P4.7 IMAGE PROCESSING OF METEOROLOGICAL RADAR DATA USING A COHERENT CLUSTERING TECHNIQUE

James Dunyak, Xiaoning Gilliam*, Arthur Doggett
Wind Engineering Research Center, Texas Tech University

Sunanda Mitra
Department of Electrical Engineering, Texas Tech University,

1. INTRODUCTION

Interpretation of meteorological radar data is often done in a subjective, qualitative manner. While features of interest can be identified and tracked in this manner, this method does not lend itself to automation or mass processing of radar data. Use of centroid tracking is one method to approach this problem, but it is limited to convective cores. We believe that more robust clustering techniques can be developed to allow a better identification of storm features. Some applications of these techniques include: identification and study of features other than just convective cores, development of object-oriented codes to build databases from large amounts of radar data, and automation to provide operational meteorologists with more comprehensive (and easily accessible) information (see [6, 9, 10]).

In this paper, we first introduce a non-convex optimization clustering technique for image processing: the deterministic annealing method [5, 8]. This method works on image segmentation based on the principle of maximum entropy, which is independent of the initial choice of configuration and without a priori knowledge of the image. However, clustering for segmentation of complex images is often problematic: the clustering methods always segment the image, but they often replace a complex image with an even more complex set of complicated segments. This is particular true with meteorological radar data. To overcome this problem, we replace the image segmentation with a rigorous statistical test. This new approach follows our earlier work on coherent structure detector [1, 2, 3, 4]. The method is based on a definition of an incoherent image and controls the test size without assumptions about the process. We apply this concept of coherency to multidimensional images for the detection and characterization of coherent image segments. Our recent analysis established the test size and provides an algorithm for calculating p-values.

* *Corresponding author address:* Xiaoning Gilliam, Dept. Civil Engineering, Texas Tech University, Lubbock, TX 79409, e-mail kgilliam@texas.math.ttu.edu

2. METHODOLOGY

Clustering using Deterministic Annealing (DA) approach ([5, 8]) focuses on defining a cost criterion to be minimized. In our case, we minimize the expected cost D under constrained "randomness" expressed by the Shannon entropy H:

$$D = \sum_x p(x) \sum_y p(y/x) d(x, y),$$

where x is the input source vector; y is the centroid vector associated with x; $p(x, y)$ is the joint probability distribution; $p(y/x)$ is the conditional probability; $d(x, y)$ is the Euclidean distance, and

$$H = -\sum_x \sum_y p(x, y) logp(x, y).$$

The minimization can be achieved by minimizing the Lagrangian $F = D - TH$, where T is the Lagrange multiplier used as a pseudo-temperature parameter in the annealing scheme. Simple mathematical analysis leads to a fixed point iterative algorithm:

$$y^{(n+1)} = f(y^{(n)}),$$

where f is define by

$$f(y) = \frac{\sum_x \left(\frac{x e^{(-\beta |x - y^{(n)}|^2)}}{\sum_y e^{(-\beta |x - y^{(n)}|^2)}} \right)}{\sum_x \left(\frac{e^{(-\beta |x - y^{(n)}|^2)}}{\sum_y e^{(-\beta |x - y^{(n)}|^2)}} \right)},$$

and where $\beta = 1/T$.

We have developed a localized DA which scales linearly with area of image since our bigger images cover more area instead of increasing in resolution. Once the suitable parameters such as the pseudo-temperature parameter β, the number of clusters, and the number of iterations have been chosen, we obtain the segmented radar image. Our next step is to identify the coherent clusters from the segmented radar image.

Unfortunately, the statistical relevance of any particular segment is not always clear. Analysis and interpretation in the presence of white noise is straightforward, but analysis of clusters in the presence of spatially correlated noise is much more complex. Thus, we develop a rigorous statistical test to detect and characterize the coherent cluster from incoherent (but spatially correlated) noise. This test, the coherent cluster detector (CCD), is based on one of the oldest methods in nonparametric statistics: the development of a randomized reference distribution. Based on the definition of an incoherent image, no explicit knowledge is needed of the spatial noise spectrum or amplitude distribution. In addition, no explicit knowledge is needed of the cluster shape or amplitude. In our case, we want to distinguish coherent clusters from incoherent spatially correlated noise in the fine structure of high-resolution images due to clouds and other concentrated scatters (insects, birds and dust), as well as inversion layers, and air density variations.

To build our randomized reference distribution, we build a large number of exemplars through random rearrangement of the original segmented radar image. The exemplars are built by applying a multidimensional Fourier transform, randomizing the phases, and transforming back. If the image is spatially correlated noise, then the exemplars have the same distribution as the original image. These exemplars are then clustered.

