Clustering ellipses for anomaly detection

Masud Moshtaghi\textsuperscript{a,e,*}, Timothy C. Havens\textsuperscript{b}, James C. Bezdek\textsuperscript{a,b}, Laurence Park\textsuperscript{c}, Christopher Leckie\textsuperscript{a,e}, Sutharshan Rajasegarar\textsuperscript{d}, James M. Keller\textsuperscript{b}, Marimuthu Palaniswami\textsuperscript{d}

\textsuperscript{a} Department of Computer Science and Software Engineering, University of Melbourne, Parkville, Melbourne, Australia
\textsuperscript{b} Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211, USA
\textsuperscript{c} School of Computing and Mathematics, University of Western Sydney, Australia
\textsuperscript{d} Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Melbourne, Australia
\textsuperscript{e} NICTA Victoria Research Laboratories, Australia

1. Introduction: clustering and ellipsoids

Hyperellipsoids (more simply, ellipsoids) occur in many areas of applied mathematics. For example, level sets of Gaussian probability densities are ellipsoids [1]. Ellipsoids also appear often in clustering [2–4] and classifier design [5,6–7]. Please be careful to distinguish the present work, wherein the input data objects are ellipsoids, from clustering algorithms such as that of Dave and Patel [4], where the output of clustering input sets of object vectors in $p$-space results in ellipsoidal prototypes.

The application that motivates the present work is the use of ellipsoids for anomaly detection. This problem occurs, for example, in wireless sensor networks (WSNs) [8–12] and motif-based patterned fabric defect detection [13]. In particular, the authors of [10] model the data collected at individual sensor nodes by sample-based ellipsoids; in [11] they develop a method for clustering sets of ellipsoids in this context; and in [12] visual tendency of assessment is used to establish the possible presence of clusters of ellipsoids. For example, Fig. 1 is a plan view of the 54 node IBRL (Intel Berkeley Research Lab) WSN installed on March 1, 2004.

Fig. 2 is a set of ellipses generated by summarizing data collected at the 54 nodes of the IBRL. This data is available at the IBRL-Website: http://db.lcs.mit.edu/labdata/labdata.html. The data used in this paper were collected from 8:00 AM to 8:00 PM from the first 18 days of March in 2009. The data consist of $668,830$ ($t =$ temperature, $h =$ humidity) pairs, each pair labeled as being collected at one of the 54 nodes. The cardinality for each of the 54 data sets was not quite the same because some of the sets had a few values missing. We conditioned the data by first rounding up the $t$, $h$ values, and then removing duplicate vectors. Each duplicate was weighted by the number of times duplicated, resulting in a total of 8503 weighted pairs. Finally, the sample mean and sample covariance matrix of the vectors associated with each node resulted in the set of 54 ellipses as shown in Fig. 2. We call this data set E54. During this collection period, node 17 showed abnormal behavior, as manifested by the visually apparent “more horizontal” ellipse that stands in stark contrast to the other 53 ellipses. This is a real data example of a second order WSN anomaly as defined in [10–12].

The data in Fig. 2 offer a glimpse of our objectives in this article. First, we develop three measures of similarity for pairs of ellipsoids that (in principle) enable us to look for clusters of ellipsoids. Second, we image reordered versions of the dissimilarity matrices induced on ellipsoidal pairs by the similarity measures with the recursive iVAT algorithm [14,15]. The images
are used to assess whether or not the data do contain clusters, and if so, how many? Using estimates of $c$, the number of clusters in the data found by iVAT and a method based on the ordered eigenvalues of $D$, we find “optimal” clusters with the single linkage (SL, [16]) and CLODD [37] clustering algorithms. Ideally, anomalies in the sets of ellipsoids will not be grouped with (sets of) typical ellipsoids. We will illustrate this procedure with three numerical examples that use both real and artificial WSN data.

This paper is organized as follows. Section 2 reviews the essential algebra and geometry of ellipsoids. Sections 3, 4, and 5 contain definitions and proofs for three measures of similarity or dissimilarity on pairs of ellipses: compound normal and transformation energy similarities, and focal distance dissimilarity. Section 6 discusses the recursive iVAT algorithm for displaying reordered dissimilarity images. Section 7 presents iVAT images for the three data sets we use to illustrate our method. Section 8 discusses clustering in the dissimilarity data produced by each measure with the SL and CLODD algorithms. Section 9 offers our conclusions and some ideas for future research.

