Clustering elliptical anomalies in sensor networks

James C. Bezdek, Timothy C. Havens, James M. Keller, Chris Leckie, Laurence Park, Marimuthu Palaniswami, Sutharshan Rajasegarar

***Abstract*— We model anomalies in wireless sensor networks with ellipsoids that represent node measurements. *Elliptical anomalies* (EAs) are level sets of ellipsoids, and classify them as type 1, type 2 and higher order anomalies. Three measures of (dis)similarity between pairs of ellipsoids convert model ellipsoids into dissimilarity data. Clusters in the dissimilarity data may correspond to normal and anomalous measurements and nodes in the network. Assessment of (clustering) tendency is facilitated by visual inspection of (VAT/iVAT) images. Two examples illustrate the potential for anomaly detection.**

Keywords- Anomaly detection, Visual assessment of clustering tendency, Elliptical similarity, wireless sensor networks

I. INTRODUCTION[1]

The application that motivates the present work is the use of ellipsoids for distributed anomaly detection in *wireless sensor networks* (WSNs) [1-2]. The authors of [2] model the data collected at individual sensor nodes by sample-based ellipsoids; and [3] presents a method for clustering sets of ellipsoids in this context. The authors of [4] further this effort by defining three measures of similarity for pairs of ellipsoids, and then use them for visual assessment of clustering tendency (VAT, [5]). Algorithm iVAT [6] augments VAT by applying a path-based distance transform to the input dissimilarity data before VAT images are made. This note combines the models in [2, 4, 6] with the objective of improving this scheme for anomaly detection in WSNs.

Figure 1 depicts a typical hierarchical WSN with subtree heads and a gateway node. Examples of WSNs include: the Great Duck Island project in Maine for monitoring sea birds; the Great Barrier Reef (GBR) project to monitor ocean parameters at reef sites in Australia; and the Intel Berkeley Research Laboratory (IBRL) deployment [2].

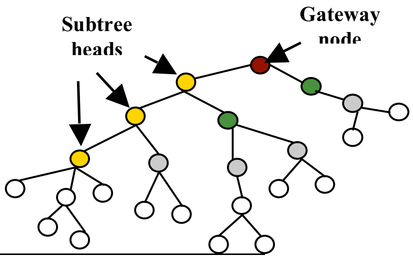


Figure 1. Typical architecture of a hierarchical WSN

II. ANOMALIES IN SENSOR NETWORKS

The literature contains many definitions of anomalies in data measurement, and there are many algorithms available for their detection [7]. Many of these algorithms are designed for large data sets and assume substantial processing capabilities, so these techniques are often inappropriate for anomaly detection in resource limited sensor networks. Consequently, there has been substantial interest and activity about research concerning techniques that can be used in WSNs: see Bettencourt, Hagberg and Larkey [8] for an introduction to some of this literature.

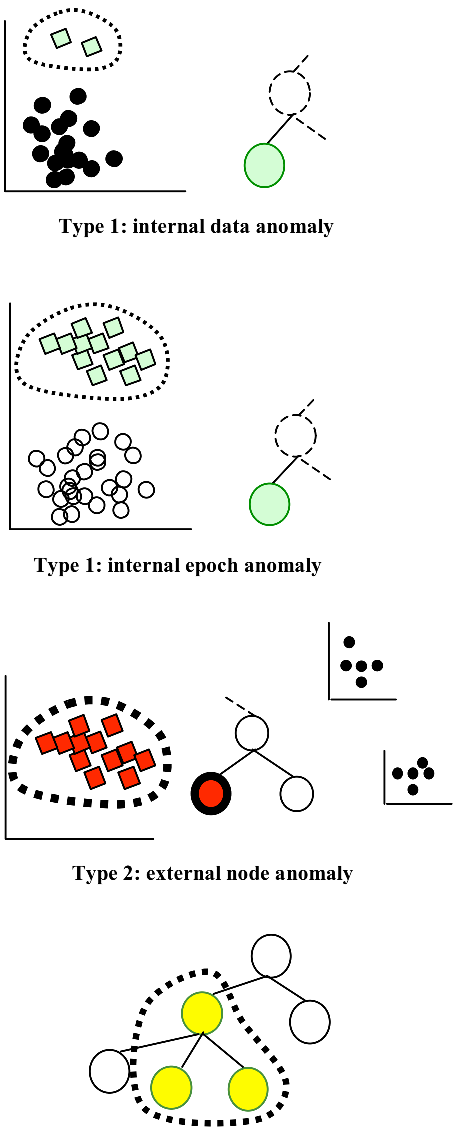
Each sensor in a WSN collects data measurements at specified time intervals over the life of the network. Various events will alter data collected at individual sensors or subtrees in a WSN. For example, the local temperature at a node or in a subnetwork might be much higher than in other parts of the network due to a fire, sunshine, or proximity to a fixed heat source. This may occur at just a few isolated times, or over an epoch within the collection window, or for the duration of data collection. During operation, anomalies at an individual node can occur occasionally (intermittent fault), or over the whole time epoch (e.g., power failure).

We classify anomalies according to the characteristics of individual data measurements made by each sensor at each node. Fig. 2 depicts the four types of anomalies recognized by our model. The upper view in Fig. 2 shows several data points (the squares) that differ a lot from the rest of the data at the same (shaded gray) node. This is a *type 1(or first order) data anomaly*, internal to this node. The second sketch in Fig. 2 illustrates an internal *type 1 (or first order) epoch anomaly*. Here a subset of data measurements (the squares) over some contiguous time epoch in the measurement window differs enough from the general trend at the node to warrant being regarded as abnormal. This could, for example, be the result of a temporary change in sensor environment. The third view in Fig. 2 illlustrates a *type 2 (or second order) external anomaly*. Here, *all* of the data (the squares in the data space of the gray node) differ from those being observed at neighbor nodes, so this is an anomalous node. The bottom view in Fig. 2 shows a subnetwork of three gray nodes that are all producing data different from their neighbor nodes. This is a *higher order (HO) external anomaly*. All four types of anomalies are well

Bezdek, Havens, Keller: U. of Missouri, USA; Leckie, Park, Palaniswami, Rajasagar: U. of Melbourne, Au. Corresponding author: jcbezdek@gmail.com

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

known in real WSN deployments. See [3] for some examples based on the IBRL data. Our model is set up to detect type 1 anomalies in the data space. type 2 and HO anomalies are detected in a new feature space of ellipsoids.



