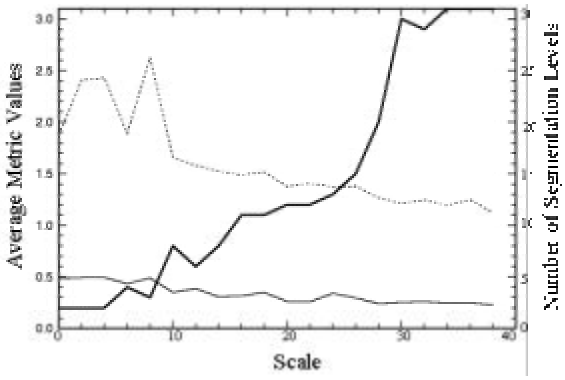

Fig. 4c. Metric plots of Fig. 1 at P_{min} = 2.
Dark curve is number of segmentations.

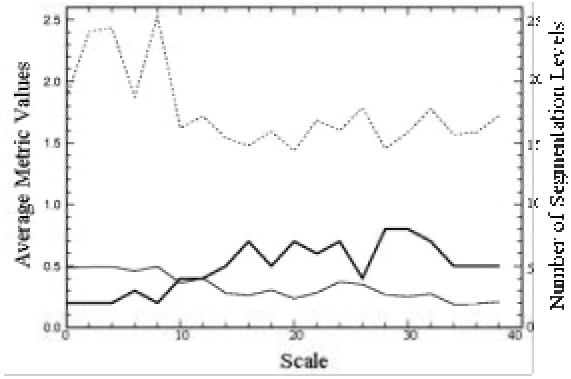

Fig. 4d. Metric plots of Fig. 1 at P_{min} = 14.
Dark curve is number of segmentations .

Turning to the SWIR HYDICE image[9] in Figure 5, we see that the histogram plot indicates 7-8 clusters and the metric plots again at P_{min} = 2 and 14 are quite similar in character to those in the MWIR image (Figs. 4c and 4d) and are consistent in preferring 7-8 levels (although the trend in the σ metric at P_{min} = 2 of gradual decrease with increasing segmentation levels is present).

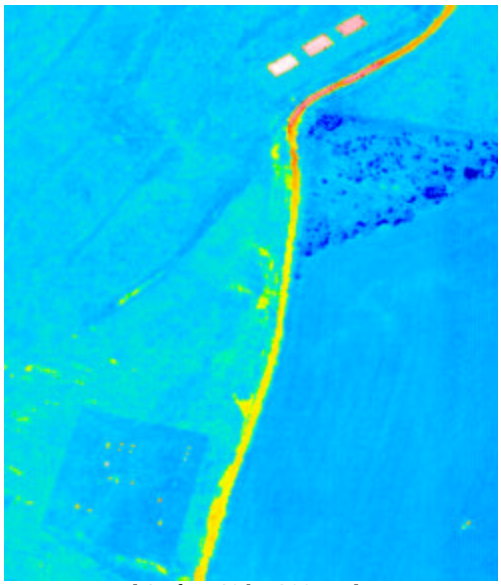

Fig. 5a. Band 8 of a 260 by 300 Hydice image.