2. Similarity measures for pairs of ellipsoids

Let vectors $\mathbf{x}, \mathbf{m} \in \mathbb{R}^p$, and let $A \in \mathbb{R}^{p \times p}$ be positive definite. The quadratic form $Q(\mathbf{x}) = \mathbf{x}'A\mathbf{x}$ is also positive definite, and for fixed $\mathbf{m} \in \mathbb{R}^p$, the level set of $Q(\mathbf{x}-\mathbf{m}) = (\mathbf{x}-\mathbf{m})'^{t}A(\mathbf{x}-\mathbf{m}) = |(\mathbf{x}-\mathbf{m})|_A^2$, for scalar $t^2 > 0$, is

$$E(\mathbf{A}, \mathbf{m}, t) = \{ \mathbf{x} \in \mathbb{R}^p | |\mathbf{x}-\mathbf{m}|_A^2 = t^2 \}$$

Geometrically, $E(\mathbf{A}, \mathbf{m}, t)$ 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 $\mathbf{m}$. Sometimes $t$ is called the “effective radius” of $E(\mathbf{A}, \mathbf{m}, t)$. When $A = I_p$, $E(\mathbf{A}, \mathbf{m}, t)$ is the surface of a hypersphere with radius $t$. Henceforth, we may omit the prefix “hyper”, using ellipsoid and sphere for all cases, $p \geq 2$. Eq. (1) defines an infinite family $(E(\mathbf{A}, \mathbf{m}, t) : t > 0)$ of concentric ellipsoids parameterized in $t$. Each member of this family can be normalized, $|(\mathbf{x}-\mathbf{m})|_A^2 = 1$. This simply shifts attention to the scaling of $A(A(t^2-1))$ without loss of generality.

Suppose we have two ellipsoids in $p$-space, $E_i$ and $E_j$, that have effective radii $t_i$ and $t_j$, centers $[\text{means}] \mathbf{m}_i$ and $\mathbf{m}_j$, and $[\text{inverse of covariance}]$ matrices $A_i$ and $A_j$. First, normalize $E_i$ and $E_j$, $A_i/t_i^2 \rightarrow A_i$, and $A_j/t_j^2 \rightarrow A_j$. We want a measure of similarity between the two ellipsoids. Let $s(E_i, E_j)$ denote the similarity between $E_i$ and $E_j$. There are many definitions of similarity functions in the literature. For our purposes the following properties are used:

\begin{align}
(s1a) & \quad s(E_i, E_i) = 1 \quad \forall \ i \\
(s1b) & \quad s(E_i, E_j) = 1 \iff E_i = E_j \\
(s2) & \quad s(E_i, E_j) = s(E_j, E_i) \quad \forall \ i \neq j \\
(s3) & \quad s(E_i, E_j) > 0 \quad \forall \ i, j
\end{align}

Functions that satisfy (2a), (3), and (4) are weak similarity measures. Functions that satisfy (2b), (3), and (4) are strong similarity measures. Strong similarity measures are also weak, and are simply called similarity measures. Weak similarity measures correspond to pseudometric dissimilarities.

An ellipsoid $E(\mathbf{A}, \mathbf{m}, 1)$ is defined by its matrix $A$ and center $\mathbf{m}$. Any ellipsoid can be created by applying a scaling, rotation, and translation of a unit (hyper)sphere. The matrix $A$ scales and rotates the underlying space so that it maps the ellipsoid into a unit sphere (hence, $|(\mathbf{x}-\mathbf{m})|_A^2 = 1$ for every point $\mathbf{x}$ on the surface of the ellipsoid). Scaling a sphere is done by matrix multiplication with a scaling matrix $S$,

$$S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & s_p \end{bmatrix}$$

Fig. 1. The IBRL wireless sensor network.

Fig. 2. Ellipsoidal summaries for 54 nodes in the IBRL: node 17 is the visually anomalous ellipse.
where $s_k$ is the scaling factor for dimension $k$, $1 \leq k \leq p$. Rotation through an angle $\theta$ is accomplished by matrix multiplication with a unitary rotation matrix $R (R^T = -R^{-1})$, which in $p=2$ space has the form

$$R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

(6)

Any point $z$ in the unit sphere can be mapped to an ellipsoid via scaling, rotation, and shift:

$$z \rightarrow x = RSz + m, \quad |z|^2 \leq 1$$

(7)

Now we are ready to discuss measures of similarity on pairs of ellipsoids. We begin with a measure that combines the three elements of similarity for geometric structures of this type.