Type 1: internal data anomaly

Type 1: internal epoch anomaly

Type 2: external node anomaly

Type HO: external subtree anomaly

Figure 2. Types of Anomalies in Sensor Networks

III. Elliptical Anomalies

In order to detect these different types of anomalies, a model is required that represents the structure of normal (i.e., non-anomalous) data in the network. Our model depends on the geometry of ellipsoids. Let vectors $\mathbf{x},\mathbf{m}\in\Re^p$, and $A\in\Re^{p\times p}$ be positive definite. The quadratic form $Q(\mathbf{x})=\mathbf{x}^TA\mathbf{x}$ is positive definite. For fixed m, let $E(A,\mathbf{m};t)=\{\mathbf{x}\in\Re^p \mid \|\mathbf{x}-\mathbf{m}\|_A^2=t^2\}$. This set is the (surface of the) hyper-ellipsoid in p-space induced by A, all of whose points are the constant A-distance (t) from its center **m**. We normalize each member of the family $\{E(A,\mathbf{m};t):\ t>0\}$ as $\|\mathbf{x}-\mathbf{m}\|_{A/t^2}^2=1$. Scaling a spheroid is done by matrix multiplication with a scaling matrix $S=\mathrm{diag}(s_k)$, where s_k is the scaling factor for dimension t, $1\le k\le p$. Rotation through an angle θ is accomplished by matrix multiplication with a unitary rotation matrix $R=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$ (in $\Re^2$). Any point **z** in the unit sphere can be mapped to an ellipsoid via scaling, rotation and shift, $\mathbf{z}\to\mathbf{x}=RS\mathbf{z}+\mathbf{m}$.

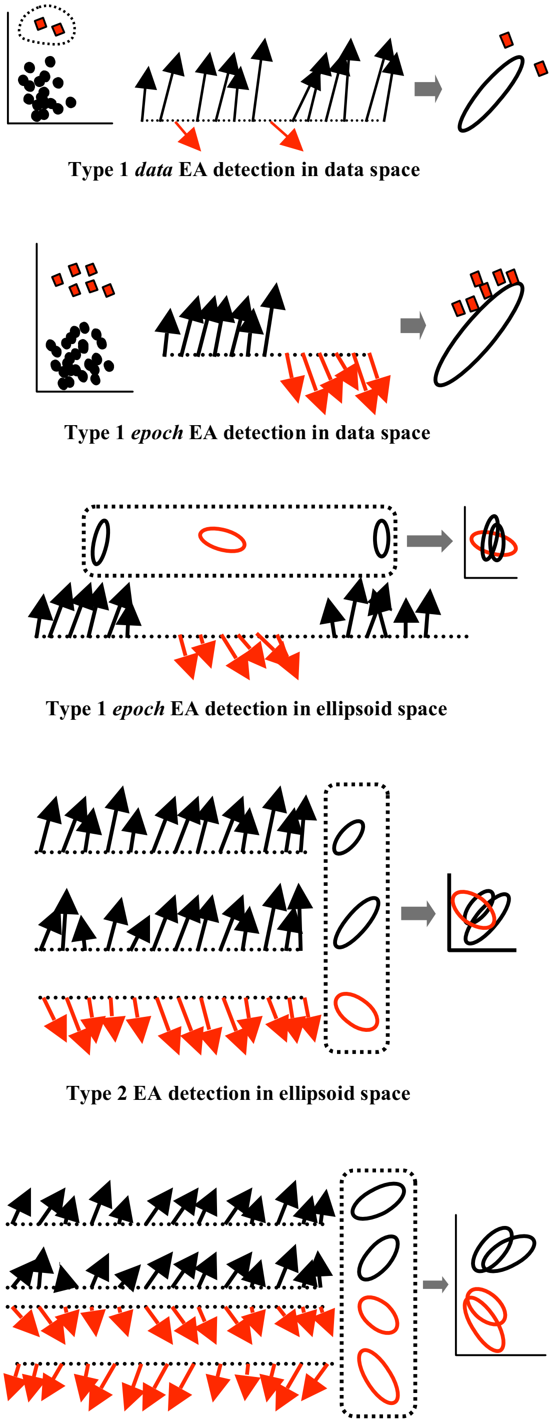
Let $X=\{\mathbf{x}_k\}\subset\Re^p$ be a collection of measurement vectors that record sensor measurements at a node, e.g., temperature, humidity, etc. Let (**m**, S) denote the *sample* mean and covariance matrix of X. Normal and anomalous points *in the data set X* are now defined, relative to the hyperellipsoidal parameters (**m**, S^{-1}) as

$$NP_{X,t}\equiv NP_X(S^{-1},\mathbf{m};t)=\{\mathbf{x}_k\in X \mid \|\mathbf{x}_k-\mathbf{m}\|_{S^{-1}}^2\le t^2\} \quad (1)$$

$$AP_{X,t}\equiv AP_X(S^{-1},\mathbf{m};t)=\{\mathbf{x}_k\in \mathbf{X} \mid \|\mathbf{x}_k-\mathbf{m}\|_{S^{-1}}^2> t^2\} \quad (2)$$

These definitions use level sets of the sample-based Mahalonobis distance to create crisp 2-partitions of X. Reminder: please don't confuse our use of the word "normal" here with its meaning in probability and statistics. Here, normal is synonomous with "not anomalous". In particular, we do *not* make any assumptions about the distribution of the data set X; (**m**, S) is just a compact, useful way to represent the collected data.