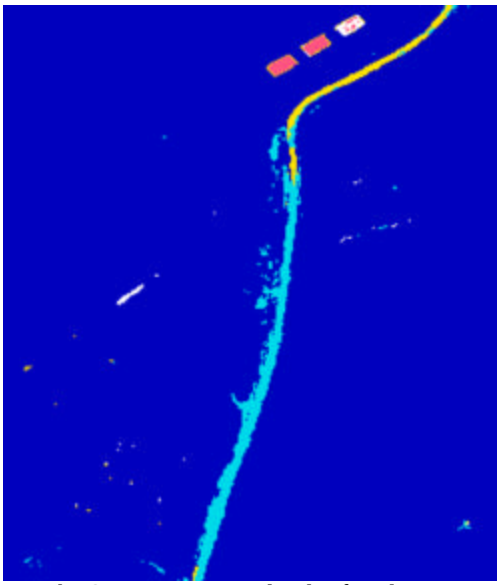

Fig. 5b. Segmentation to 5 levels of Hydice image

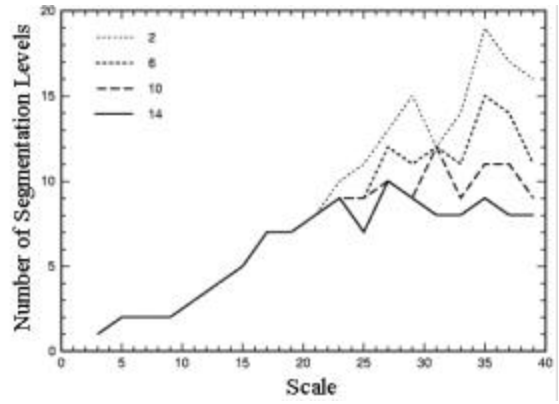

Fig. 5c. Number of histogram peaks or segmentation levels versus scale at 4 values of P_{min} for the Hydice image.

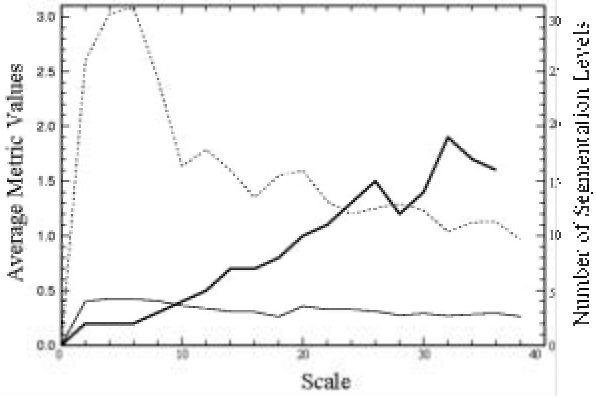

Fig. 5d. Metric plots for the Hydice image at P_{min} = 2. Dark curve is number of segmentations.

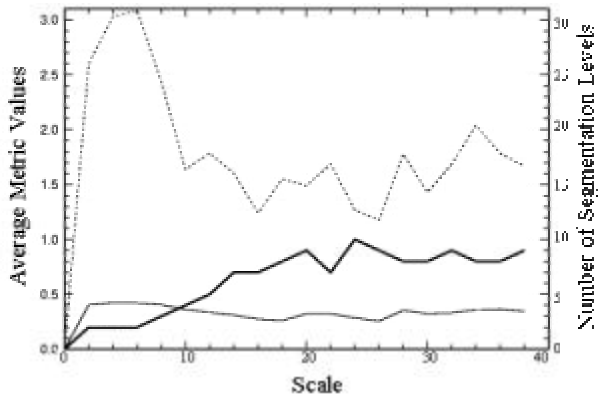

Fig. 5e. Metric plots for the Hydice image at P_{min} = 14. Dark curve is number of segmentations.

A second potential use of our merit function is to compare segmentations at the same number of levels of different algorithms. In Figure 6, our histogram-based segmentations are compared to k-means segmentations at 7 levels for the MWIR image (6a and 6b) and at 6 levels for a SWIR hydice image (6c and 6d) similar to that used earlier. The K starting seeds are from a randomly chosen pixel in each segment of the histogram-based segmentation. The final results (6b and 6d) are after 15 iterations of the k-means process[3]. Given that the histogram-based results are here based on two PC values per pixel, while the k-means results use full band profiles per pixel (74 bands for the MWIR and 210 for the SWIR), and further given that the k-means algorithm iterates an improved Euclidean compactness for each cluster, one would expect lower (improved) metric values for the k-means. Indeed this is the case for both images and both the F and σ measures.