3. Compound similarity

We use subscripts 1 and 2 for a pair of ellipsoids. Our first measure of similarity is a compound measure, the product of three exponential factors that satisfy requirements (2a), (2b), (3), and (4) for strong similarity. This measure of similarity is built by considering the location, orientation, and shape of an ellipse. The geometric rationale and limit behavior of each factor are discussed next.

Location: Positional similarity for $(E_1, E_2)$ is a function of their mean separation, i.e., the distance $|m_1 - m_2|$ between their centers. Use any vector norm on $\mathbb{R}^p$ to define

$$g_1(E_1, E_2) = e^{-|m_1 - m_2|}$$

(8)

The function $g_1$ satisfies (3) and (4), and $E_1 = E_2 \Rightarrow e^{-|m_1 - m_2|} = 1$, but $m_1 = m_2$ does not imply that $E_1 = E_2$ (see the left view of Fig. 3). In the limit, $e^{-|m_1 - m_2|} \rightarrow 0$ as $|m_1 - m_2| \rightarrow \infty$. The center and right views in Fig. 3 illustrate the most common situation for sample-based ellipsoids, viz., that they have different effective radii, orientation, and centers.

Orientation: For $E_1 = (A_1, m_1, 1)$ and $E_2 = (A_2, m_2, 1)$ we define the orientation of each ellipse using the rotation matrices $R_1$ and $R_2$, the eigenvector matrices of $A_1$ and $A_2$. The angle of rotation between the two ellipsoids is found by projecting the basis vectors of $R_1$ onto the associated basis vectors of $R_2$, where the association is established by ranking each basis with the ordered eigenvalues of its matrix. The angles are found by calculating the vector of angles between associated eigenvectors as

$$\theta = \arccos(\text{diag}(R_1^T R_2))$$

(9)

Rotational similarity between $E_1$ and $E_2$ is assessed by measuring the set of angles between the associated eigenvector pairs. More specifically, let $\sin \theta = (\sin \theta_1, \ldots, \sin \theta_p)^T$ and define

$$g_2(E_1, E_2) = e^{-|\sin \theta|}$$

(10)

The function $g_2$ satisfies (3) and (4). When $E_1 = E_2$, their associated eigenvectors are parallel; hence $E_1 = E_2 \Rightarrow e^{-|\sin \theta|} = 1$. When the principal vectors are perpendicular, $g_2$ takes its maximum value, which is $e^{-|\sin \theta|} = e^{-\bar{p}}$ when the norm is Euclidean. But, again, $\theta = 0$ does not imply that $E_1 = E_2$ (e.g., see the left view of Fig. 3).

The final piece in our similarity puzzle is shape. The eigenstructure of $A$ is the key to assessing the shape of $E(A, m, 1)$. Invariants of $A$ to rotation and translation that might be useful in comparing the shapes of two ellipsoids include: $\text{tr}(A)$ (sum of the eigenvalues to total variance); $\text{det}(A)$ (product of the eigenvalues to generalized variance); scatter volume (a function of $\text{det}(A)$); and normalized scatter volume (a “decorrelated” function of $\text{det}(A)$). Each of these functions has some merit, but also a deficiency. For example, let $p = 3$ and suppose that $\{1, 1, 27\}$ and $\{3, 3, 3\}$ are the eigenvalues for $E_1$ and $E_2$. The determinants and scatter volumes for $E_1$ and $E_2$ are equal, but $E_1$ and $E_2$ have quite different shapes. $E_1$ is circular in cross section, but very elongated in its principal direction, while $E_2$ is a sphere. We turn to a different function of the eigenvalues to characterize shape similarity for a pair of ellipsoids.

