Consistent Clustering of Radar Reflectivities using Strong-Point Analysis: A Prelude to Storm Tracking

B Root*, **T-Y You, M Yeary**
Atmospheric Radar Research Center & University of Oklahoma
Norman, Oklahoma 73072

1. Introduction

Tracking multiple objects over a time series requires that the objects are consistent and smoothly varying in position. Drastic changes to the position at each time-step hinders the tracking algorithm. For radar images, these objects are features that are identified by an image processing clustering algorithm.

Therefore, a clustering algorithm must produce stable and consistent results in order to maximize the skill of the tracking algorithm. If the clustering algorithm output does not change continuously with changes in the input images, then any object-based tracking algorithms will be severely hampered.

Similarly, another important feature of image processing clustering algorithms is the relationship between the output clusters and the input parameters. For operational usability, the clustering algorithm should have an obvious relationship between the input parameters and the output results. In other words, it is undesirable for a clustering algorithm to have dramatically varying results with small changes to input parameters. This property is the "tunability" of the algorithm.

This work presents a threshold-based image processing clustering algorithm called Strong Point Analysis (SPA). SPA was developed with the goal of achieving the stated desired properties. SPA uses a statistically-driven approach for determining thresholds, as well as a feature to dynamically re-evaluate the thresholds in sub-regions of the image being processed.

The goal of the algorithm is to consistently and reliably identify the salient features of an image. This project demonstrates the properties of SPA on radar reflectivity images from several Weather Surveillance Radar 1988 Doppler (WSR-88D) sites. The radar images sample a wide variety of weather events, so as to determine independence of SPA from shape and size biases. A few of these images are showcased in this manuscript.

*ben.root@ou.edu

2. Methods

At the most fundamental level, the algorithm of SPA can be reduced down to the following description:

1. Identify and network together neighboring pixels that are "Strong Points". An isolated network is a cluster.
2. Grow each cluster's region with neighboring pixels that are "Weak Points."
3. Repeat within each cluster.

This is the engine of the SPA algorithm. A definition of how a pixel can be a "strong point" or a "weak point" is needed for this engine. Also, a condition is needed to end the recursion stated in step three. These definitions can be implementation-specific in order to gain certain properties, or to tailor to a particular dataset. The following definitions are largely heuristical in order to accomplish the stated goals.

A "strong point" is defined to be a pixel that has a value, z, that is greater than or equal to the mean value of the image, μ, plus the quantity of the "upper sensitivity level", U, times the standard deviation of the image, σ. This threshold is set to be no larger than the maximum pixel value in the image.

$$z \geq \mu + U\sigma \tag{1}$$

A "weak point" is defined to be a pixel that has a value, z, that is greater than or equal to the mean value of the image, μ, minus the quantity of the "lower sensitivity level", L, times the standard deviation of the image, σ. This threshold is set to be no lower than the minimum pixel value in the image. In addition, the pixel must neighbor a strong point.

$$z \geq \mu - L\sigma \tag{2}$$

The condition used for controlling the recursion depth of step three is the "subcluster level" parameter. A value of zero would indicate that no recursion should be done. In addition, if a parent cluster has less than six pixels, then no children is sought. Finally, if the parent cluster produces no children upon recursion, then the parent cluster is retained.

There are five input parameters that are used to control the behavior of the SPA algorithm: upper and lower sensitivities, subcluster level, reach and pixel promotion.

The upper sensitivity level, U, is used for determining the strong points. It has units of standard deviations. Higher values allows for smaller chunks of clusters and the elimination of smaller-value clusters. The lower sensitivity, L, is used for determining the weak points. Like the upper sensitivity level, it also has units of standard deviation. Larger values allows for a "more complete" region-growing of the cluster. The region-growing method for SPA is admittedly rudimentary. The weak points that neighbors one or more strong point are "padded" onto the network of strong points. This parameter has very little impact on overall cluster results, but can make a significant improvement when SPA is used on low-resolution images. A distance criteria is used for determining the pixels that neighbors a given pixel. This is controlled by the reach parameter, r.

The subcluster level is used for controlling the number of recursive iterations of SPA is applied to the image. Larger integer values lead to more chunks of clusters and the elimination of flat (very low standard deviation) clusters. Larger values forces SPA to separate out the nuances of complex images.