Setting thresholds on the Mahalonobis distances in equation (1), and hence (2), can be done in various ways. Three definitions of *elliptical anomalies* (EAs) are used in [2] as the basis for identification of type 1, type 2 or HO anomalies. For example, the *Elliptical Cardinality Anomalies* (ECAs) of X are defined as follows. Let c ∈ {0, 0.01, 0.02, .., 1}. Compute the n distances $\{d_k\}=\left\{\|\mathbf{x}_k-\mathbf{m}\|_{S^{-1}}^2\right\}$ and sort them in ascending order, $d_{(1)}\le d_{(2)}\ \ldots\le d_{(\lfloor cn\rfloor)}\le\ \ldots\ \le d_{(n)}$. The sets $NP_{X,(\lfloor cn\rfloor)}$ and $AP_{X,(\lfloor cn\rfloor)}$ are the 100c% ECA partition of X. The hyperellipsoid that contains 100c% of the samples in X has an effective radius $t=d_{(\lfloor cn\rfloor)}$. For example, if X has n = 250 samples, then $NP_{X,(\lfloor 0.9n\rfloor)}$, $NP_{X,(\lfloor 0.95n\rfloor)}$, and $NP_{X,(\lfloor 0.99n\rfloor)}$ contain, respectively, 225, 237 and 247 of the 250 samples in X. That is, these three sets contain the closest (in Mahalonobis distance) 90%, 95% and 99% of the points to the sample mean. Conversely, they identify, respectively, the 25, 13 and 3 points furthest from the sample mean as companion anomaly sets.We will call the family $\left\{NP_{X,(\lfloor cn\rfloor)}\cup AP_{X,(\lfloor cn\rfloor)}:c=0.01,\ldots,1\right\}$ the elliptical cardinality anomalies of X. The sets $NP_{X,t}$ and $AP_{X,t}$ are used to detect elliptical anomalies as depicted graphically in the upper two views of Fig. 3.



Type 1 *data* EA detection in data space

Type 1 *epoch* EA detection in data space

Type 1 *epoch* EA detection in ellipsoid space

Type 2 EA detection in ellipsoid space

HO EA detection in ellipsoid space

Figure 3. Anomaly detection in data and ellipsoid spaces

The lower three views in Fig. 3 show a representation of the WSN data by their elliptical parameters. The ellipses become the data that are used for anomaly detection. For internal node assessment, windowed subsets of data vectors are converted to sample-based ellipsoids. For whole node (type 2) and substree (type HO) anomaly detection, all the data collected at each node over a specified time interval are represented by one sample-based ellipsoid. Our visual anomaly detection approach requires a way to assess elliptical similarity, the topic we turn to next.

IV. SIMILARITY MEASURES FOR ELLIPSOIDS

Suppose we have two ellipsoids in p-space, E_i and E_j, that have effective radii t_i and t_j, centers [means] $\mathbf{m}_i$ and $\mathbf{m}_j$ and [covariance] matrices A_i and A_j. First, normalize E_i and E_j, $A_i/t_i^2 \leftarrow A_i$ and $A_j/t_j^2 \leftarrow A_j$. Then $s(E_i, E_j)$ is a similarity measure for E_i and E_j. if, and only if, three conditions are satisfied: (s1) $s(E_i,E_j)=1 \Leftrightarrow E_i = E_j$; (s2) $s(E_i,E_j)=s(E_j,E_i) \;\; \forall\, i \neq j$; and (s3) $s(E_i,E_j)>0 \;\; \forall\, i, j$. We summarize three measures of similarity between ellipsoid pairs that are presented and analyzed in [4].

Compound Similarity Let $\boldsymbol{\alpha} = \{\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_p\}$ and $\boldsymbol{\beta} = \{\beta_1 \leq \beta_2 \leq \ldots \leq \beta_p\}$ be the ordered eigenvalues of A_i and A_j for ellipsoids $E_i = (A_i, \mathbf{m}_i, 1)$ and $E_j = (A_j, \mathbf{m}_j, 1)$, and $\boldsymbol{\alpha}^* = (1/\sqrt{\alpha_1},\ldots,1/\sqrt{\alpha_p})^T$ and $\boldsymbol{\beta}^* = (1/\sqrt{\beta_1},\ldots,1/\sqrt{\beta_p})^T$. The vector of angles between paired eigenvectors of A_i and A_j is $\theta = \arccos(\mathrm{diag}(R_i^T R_j))$. The *normalized compound similarity* between (E_i, E_j) is

$$s_{cn}(E_i,E_j) = e^{-\left(\left\|\mathbf{m}_i - \mathbf{m}_j\right\|^2_{(A_i+A_j)^{-1}} + \left\|\sin\theta\right\| + \left\|\boldsymbol{\alpha}^* - \boldsymbol{\beta}^*\right\|\right)} \qquad (3)$$

where $\left\|\mathbf{m}_i - \mathbf{m}_j\right\|^2_{(A_i+A_j)^{-1}} = (\mathbf{m}_i - \mathbf{m}_j)^T (A_i + A_j)^{-1} (\mathbf{m}_i - \mathbf{m}_j)$.

Transformation Energy Similarity. A point in the space of ellipsoid E_i can be mapped to a common co-ordinate space by scaling the point by S_i^{-1}, reversing the rotation by R_i^{-1}, then shifting the point away from the origin by translation by $\mathbf{m}_i$. Within this common space the point can then be mapped into the space of E_j by shifting the point by $\mathbf{m}_j$, rotating by R_j and scaling by S_j. The mapping is summarized as

$$\begin{aligned}\mathbf{x}_j &= f(\mathbf{x}_i \mid E_i, E_j) = S_j R_j (R_i^{-1} S_i^{-1} \mathbf{x}_i - \mathbf{m}_i + \mathbf{m}_j) \\ &= \underbrace{S_j R_j R_i^{-1} S_i^{-1}}_{M_{ij}} \mathbf{x}_i + \underbrace{S_j R_j (\mathbf{m}_j - \mathbf{m}_i)}_{\mathbf{d}_{ij}} = M_{ij}\mathbf{x}_i + \mathbf{d}_{ij},\end{aligned}$$

where $E_i = E(R_i S_i^{-2} R_i^T, \mathbf{m}_i; 1)$, $E_j = E(R_j S_j^{-2} R_j^T, \mathbf{m}_j; 1)$. Now define $\left\|f(E_i,E_j)\right\|_2 = \max\left\{f(\mathbf{z} \mid E_i, E_j) : \; \mathbf{z} \in \Re^p ; \; \left\|\mathbf{z}\right\|_2 = 1\right\}$. The *transformation energy* similarity function is

$$s_{te}(E_i,E_j)=1\Big/\max\left\{\left(\left\|f(E_i,E_j)\right\|_2,\left\|f(E_j,E_i)\right\|\right)_2\right\}. \tag{4}$$