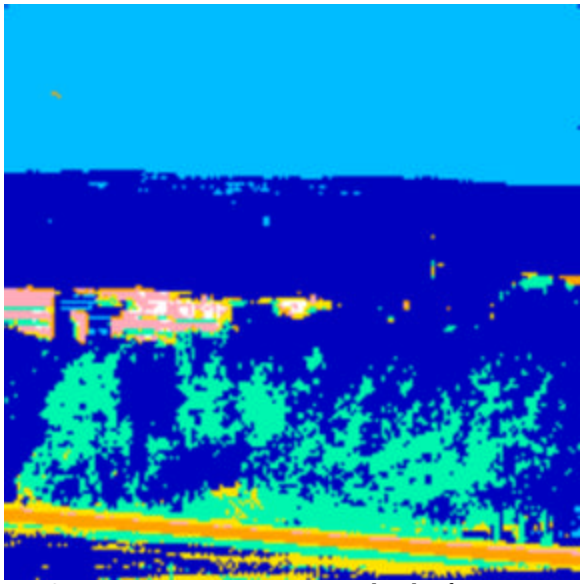

Fig. 6a. Histogram segmentation to 7 levels of MWIR image. Average metrics are: F = 0.27 σ = 1.73

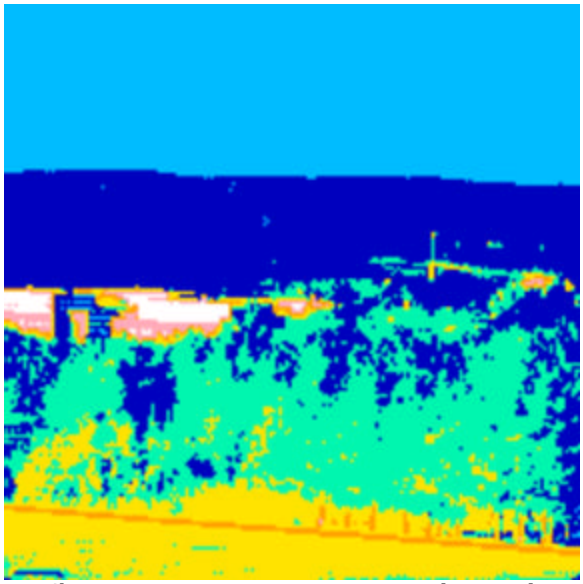

Fig. 6b. K-means segmentation using a random pixel in each segment as starting center. F = 0.21 σ = 1.22

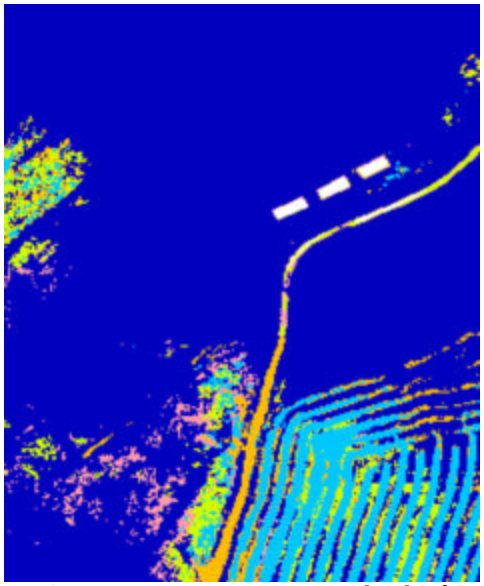

Fig. 6c. Histogram segmentation to 6 levels of SWIR image. Average metrics are: F = 0.34 σ = 1.60

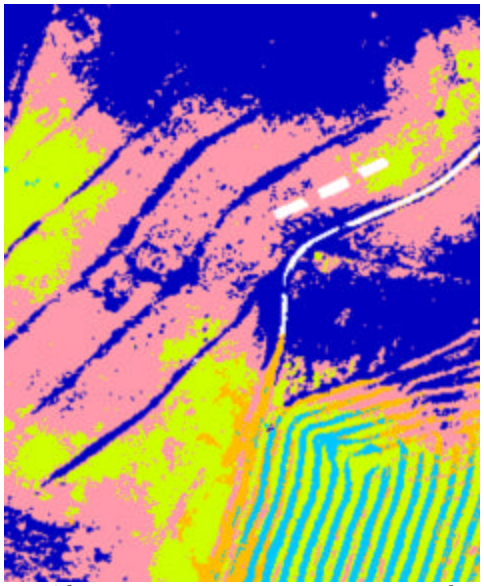

Fig. 6d. . K-means segmentation using a random pixel in each segment as starting center. F = 0.26 σ = 1.46