The pixel promotion effects the "completeness" of the region-growing portion of the algorithm. Larger values signify more padding. Pixels that do not satisfy the primary definition of a weak point can still be labeled as a weak point if it is sufficiently surrounded by strong points. The larger the parameter is, the more strong points are needed to be neighboring so that a pixel may be promoted to weak point. This parameter has, at most, only an indirect effect on the number of clusters.

3. Results

Radar reflectivity images from several WSR-88D sites were provided as input to the SPA algorithm in order to test its ability to produce consistent clustering results. First, the behavior of the algorithm for changing input data is studied while holding the input parameters constant. Of particular interest are the clusters it produces for a time-series of images. Figure 1 shows the reflectivities and cluster results from KINX in Tulsa, Oklahoma on May 30th, 2004 at 0147, 0202, and 0207 GMT. In this time series, the images are largely similar. However, they are different enough to test SPA's behavior while input parameters are held constant. SPA identified these clusters using an upper sensitivity Level of 1.1, a lower sensitivity Level of 0.9, a pixel promotion parameter of 1, a padding parameter of 2.5, a reach of 1.5, and a subcluster level of 1. Clusters are outlined by a thick, solid, yellow line. Note that these outlines circumscribes more pixels than what SPA identified for a cluster. This is merely a limitation of the author's graphical programming ability, and not the fault of SPA.

It is also important to study the behavior of the algorithm with changes to the input parameters for a given input image. Figure 2 shows the reflectivity map from the KDDC radar site in Dodge City, Kansas on July 8th, 2004 at 0028 GMT. This reflectivity map shows a very strong storm system northeast of the radar site that is composed of several possible segments.

Figure 3 shows several clustering results from SPA as a function of changing values of the upper sensitivity level and the subcluster level. The other input parameters were kept constant and are the same as they were for the previous result. The figures are laid out in a grid such that the subcluster level parameter, n, increased from 0 (left) to 2 (right), and the upper sensitivity level, U, increased from 0.5 (top) to 1.5 (bottom).

4. Conclusions

In a situation where the input images are slowing changing, such as in figure 1, the clusters that are returned for each time-step appear reasonable and consistent. An undesirable behavior would have been for the size scale of the clusters to keep changing over time, and for pixels to swap membership between clusters. This did not occur in figure 1, and thus SPA appears to have a consistency property to its results. In addition, this time series hints that the parameters of the algorithm can be held constant over long periods of time, which would be useful for large scale, automated data analyses.

In terms of the dependency of the clusters upon changes in the input parameters, figure 3 shows that the results have an almost hierarchal characteristic to them. The clusters monotonically changes with the changes in the upper sensitivity level and the subcluster level. This is a useful property for tweaking the parameters to produce desirable results.

What is important with both of these observations is that the ability of the algorithm to consistently identify objects in images is beneficial to an object tracking algorithm. Consistent identification of objects means that the centroid of the object (often the primary information used in tracking) will have reduced noise for its location. Additionally, because the clusters have stable partitioning, the tracking algorithm will have less noise from the cluster centroid disappearing and then reappearing over time.

Acknowledgement Radar data was obtained using the HDSS website of the National Climatic Data Center's HDSS Access System (HAS). NCDC's Java NEXRAD Tools was then used to read the Level-2 data files to produce images and to produce NetCDF files, which was used as input to SPA.