Bhattacharya coefficient similarity. The Bhattacharya similarity coefficient between two ellipsoids is

$$\begin{aligned} s_{bc}(E_i,E_j) &= e^{-\frac{1}{8}\left\|\mathbf{m}_i-\mathbf{m}_j\right\|^2_{\left[(A_i+A_j)/2\right]^{-1}}} \\ &+\frac{1}{2}\ln\left[\det\left((A_i^{-1}+A_j^{-1})/2\right)\Big/\sqrt{\det A_i^{-1}\det A_j^{-1}}\right] \end{aligned} \tag{5}$$

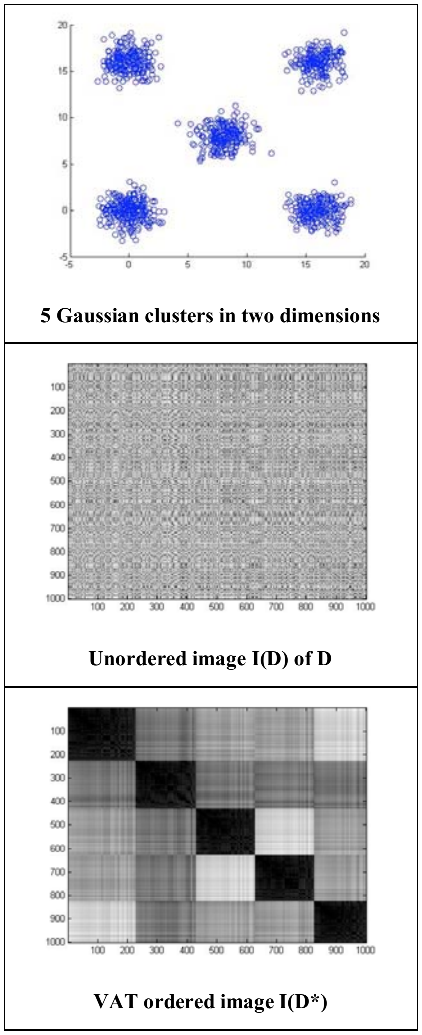
The authors of [4] present several numerical examples that compare and contrast these three measures of ellipsoidal similarity on various sets of ellipsoids. The problem that these measures is designed to solve is the detection of anomalies as shown in the lower three views of Fig. 3. To assess what anomalies may exist in sets of ellipsoids, we turn to a method based on visual examination of dissimilarity data.

V. Visual Assessment of Clustering Tendency

Many visual methods for clustering tendency begin with the *reordered dissimilarity image* (RDI). The intensity of each pixel in an RDI corresponds to the dissimilarity between the addressed row and column objects. An RDI is "useful" if it highlights potential clusters as a set of "dark blocks" along its diagonal. Each dark block represents a group of objects that are fairly similar. The method used here is called VAT [5], which creates and displays a grayscale image whose ij-th element is a scaled dissimilarity value between objects o_i and o_j. Each element on the diagonal of the VAT image is zero. The off diagonal values range from 0 to 1. If an object is a member of a cluster, then it also should be part of a submatrix of similar values, whose diagonal is superimposed on the diagonal of the image matrix.

The uppermost view in Fig. 4 is a scatterplot of n=1000 data points in two dimensions drawn from a mixture of five normal distributions. These object data were converted to $D_{1000\times1000}=D=[d_{ij}]=\left[\left\|\mathbf{x}_i-\mathbf{x}_j\right\|\right]$ using the Euclidean norm. The c = 5 visually apparent clusters in the upper view of Fig. 3 are suggested by the 5 distinct dark diagonal blocks in the lower view, I(D*), which is the VAT RDI of the data after reordering. Compare this to the image seen in the center view, which is the image I(D) of the dissimilarities in random order. It is clear from this example that reordering is necessary to reveal the structure of the underlying data.

However, there are many data sets whose clusters are not well defined, or are confusingly arranged. This type of substructure is not readily captured by VAT reordering and its subsequent image may not be so useful. This is not surprising, because VAT reordering is explicitly tied to the type of data favored by single linkage clustering [9]. In an effort to overcome this limitation, the authors of [6] introduced a transformation of the input dissimilarity matrix that is applied to D before submitting it to the VAT algorithm.



5 Gaussian clusters in two dimensions

Unordered image I(D) of D

VAT ordered image I(D*)