5. CONCLUSION

We have described our initial approaches to and evaluations of two major issues in classification/ segmentation algorithms. The first issue concerns the desired number of segmentation levels which is explored by means of plots of histogram peaks versus the scaling parameter used to map into integer bins. By taking into account the role of $P_{min,}$ the minimum definition of a peak in the histogram, we have demonstrated the viability of this approach as seen by the roll-off in the number of segmentation levels versus scale plots with P_{min} as a parameter. The second issue is that of devising a merit function for assessing segmentation quality. Our approach is based on statistical tests used in the ACTS[4] algorithm. The two tests compare a parent cluster of pixels to a pair of child clusters formed from the parent as to the fraction closer to the parent cluster center (F test) and a ratio of standard deviations of the child clusters divided by the standard deviations of the parent cluster (σ test). Our candidate merit function treats each pair of segments in a given segmentation as a two-child/parent group and computes the test values. The final metrics are the average F and σ over all pairs.

Although the merit function looks promising, several problems emerge. The σ metric tends to fall gradually at larger numbers of segmentation levels (see Figs. 3c, 4c, 5d) and the F metric is generally rather flat after some fixed number of levels. We have also compared two segmentations at the same number of levels with the merit function (Fig. 6) with reasonable results.

Further efforts to test and possibly modify the merit function will focus on segmenting hyperspectral imagery having ground truth available and will compare histogram-based segmentations with those from the k-means, ACTS algorithm, and a variation on the ACTS algorithm[10].

ACKNOWLEDGMENTS

The camera used to collect the MWIR data were designed and fabricated by William Ewing, Toby Reeves and Steven DiSalvo of our laboratory. The HYDICE data was provided by the Spectral Information Technology Application Center (SITAC). This work was carried out under Air Force Task 2305BN00.

REFERENCES

1. J. Silverman, C. E. Caefer, J.M. Mooney, M.M. Weeks, and P. Yip, "An automated clustering/segmentation of hyperspectral images based on histogram thresholding", in *Imaging Spectrometry VII*, Michael R. Descour, Sylvia S. Shen, Editors, Proceedings of SPIE Vol. 4480, 65-75 (2002).

2. J. Silverman, S.R. Rotman and C.E. Caefer, "Segmentation of Hyperspectral Images from the Histograms of Principal Components", *in Imaging Spectrometry VIII*, Sylvia S. Shen, Editor, Proceedings of SPIE Vol. 4816, 270-277 (2002).

3. R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, Chap. 9, Academic Press, San Diego (1997).

4. N. Viovy, "Automatic Classification of Time Series (ACTS): a new clustering method for remote sensing time series", *Int. J. Remote Sensing* **21**, pp. 1537-1560 (2000).

5. J.A. Richards and J. Xiuping, *Remote Sensing Digital Image Analysis: An Introduction*, Chap. 9, Springer-Verlag, Berlin, 1999.

6. F. M. Wahl, *Digital Image Signal Processing*, Chap. 5, Artech House, Boston, 1987.

7. V.E. Vickers, “Plateau equalization algorithm for real-time display of high-quality infrared imagery”, *Opt. Eng.* **35(7)**, pp. 1921-1926 (1996).

8. J. M. Mooney, V. E. Vickers, M. An, and A.K. Brodzik, "High throughput hyperspectral infrared camera", *J. Opt. Soc. Am. A,* **14(11)**, pp. 2951-2961 (1997).

9. The HYDICE imagery, taken over the CART/ARM Site Lamont, was provided by the Spectral Information Technology Application Center (SITAC). HYDICE is a push broom imaging spectrometer with 210 spectral bands over the 0.4 to 2.5 micron spectral range.

10. K. Huang, “The use of a newly developed algorithm of divisive hierarchical clustering for remote sensing image analysis”, *Int. J. Remote Sensing* **23**, pp. 3149-3168 (2002).