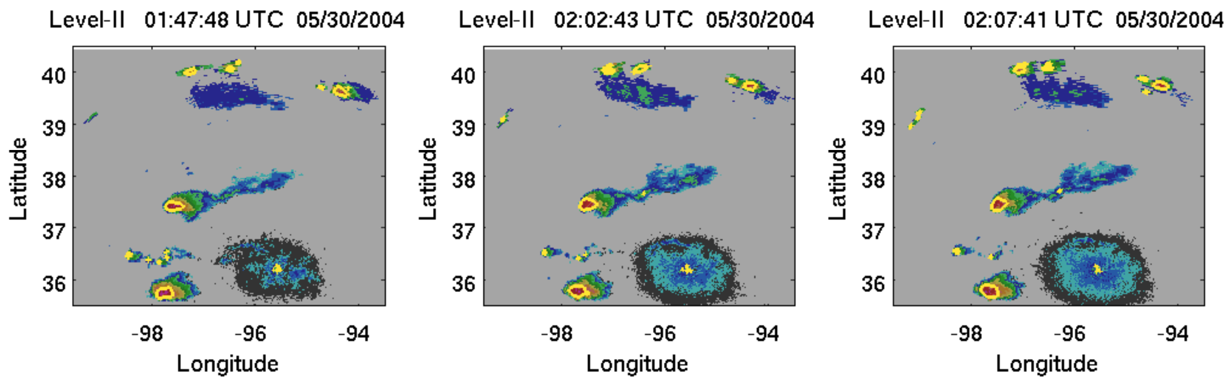

Figure 1: Time series of radar images from KINX in Tulsa, Oklahoma on May 30th, 2004 at 0147, 0202, and 0207 GMT. The clusters are indicated by the thick, yellow lines.

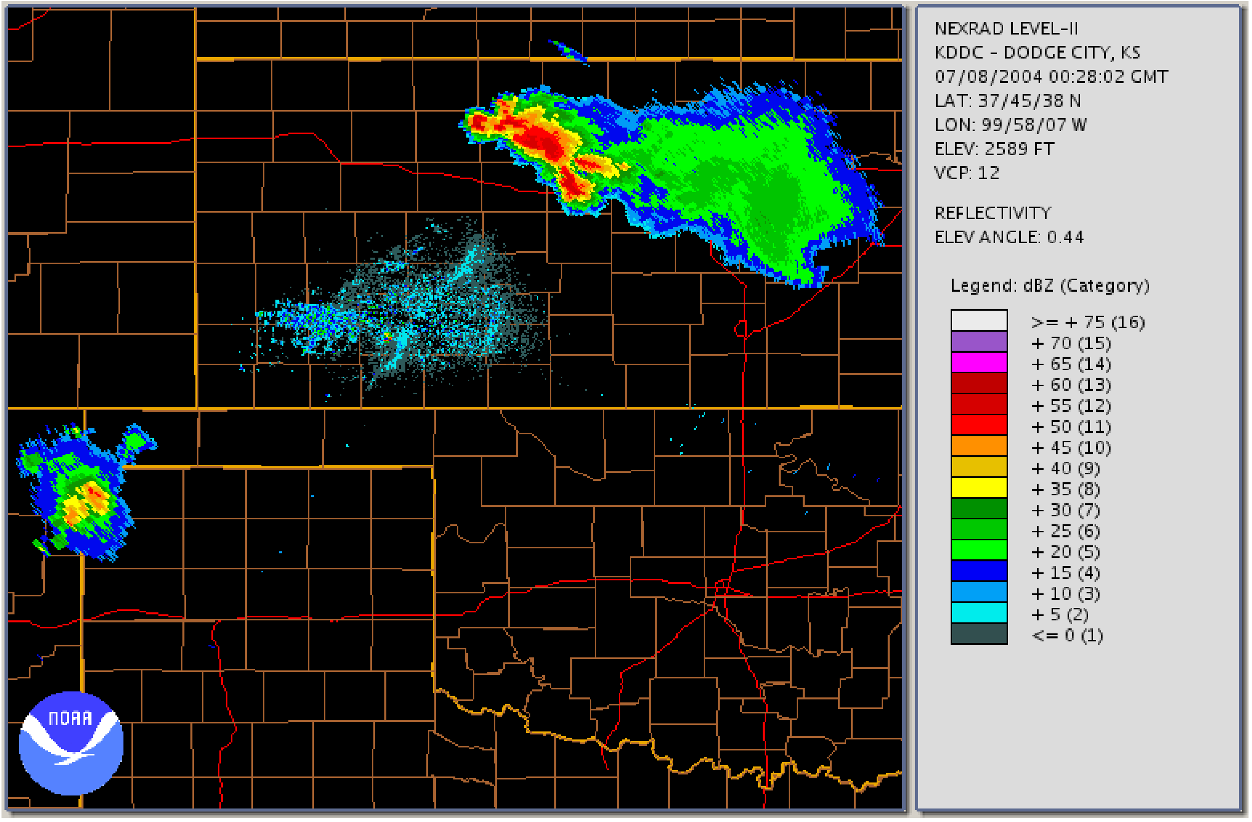

Figure 2: Radar reflectivity from KDDC in Dodge City, Kansas on July 8th, 2004 at 0028 GMT.