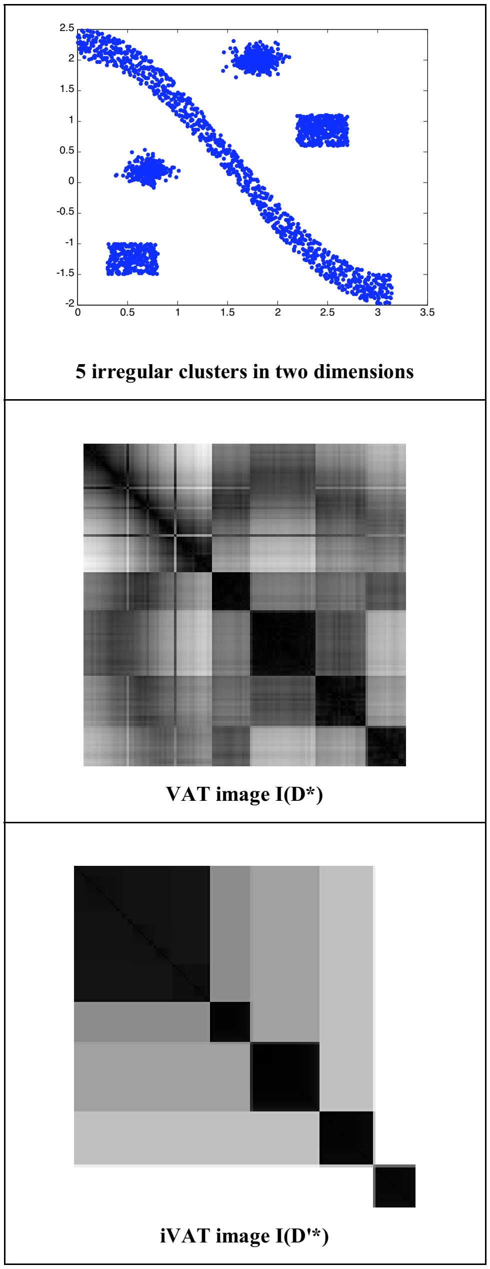
Figure 4. Unordered and VAT images of the five Gausaian clusters

The iVAT transformation replaces each dissimilarity d_{ij} in D by the minimax path distance between nodes i and j (objects i and j in the input data). Let P_{ij} be the set of all possible paths from o_i to o_j. The path based distance between o_i and o_j is seen in the transform

$$D'=[d'_{ij}=\underbrace{\min}_{p\in P_{ij}}\{\underbrace{\max}_{1\le h<|p|}\{d_{p[h]p[h+1]}\}\};1\le 1,j\le n\}] \tag{6}$$

where p[h] denotes the object at the h-th position in path p, and |p| is the length of this path.

The upper view of Fig. 5 shows a scatterplot of data that contains c = 5 visually apparent clusters. The object data were converted to D using the Euclidean norm. The middle view in Fig. 5 is the VAT image, I(D*) of the input matrix D. The VAT image does not possess nice clean visual structure like that in the bottom view in Fig. 4, even though both sets of object data have five pretty compact and well separated clusters. The problem in the data set of Fig. 5 is the large irregularly shaped cluster in the center of the data, which occupies the upper left portion of the VAT image. You might talk yourself into seeing 5 dark blocks along the diagonal of this image, but it's arguable whether there is any clearly evident structure in I(D*). The pixels corresponding to the stripe cluster are jumbled in appearance.



5 irregular clusters in two dimensions

VAT image I(D*)

iVAT image I(D'*)

Figure 5. VAT and iVAT images of the data set in the upper view

The bottom view in Fig. 5 is the iVAT image I(D'*) of this data set, made by first transforming D→D' with the path based distance transform in (6), and then applying VAT to D'. No prompting is needed for you to see the improvement made in the image. The iVAT image clearly shows the underlying structure of the input data. Examples such as this encouraged us to add the iVAT transform to the elliptical similarity algorithms reported in [3]. Are the VAT and iVAT orderings of an input data set are always the same? The answer is *yes*. We have a proof of this that will appear in a forthcoming paper. Is an iVAT image is always at least as "good" as the VAT image of the same D. We think this is true, but this is a subjective judgment, not a provable assertion.

We next give a concise statement of VAT/iVAT. In general, the functions arg max and arg min in Steps 1 and 2 are set valued, and when the sets contain more than one pair of optimal arguments, any optimal pair can be selected. The VAT/iVAT reordering is stored in the array P = (P(1), ..., P(N)). Applying VAT to D results in the image I(D*); applying iVAT to D results in the image I(D'*).

VAT/iVAT: Visual Assessment of Tendency [5, 6]

```
Input:  Dissimilarities $D_{n \times n}$ for $O = \{o_1,\ldots,o_n\}$;
        (Convert similarity data $S_{n \times n}$ as $D = [1]-S$)

    Step 1. $K = \{1,\ldots,n\}$; select $(i,j) \in \arg\max_{p \in K, q \in K} \{D_{pq}\}$ ;
        Set $P(1) = i$; $I = \{i\}$; and $J = K - \{i\}$.
    Step 2. For $t = 2,\ldots, n$: select $(i,j) \in \arg\min_{p \in I, q \in J} \{D_{pq}\}$ ;
        $P(t) = j$;  $I \leftarrow I \cup \{j\}$ and $J \leftarrow J - \{j\}$.
    Step 3. Form the ordered dissimilarity matrices
        [VAT]: D*:  $d^*_{ij} = d_{P(i)P(j)}$, for $1 \le i, j \le n$.
        [iVAT]:  $D'^* = [0]_{n \times n}$; for $r = 2; \ldots ; n$ do
            $j = \underbrace{\arg\min}_{k=1,\ldots,r-1} \{D^*_{rk}\}$
            $D'^*_{rc} = D^*_{rc}, c = j$
            $D'^*_{rc} = \max\{D^*_{rj}, D'^*_{jc}\},\ \ c = 1,\ldots,r-1,\ \ c \ne j$
            For $2 \le <j \le n$; $i < j$: $D'^*_{ji} = D'^*_{ij}$
    Step 4. Display I(D*) and I(D'*), scaled so that
        $\underbrace{\max}_{1 \le i, j \le n} \{d^*_{ij}\}$, $\underbrace{\max}_{1 \le i, j \le n} \{d'^*_{ij}\}$ = white and 0 = black.
```

Constructing the iVAT matrix D' matrix with (6) can be computationally expensive ($O(n^3)$). The recursive computation of D'* given here does not alter the VAT order of D*, has the same ordering as D*, and is $O(n^2)$. [10, 11].