At each exemplar cluster, order statistics are accumulated from the group of n maximum values of the exemplars x_i^e. These order statistics are then used to provide thresholds and corresponding p-values [7] for comparison with x_i, the original image at each clusters. For a specific level of significance α (size of the test) and the number of clusters N, we use the indicator function $I_{(.)}$ to estimate the threshold

$$t_{CCD}(N, n, \alpha) = \left\{ \sup\ t : \left(\frac{1}{n} \sum_{i=1}^{n} I_{x_i^e > t} \right) > \alpha \right\}.$$

At each cluster in the original image, we use these thresholds to determine the associated p-value. Only those clusters that have sufficiently small p-value may be viewed as statistically relevant. Through this threshold technique, we can decide if the original cluster is statistically relevant. The coherent cluster detector controls the false alarm γ per cluster and scales linearly with image area, since controlling false alarm per image is impractical and inappropriate for large area images. In our statistical test, if the cluster is incoherent, the exemplars and the original clustered image have the same distribution. If, on the other hand, the cluster is coherent, energy from the coherent cluster component is spread through the entire image when random rearrangements of radar image were used. The threshold used by the coherent cluster detector is not greatly affected by the presence of the statistically relevant coherent structure at the cluster. This results in a high probability of detection.

3. EXAMPLES

To illustrate the coherent clustering technique, we first consider the image of a known coherent signal embedded in colored spatial noise. The noise component was obtained by passing Gaussian white noise through a spatial linear. Figure 1(a) shows the original image as well as each location of the clusters (indicated in $\otimes$), when $\beta = 39$. Figure 1(b) indicates the identification of each clusters. Figures 1(c) and 1(d) show the image for an exemplar and the identification of each cluster in the exemplar. Note that the coherent signal is spread through the entire image. Using our coherent clusters detector, the resulting p-value map is obtained in Figure 1 (e). We are able to recover the coherent signal from the image. This technique works well for moderate to low noise levels, but higher noise levels require noise removal techniques.

Now we apply our coherent clustering technique to the radar image. The image used in this study was collected from WSR-88D Doppler radar. Figure

2 (a) shows radar reflectivity data during a thunderstorm outflow. Notice the strong reflectivity at the location of $x_{axis} = 0.25$ and $y_{axis} = 0.5$. The p-value map for the radar image is given in Figure 2 (b), indicates the statistically relevant information (coherent clusters) in the image data. The performance of the statistical test can be evaluated by a comparison study of the coherent cluster detector and the traditional optimal detector for moderately high signal to noise ratio with known signal and noise setting. Noise characteristics, the image segment properties, and the clustering algorithm will determine this performance.

Our aim is to detect and characterize coherent structures in the image in time- evolving settings. Sequences of images may be used for the detection and prediction of the storm event using the p-value map. Another viewpoint of incoherency may also be considered based on Poisson noise that follows from ground clutter, aircraft, birds, insects, and precipitation events. The advantage of the coherent clustering techniques for image data fusion is its nonparametric nature. It requires no explicit knowledge of the spatial noise spectrum or the Poisson amplitude distribution.

ACKNOWLEDGMENTS

This work was performed under the Department of Commerce National Institute of Standards and Technology/Texas Tech University Cooperative Agreement Award 70NAB8H0059.

References

[1] Dunyak, J., Gilliam, X., Doggett, A., and Smith, D.: under review, A statistical technique for identifying intermittent turbulence in the atmospheric boundary layer.

[2] Dunyak, J., Gilliam, X., Peterson, R., and Smith, D.: 1998a, Coherent gust detection by wavelet transform, *J. Wind Engineering and Industrial Aerodynamics*, **77-78**, 467-478.

[3] Dunyak, J., Gilliam, X., Peterson, R., and Smith, D.: 1998, Detection of coherent structures in long time series, *Proceedings of the 1998 Conference on the Mathematical Theory of Networks and Systems*, Padova, Italy, July.

[4] Gilliam, X.: 1998, Statistical techniques for identification of coherent structures by wavelet analysis, *Ph.D. Dissertation.*

[5] Joshi, S.: 1999, *Deterministic annealing for efficient clustering and vector quantization*, Master Thesis, Texas Tech University, Lubbock, Texas.

[6] Moller, A. R., C. A. Doswell III, M. P. Foster, and G. R. Woodall, 1994: The operational recognition of supercell thunderstorm environments and storm structures. *Wea. Forecasting*, **9**, 327-347.