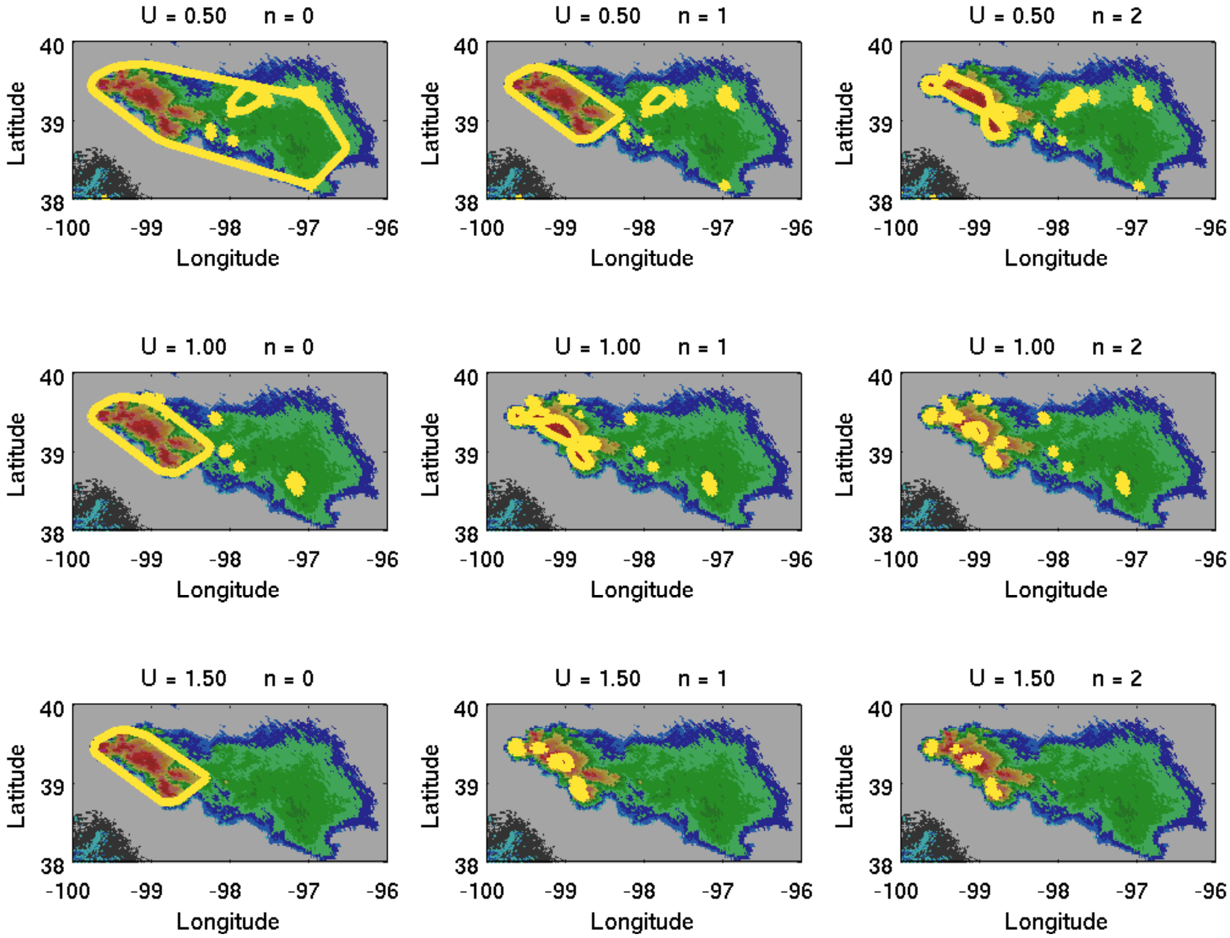

Figure 3: Demonstration of SPA output over a gamut of input parameter values, while holding the input image constant. The input image is a subsection of the radar reflectivity image from KDDC in Dodge City, Kansas on July 8th, 2004 at 0028 GMT. The input parameter "subcluster level" increases from 0 (left) to 2 (right). The input parameter "upper sensitivity level" increases from 0.5 (top) to 1.5 (bottom). Clusters are indicated by the thick, yellow lines.