Let $\alpha = \{x_1 \leq x_2 \leq \cdots \leq x_p\}$ and $\beta = \{\beta_1 \leq \beta_2 \leq \cdots \leq \beta_p\}$ be the ordered eigenvalues of $A_1$ and $A_2$, and recall that the semi-axial length of $E_1 = E(A_1, m_1; 1)$ from $m_1$ to its surface in the $k$th direction is $1/\sqrt{\alpha_k}$, and likewise for $E_2$. If the ellipsoids have the form $E = E(A, m, t)$, we adjust the eigenvalues to account for the transformation $A^T t^{2} A$ that normalizes them to the form $E(A, m, 1)$. Letting $\alpha^* = (1/\sqrt{\alpha_1}, \ldots, 1/\sqrt{\alpha_p})^T$ and $\beta^* = (1/\sqrt{\beta_1}, \ldots, 1/\sqrt{\beta_p})^T$, define, for any vector norm, $g_3(E_1, E_2) = e^{-|\alpha - \beta^*|}$ (11)

Function $g_3$ satisfies (3) and (4), and $E_1 = E_2 \Rightarrow g_3(E_1, E_2) = 1$, because the (ordered) eigenvalues of the two ellipsoids are equal, but again, the converse is not guaranteed. Switiching notation from (1, 2) to $(i, j)$, we define the product of $g_1, g_2$, and $g_3$ as the compound similarity between the ellipsoid pair $(E_i, E_j)$:

$$s_i(E_i, E_j) = e^{-|m_i - m_j|} e^{-|\sin \theta_i|} e^{-|\alpha_i - \beta^*_j|}$$

(12)

We summarize the properties of $s_i(E_i, E_j)$ in

**Proposition 1.** Let $E_i = E(A_i, m_i; t_i)$ and $E_j = E(A_j, m_j; t_j)$ be (hyper-)ellipsoids in $p$-space. Let the eigenvalue–eigenvector pairs for $E_i$ and $E_j$ be $\{0 \leq x_1 \leq x_2 \leq \cdots \leq x_p\} \leftrightarrow \{v_{1}, u_{2}, \ldots, u_{p}\}$ and $\{0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_p\} \leftrightarrow \{v_i, v_{2}, \ldots, v_p\}$. Let $\alpha = \{x_1 = x_{m_1}, x_2 = x_{m_2}, \ldots, x_p = x_{m_p}\}$, $\beta = \{\beta_1 = \beta_{m_1}, \beta_2 = \beta_{m_2}, \ldots, \beta_p = \beta_{m_p}\}$. $\alpha^* = (1/\sqrt{\alpha_1}, \ldots, 1/\sqrt{\alpha_p})^T$, and $\beta^* = (1/\sqrt{\beta_1}, \ldots, 1/\sqrt{\beta_p})^T$. Define $\theta$ as the angle between eigenvector pairs $u_i$ and $v_i$, $1 \leq i \leq p$. Then

$s_i(E_i, E_j) = 1 \Leftrightarrow E_i = E_j$ (13)

$s_i(E_i, E_j) = s_i(E_j, E_i) \quad \forall \ i \neq j$ (14)

$s_i(E_i, E_j) > 0 \quad \forall \ i, j$ (15)

![Fig. 3. Different effective radii and centers for two ellipsoids.](image-url)
4. Transformation energy similarity

Consider each ellipsoid as having its own space spanned by its eigenvector basis with origin at its center. We can construct a function that maps a point from one ellipsoid space to another via the common space between them. A point in the space of ellipsoid \(E_i\) can be mapped to the 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 \(m_i\). Within this common space the point can then be mapped into the space of \(E_j\) by shifting the point by \(m_j\), rotating by \(R_j\) and scaling by \(S_j\). The mapping is summarized as

\[
x_j = f(x_i | E_i, E_j) = S_j R_j (R_i^{-1} S_i^{-1}) x_i - m_i + m_j
\]

(17)

where \(E_i = E(R_i S_i^{-1} R_i^T, m_i; 1), E_j = E(R_j S_j^{-1} R_j^T, m_j; 1)\). The norm of this function is a measure of the "amount of energy expended" to make the transformation,

\[
\|f(E_i, E_j)\|_2 = \max_{z \in \mathbb{R}^p} \{\|f(z | E_i, E_j)\|_2\}
\]

(18)

Since (18) is not necessarily symmetric in the arguments of \(f\), \(\|f(E_i, E_j)\|_2 \neq \|f(E_j, E_i)\|_2\), we account for asymmetry by utilizing the norms in both directions. Thus, we define the transformation energy similarity function as

\[
s_{\|0\|}(E_i, E_j) = 1 / \max \left\{ \frac{\|f(z | E_i, E_j)\|_2}{\|z\|_2} \right\}
\]

(19)

Proposition 2. \(s_{\|0\|}(E_i, E_j)\) is a similarity function satisfying Eqs. (2), (3), and (4).