VI. Visual Assessment of Elliptical Anomalies

Now we are ready to assemble our approach to detecting possible (elliptical) anomalies using data collected by wireless sensor network nodes. Let E denote a set of n ellipsoids in p-space. For each $(E_i, E_j) \in E \times E$, we compute $s_{ij} = s(E_i, E_j)$ with any of the three measures of similarity described in Section IV, and array these n^2 values as the n×n similarity relation matrix $S = [s_{ij}]$. The transformation $D = [d_{ij}] = [1 - s_{ij}]$ yields a dissimilarity relation on $E \times E$. Consequently, applying the VAT or iVAT algorithms to D will yield RDIs that can be used to assess clustering tendencies of the ellipsoids in $E \times E$.

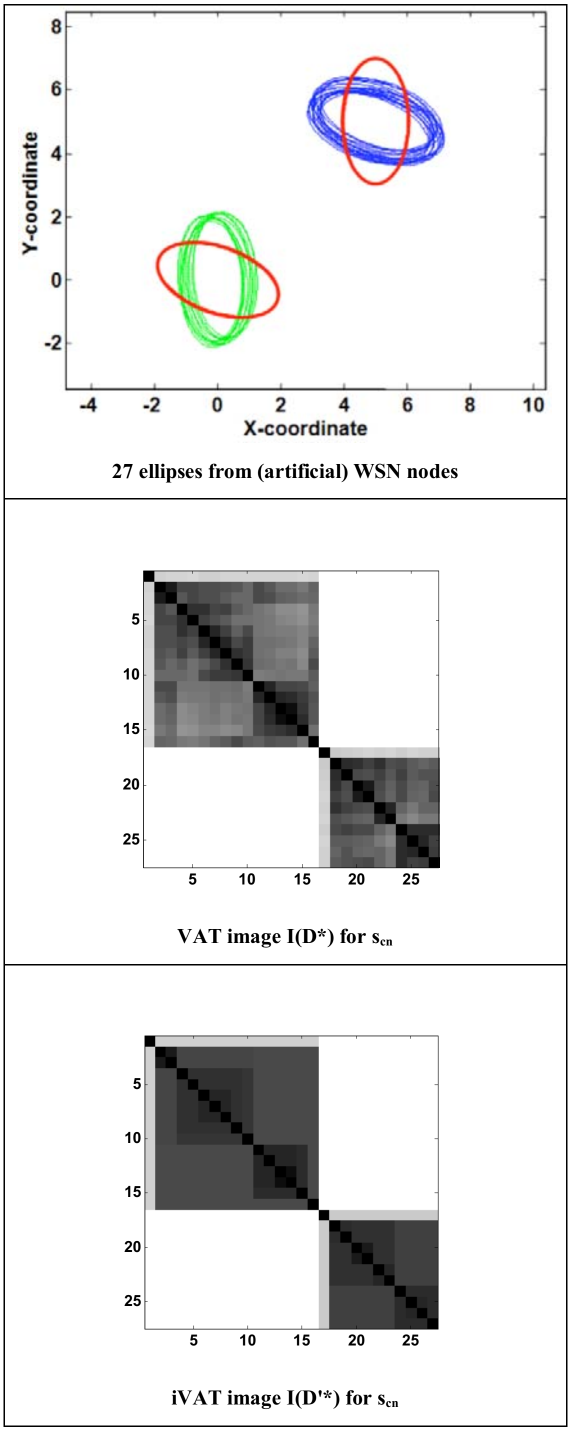
First we can apply this strategy at a single node. Thus, if E_j is the set of n ellipsoids collected at node j in the WSN, we can examine the VAT/iVAT images of E_j to assess the possibility that there are type 1 epoch anomalies at node j (some ellipsoids at this node differ significantly from the rest, as shown in the middle view of Fig. 3).

At the next level, we can aggregate (single) ellipsoids from different nodes in the WSN, and look for type 2 anomalies in E (almost all of the ellipsoids are different from those at neighbor nodes). And finally, we can represent each node in the entire WSN by an ellipsoid, and search for subsets of ellipsoids that seem abnormal (or at least different). Such a subset corresponds to a subtree of nodes that are behaving differently than their neighbors - that is, a higher order EA.

Example 1. Our first example is shown in Fig. 6. The upper view in this Fig. shows the (simulated) input data to be a set of 27 ellipsoids in two dimensions. There are 11 ellipses in the lower left primary cluster, and 16 ellipses in the upper right primary cluster. Each primary cluster has one aberrant ellipse, so each primary cluster has two subclusters. These 27 ellipses can represent several different scenarios. For example, they might be sets of ellipses built from data collected at just 2 nodes taken over consecutive time intervals, with one set of intervals longer than the other. In this case, each of the nodes has a single type 1 epoch anomaly, the case represented by the middle graphic in Fig. 3.

A second possibility is that these 27 ellipsoids are single representatives of data collected over some fixed time interval at 27 nodes in a WSN. In this case the two main subsets would correspond to subtrees whose sensors were collecting data that differed from each other, and the single (rotated) ellipse centered (more or less) at the locations of the two clusters would then be a type 2 EA within the overall higher order anomaly structure seen in the data. You might argue that there are only two kinds of ellipses in this data, since the single type 2 anomalies in each subset appear quite similar to all but the single ellipse in the other subset. But they are similar only in orientation and shape; their centers are far apart. We hope that this feature will be seen by our measures of elliptical similarity. And it is.

Space constraints limit our presentation in Fig. 6 to the VAT and iVAT images of this data set for just the normalized compound similarity measure s_{cn} shown in equation (3). The center view is the VAT image I(D*), the lower view is the iVAT image I(D'*).



27 ellipses from (artificial) WSN nodes

VAT image I(D*) for s_{cn}

iVAT image I(D'*) for s_{cn}

Figure 6. Data and VAT/iVAT images for type 2 and HO elliptical anomalies using the normalized compound similarity measure (3).