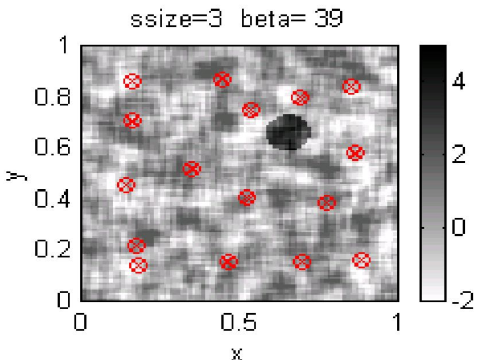
[7] Ott, R.: 1993, *An introduction to statistical methods and data analysis*, Duxbury Press, Belmont, California.

[8] Ross, K.: 1998, Deterministic annealing for clustering, compression, classification, regression, and related optimization problems *Proceedings of the IEEE*, **86**, 11, 2210-2239.

[9] Stumpf, G. J., Witt, A., E. D. Mitchell, P. L. Spencer, E. D. Mitchell, J. T. Johnson, M. Eilts, K. Thomas, and D. Burgess, 1998: The National Severe Storms Laboratory Mesocyclone Detection Algorithm for the WR-88D. *Wea. Forecasting*, **13**, 304-326.

[10] Witt, A., M. Eilts, G. J. Stumpf, E. D. Mitchell, J. T. Johnson, and K. Thomas, 1998: Evaluating the performance of the WSR-88D severe storm detection algorithm. *Wea. Forecasting*, **13**, 513-518.

Figure 1 (a) Signal + Noise Image With Clusters, (b) Identification of the Clusters



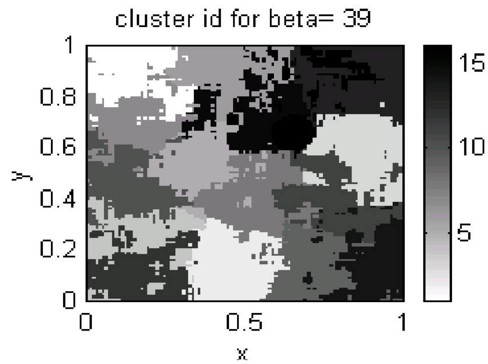
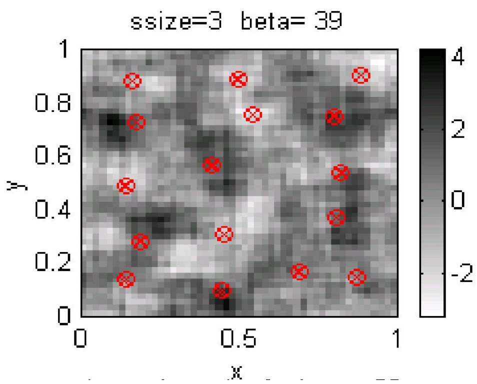


Figure 1 (c) Exemplar - Signal + Noise Image With Clusters, (d) Exemplar -Identification of the Clusters



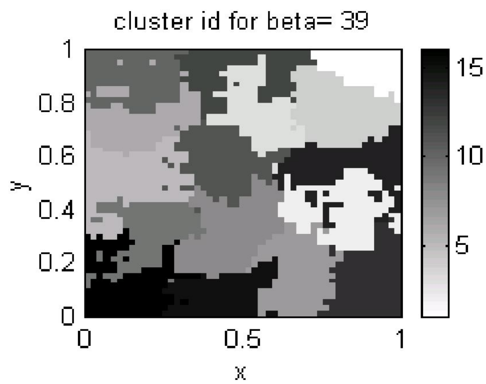


Figure 1 (e) Revovered Signal (p-value map)

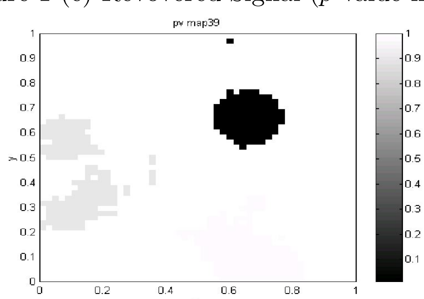
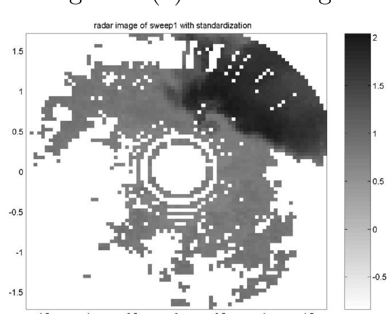


Figure 2 (a) Radar Image



(b) P-value maps (Coherence Clusters)