Proof. If \(E_i = E_j\), then \(s_{\|0\|}(E_i, E_j) = 1\), and Eqs. (2), (3), and (4) are satisfied. Now, if \(E_i \neq E_j\), then \(s_{\|0\|}(E_i, E_j) < 1\) and Eqs. (2), (3), and (4) are satisfied.

Lemma 1. If \(A\) is any invertible matrix, then \(\max \{|A|_2, |A^{-1}|_2\} \geq 1\).

Proof. Let \(X = A^{-1}\) has singular values \(0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_t\), then the spectral norm of \(X\) is \(\|X\| \geq \sigma_1\), and the spectral norm of \(X^{-1}\) is \(\|X^{-1}\| \geq \frac{1}{\sigma_t}\). Therefore, \(\|X\|_2 \cdot \|X^{-1}\|_2 \geq 1\).

Lemma 2. If \(A\) and \(B\) are invertible matrices and \(A, B\) have singular values \(0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_t\), then \(\|A - B\|_2 \leq \max \{\|A\|_2, |A|_2, |B|_2\}\).

Proof. Let \(X = A - B\) has singular values \(0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_t\), then \(\|X\|_2 \leq \sigma_t\), and \(\|X^{-1}\|_2 \leq \frac{1}{\sigma_1}\). Therefore, \(\|X\|_2 \cdot \|X^{-1}\|_2 \leq 1\).

Remark. If (18) must be solved many times, this problem can become computationally expensive. When the number of comparisons needed is large, we can replace (19) by an approximation to \(s_{\|0\|}(E_i, E_j)\), which avoids this difficulty by observing that \(\|g_i\|_2 \leq \|g_i + g_j\|_2 \leq \|g_i\|_2 + \|g_j\|_2\), which suggests taking \(\|f(z | E_i, E_j)\|_2 \approx \|f_1(z | E_i)\|_2 + \|f_2(z | E_j)\|_2\), where \(f_1(z | E_i) = M_i x_i\) and \(f_2(z | E_j) = d_0\) as...
in (17). The value of each term in this approximation is given by the largest singular value of $M_\phi$, $|f_1(E_i, E_j)| = |M_\phi x_i| = \sigma_{i,1}$ and likewise for the second term. Using the Euclidean norm for $d_{ij}$ then gives
\[
s_{dis}(E_i, E_j) = 1/\max\left\{ (\sigma_{i,1} + |d_{ij}|, \sigma_{j,1} + |d_{ij}|) \right\} \leq s_{rel}(E_i, E_j) \leq 1/\max\{\sigma_{i,1}, \sigma_{j,1}\}
\]  
with equality when $|d_{ij}| = |d_{ij}|$, or, equivalently, $m = m_i$.

The approximation function $s_{dis}(E_i, E_j)$ satisfies requirements (3) and (4), and it is clear that when $E_i = E_j$, $s_{dis}(E_i, E_j) = 1$. However, $s_{rel}(E_i, E_j)$ is an upper bound on $s_{dis}(E_i, E_j)$, so $s_{rel}(E_i, E_j) = 1$ does not guarantee that $s_{dis}(E_i, E_j) = 1$. We do not have an estimate for the tightness of this upper bound, but we did compute this approximation in all of our numerical experiments as a check on the exact value and, in most cases, the approximation is pretty close to its upper bound. So, when $n$ is large, the approximation $s_{dis} \approx s_{rel}$ is a good alternative to using (19) directly.

5. Focal similarity

Our third measure of similarity begins by recalling that every plane ellipse can be constructed by tracing the curve whose distance from a pair of foci $f_1$ and $f_2$ is some positive constant $c(t)$, which depends on the effective radius $t$.

This construction is shown for a two-dimensional ellipse in Fig. 4, with effective radius $t$ so that $p(t) + q(t) = c(t)$ for the ellipse $E(A, m, t)$. The foci always lie along the major axis of the ellipse, which is the linear span of the eigenvector of $A$ corresponding to the maximum eigenvalue. We denote the line segment with endpoints $f_1$ and $f_2$ by $f_{12}$ and call this the focal segment of $E(A, m, t)$.