What can you see? First, both images indicate that there are 4 clusters in this data. Each primary cluster block weakly contains the secondary structure of the type 2 anomalies - one rotated ellipse in each cluster. These are seen as single pixels adjacent but somewhat related to the larger dark blocks that correspond to the major group for each node. And third, the iVAT image in the lower view is *a little better* than the VAT image above it in that some of the ambiguity within the two major blocks is resolved in the iVAT view. We don't see the dramatic improvement provided to VAT by iVAT illustrated in Fig. 5, but we do see slight improvement.**Example 2.** Our second example concerns the set of 40 ellipses shown in Fig. 7. The lower left portion of the data comprises 25 ellipses that are centered at roughly the same coordinates. Ten of the twenty five ellipses in this cluster have been rotated about 90 degrees. The upper right part of this plot contains 15 ellipses that are quite similar to each other. Most observers will agree that there are three clusters (of ellipses) in this data set. In terms of elliptical anomalies there are again several possibilities, depending on how the data arose. For concreteness, assume that each ellipse represents the data collected by one node in a WSN over some fixed interval of time. In this case the data correspond to a network with two higher order anomalies - i.e., a network with three subtrees of nodes that are behaving differently from each other but which have good similarity within their subtree.

Figures 8 and 9 show the VAT and iVAT images of the dissimilarity matrix of this data built by each of our three measures of elliptical similarity. If you accept our hypothesis that there are three primary clusters in this data, this structure is evident in the left views of these two figures - that is, the images built with the normalized compound similarity measure at equation (3). The VAT and iVAT images for this measure both have three overriding dark blocks along their diagonals. And each of the three blocks appears in these views to have internal substructure.

Images for the Bhattacharya similarity in equation (4) - the center views in Figs. 8 and 9 - also have three primary dark blocks. Moreover, the secondary structure in these two views shows the two clusters of 10 and 25 ellipses in the lower left part of Fig. 7 quite nicely. The iVAT view in the center of Fig. 9 shows the three sets of ellipses perfectly: three black blocks of just the correct sizes. The image shows that the set of 25 in the upper portion of the image are two subsets of sizes 10 and 15 that are much more closely related to each other than to the other 15 ellipses. This is because these two sets are centered at roughly the same coordinates. We think this view of the data captures and represents its cluster structure completely and correctly.

The images corresponding to the transformation energy in equation (5) are less informative than those from the other two similarity measures. We are not sure if this will always be the case, since these are just the first few tests of these measures on data sets of this type.

Finally, we may again note that the iVAT images do seem to offer slight to moderate improvements over their VAT predecessors. However, there is noticeable improvement, encouraging us to believe that the path based distance used by iVAT is a useful change to the standard VAT implementation.

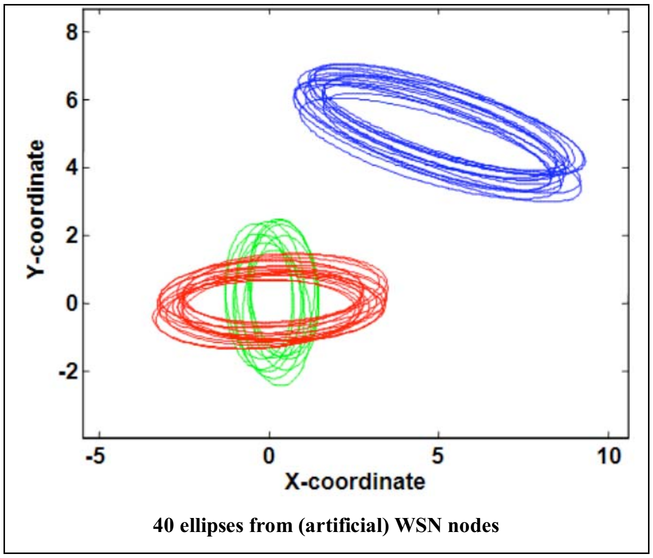


Figure 7. Simulated data for two HO elliptical anomalies

VII. Conclusions and Discussion

The sensors in almost all WSNs are small and cheap. The price we pay for this convenience is that they have very limited computational capability, and communication of data to other nodes in the network is difficult and impractical. This suggests the use of distributed models to minimize power consumption and communication overhead. Networks of this type suffer from various failures and environmental changes that contaminate the data they gather. Moreover, it is entirely possible that networks are maliciously attacked. Thus, anomaly detection is both necessary and desirable.

We defined four types of anomalies in WSNs. The simplest are type 1 data and epoch anomalies that occur at a single node. A type 2 anomaly is an entire node that seems abnormal within the context of other nodes in the network. And finally, higher order anomalies are subtrees within a network whose behavior is different from other parts of the same network.

Our model for anomaly detection resides in the geometry of ellipsoidal representation. Ellipsoids are an efficient, flexible way to compactly represent data collected by individual sensors. For example, a sensor that measures five variables and is sampled every 30 seconds for a day yields 60,120 real numbers, whereas the sample based ellipsoid built from this data comprises 20 real numbers. The advantage of the ellipsoidal representation for distribution and analysis of overall network performance is significant.

We introduced three measures of similarity for sets of ellipses. Visual VAT and iVAT displays of the similarity data are used to assess potential cluster structure in the data they processed. This, in turn, suggests how to detect various types of anomalies when the ellipses come to us in the context of elliptical representation of data from wireless sensor networks. The next two steps in the progression of this research are: (i) to *find* the clusters that are suggested by VAT/iVAT images (don't forget, these images simply suggest clusters, they do not partition the objects represented by the ellipsoids); and (ii) to apply the system described in this note to real data from a deployed wireless sensor network such as the IBRL data.

REFERENCES

[1] S. Rajasegarar, C. Leckie and M. Palaniswami. "CESVM: Centered hyperellipsoidal support vector machine based anomaly detection", *Proc. IEEE ICC 2008*, 1610-1614, 2008.

[2] S. Rajasegarar, J. C. Bezdek, C. Leckie and M. Palaniswami. “Elliptical Anomalies in Wireless Sensor Networks”, *ACM TOSN*, 6(1), 1550-1579, 2009.

[3] M. Moshtaghi, S. Rajasegarar , C. Leckie and S. Karunasekera. "Anomaly detection by clustering ellipsoids in wireless sensor networks", *Proc.* ISSNIP 2009, pp. 7-10, Melbourne, Australia, 2009.