If $|x_m| \leq |x_d|$ are the minimum and maximum eigenvalues of $A$ with corresponding orthogonal eigenvectors $\{u_m, u_d\}$, the locations of the foci are $f_{1,2} = m \pm \frac{1}{2} \sqrt{((z_m - z_d)/2z_mz_d)u_m}$. The focal similarity between $E_1$ and $E_2$ is defined as the average of a set of four distances. Each component is defined by a distance to one of the focal segments $e_{12}$ or $f_{12}$.

Let $d(x,y) = |x - y|$ be the Euclidean distance between vectors in $\mathbf{R}^d$. We have two focal segments, $e_{12}$ with endpoints $e_1$ and $e_2$, and $f_{12}$ with endpoints $f_1$ and $f_2$. We compute four default distances:
\[
\delta_1 = \text{min}(d(e_1, f_1), d(e_1, f_2)) \\
\delta_2 = \text{min}(d(e_2, f_1), d(e_2, f_2)) \\
\delta_3 = \text{min}(d(f_1, e_1), d(f_1, e_2)) \\
\delta_4 = \text{min}(d(f_2, e_1), d(f_2, e_2))
\]  

One or more of the values in (22)–(25) may be replaced using the following heuristic. For each foci “$f$”, if the orthogonal projection of “$T$” to the linear span of the opposing maximal eigenvector falls on the opposing focal segment then we replace the appropriate default distance by this distance; otherwise we find the minimum distance between “$f$” and the two opposing foci and use it in one of the Eqs. (22)–(25). In other words, if the orthogonal projection does not fall on the opposing focal segment it will not be considered in the calculation. We now define the focal distance between the ellipsoids $E_1$ and $E_2$ as the average of these four distances:
\[
d_{f\text{dis}}(E_1, E_2) = \text{dis}(e_{1,2}, f_{1,2}) = \frac{\delta_1 + \delta_2 + \delta_3 + \delta_4}{4}
\]  

Now let $E_1(A_1, m_1; t_1)$ and $E_2(A_2, m_2; t_2)$ be non-degenerate ellipsoids in $\mathbf{R}^d$. Let $\alpha = \{x_1 \leq x_2 \leq \cdots \leq x_p\}$ and $\beta = \{y_1 \leq y_2 \leq \cdots \leq y_q\}$ be the eigenvalues of $A_1$ and $A_2$. Adjust the eigenvalues to $\alpha' = (1/\sqrt{x_1}, \ldots, 1/\sqrt{x_p})$ and $\beta' = (1/\sqrt{y_1}, \ldots, 1/\sqrt{y_q})$. There are $p(p-1)/2$ focal segments for the two-dimensional ellipses spanned by each pair of eigenvectors of $E_1$ and $E_2$. Thus, there are $(p-1)$ “ordered” focal distances between pairs of focal segments of the two ellipsoids. We define the generalized focal distance between $E_1$ and $E_2$ as the average of the plane focal distances:
\[
d_{f\text{dis}}(E_1, E_2) = \frac{1}{p-1} \sum_{i=1}^{p-1} \text{dis}(e_{ij}, f_{ij+1})
\]  

Note that when $p = 2$, (26) reduces to (25). The generalized focal distance $d_{f\text{dis}}$ is almost a metric on pairs of ellipsoids. We say almost because for the simplest case, $p = 2$, it can happen that $d_{f\text{dis}}(E_1, E_2) = 0$ but $E_1 \neq E_2$. Recall that a pseudometric on a set $X$ is a non-negative real-valued function $d: X \times X \to [0, \infty)$ such that, for $x, y, z \in X$,

- pm1. $d(x,x) = 0$
- pm2. $d(x,y) = d(y,x)$ (symmetry)
- pm3. $d(x,z) \leq d(x,y) + d(y,z)$ (subadditivity/triangle inequality)

Proposition 3. $d_{f\text{dis}}$ in (25) is a pseudometric on pairs of $p$-dimensional ellipsoids.