[4] T. C. Havens, L. Park, J. C. Bezdek, C. Leckie, J. M. Keller, S. Rajasegarar and M. Palaniswami. "Clustering hyperellipsoids: a visual approach", in review, *Patt. Recog*, 2010.

[5] J. C. Bezdek and R. J. Hathaway. "VAT: A tool for visual assessment of (cluster) tendency". In *Proc. 2002 Int. Joint Conf. on Neural Networks*, Honolulu, HI, pp. 2225-2230, 2002.

[6] L. Wang, T. V. U. Nguyen, J. C. Bezdek, C. Leckie and K. Ramamohanarao. "iVAT and aVAT: enhanced visual analysis for cluster tendency assessment", in review, PAKDD, 2010.

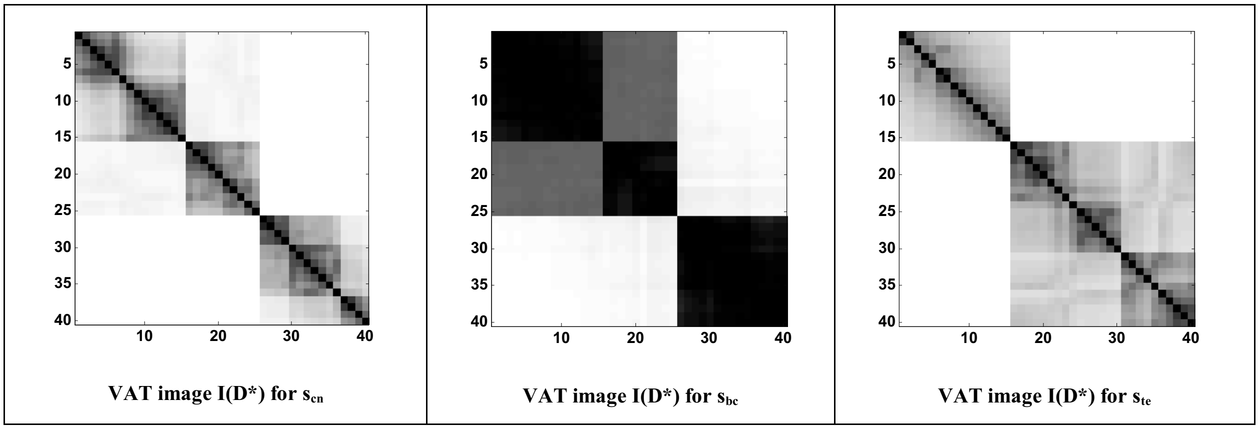
[7] V. Hodge. and J. Austin. "A survey of outlier detection methodologies", *AI Review,* Springer Netherlands, pp. 85 – 126, 2004.

[8] L. M. A. Bettencourt, A. A. Hagberg and L. B. Larkey. "Separating the wheat from the chaff: practical anomaly detection schemes in ecological applications of distributed sensor networks", *Proc. Distributed Computing in Sensor Systems*. Santa Fe, pp. 223–239, 2007.

[9] J. C. Bezdek, J.M. Keller, R. Krishnapuram and N. R. Pal. Fuzzy models and algorithms for pattern recognition and image processing. Kluwer, Norwell MA, 1999.

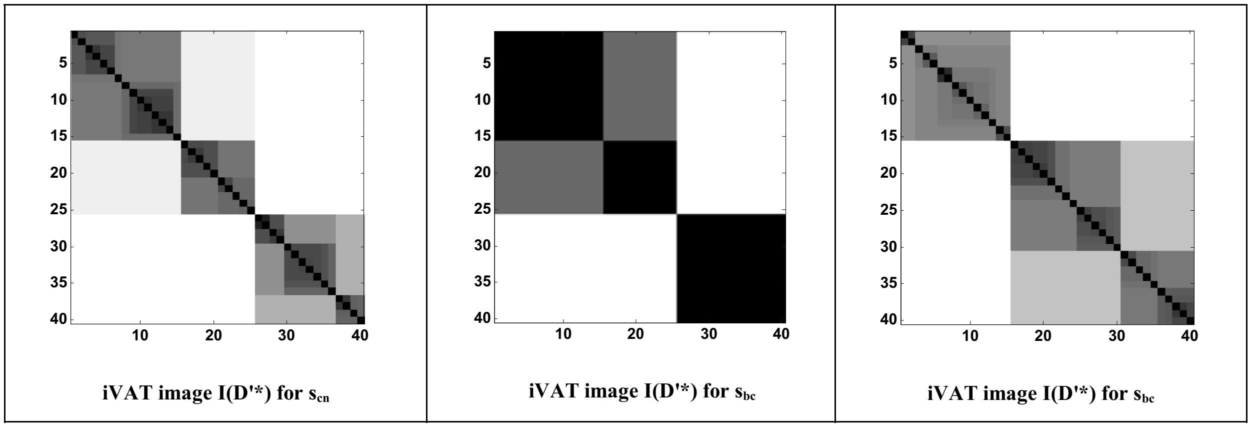
[10] T. C. Havens, J. C. Bezdek and J. M. Keller. "A new implementation of the co-VAT algorithm for visual assessment of clusters in rectangular relational data", in press, *Proc. 10th Conf. on Artificial Intelligence and Soft Computing* (ICAISC 2010), 2010.

[11] T. C. Havens and J.C. Bezdek. A recursive formulation of the improved visual assessment of cluster tendency (iVAT) algorithm. In review, *IEEE TKDE*, 2010.



VAT image I(D*) for s_{cn} **VAT image I(D*) for s_{bc}** **VAT image I(D*) for s_{te}**

Figure 8. VAT images for the data in Figure 7



iVAT image I(D'*) for s_{cn} **iVAT image I(D'*) for s_{bc}** **iVAT image I(D'*) for s_{bc}**

Figure 9. iVAT images for the data in Figure 7