Proof. Since each of the four factors in (25) is a metric, their average is symmetric, and it is easy to check that it also satisfies the triangle inequality, so pm2 and pm3 hold. Since each factor of (25) is a metric, pm1 holds for any ellipse, $d_{f\text{dis}}(E_1, E_1) = 0$. To see that $d_{f\text{dis}}$ is not positive definite, consider the construction of an ellipse $E_1$ as shown in Fig. 4, where, for $t = t_1, r(t_1) + q(t_1) = c(t_1)$. Now construct a second ellipse $E_2$ with the same focal segment but a different radius $t_2$, so that for $E_2, r(t_2) + q(t_2) = c(t_2)$. This pair of ellipses has the same focal segment, so $d_{f\text{dis}}(E_1, E_2) = 0$, but since $c(t_1) \neq c(t_2), E_1 \neq E_2$. Hence, $d_{f\text{dis}}$ is a pseudometric. □

Corollary 1. Let $D_{f\text{dis}}$ be a collection of $n^2$ normalized values of $d_{f\text{dis}}$ in (25) on $n$ ellipsoids, say $[D_{f\text{dis}}] = [d_{f\text{dis}}(E_i, E_j)]_{i,j \in \{1, \ldots, n\}}$. Define $s_{f\text{dis}} = 1 - D_{f\text{dis}}$. Since $d_{f\text{dis}}$ is a pseudometric, $s_{f\text{dis}}(E_i, E_j)$ satisfies Eqs. (3) and (4). In view of Proposition 3, we know that $s_{f\text{dis}}(E_i, E_i) = 1$ does not imply that $E_i = E_j$, so $s_{f\text{dis}}$ is not a strong similarity measure. But $s_{f\text{dis}}$ satisfies (2a), so it is a weak similarity measure. We call $s_{f\text{dis}}$ the focal similarity between $E_i$ and $E_j$.

Since the generalized focal distance is the average of $(p-1)$ focal distances that all satisfy Corollary 1, the generalized focal distance also satisfies the statement made in Corollary 1.

Corollary 2. Let $D_{f\text{dis}}$ be a collection of $n^2$ normalized values of $d_{f\text{dis}}$ in (25) on $n$ ellipsoids, say $[D_{f\text{dis}}] = [d_{f\text{dis}}(E_i, E_j)]_{i,j \in \{1, \ldots, n\}}$. Define $s_{f\text{dis}} = 1 - D_{f\text{dis}}$. Since $d_{f\text{dis}}$ is a pseudometric, $s_{f\text{dis}}(E_i, E_j)$ satisfies Eqs. (3) and (4). In view of Proposition 3, we know that $s_{f\text{dis}}(E_i, E_i) = 1$ does not imply that $E_i = E_j$, so $s_{f\text{dis}}$ is not a strong similarity measure. But $s_{f\text{dis}}$ does satisfy (2a), so it is a weak similarity measure. We call $s_{f\text{dis}}$ the generalized focal similarity between $E_i$ and $E_j$.

Fig. 5 depicts a few cases for the four distances comprising (25), where for ease of interpretation, we show focal segments and
Some examples of ellipsoidal similarity.

and

with five simple examples. Normalization of distances without notation. The solid green lines represent endpoint projections that land on opposing focal segments (so red lines may represent “replacement distances” for each of (22)–(25)).

Fig. 5 (a) has four equal distances (or put another way, one distinct distance); (b) and (c) have two distinct distances; (d) and (e) have three distinct distances; and (f) has four distinct distances. There are other cases, but these suffice to explain the concept. While (25) appears to be quite different from the first two similarity measures, the focal distance does account for location, shape, and orientation.

Example 1. To conclude this section we compare \( s_{cn}, s_{fs}, \) and \( s_{te} \) with five simple examples. Normalization of \( E_i \) and \( E_j \) via \( A_i/t_i^2 \rightarrow A_i \) and \( A_j/t_j^2 \rightarrow A_j \) is always done for synthetically generated ellipsoids. For sample-based ellipsoids, \( t \) is not known, but is not needed for a complete analysis. We used Matlab’s constrained nonlinear optimization, fmincon, which is a Trust-Region Reflective Algorithm discussed in [36], when computing the transformation energy similarity function at (18).

Table 1 contains five pairs of ellipses, and shows beneath them the value of each similarity coefficient for each pair. Look at these five examples, decide for yourself, which pairs of ellipsoids are “most different” and “most similar”, and then compare your assessment to the ones rendered by the numerical indices, whose maximums and minimums are highlighted in boldface type.

Fig. 6 plots the values of the three coefficients for cases \( A–E \) shown in Table 1. On the horizontal axis the ticked labels are \( A=1, B=2, C=3, D=4, E=5 \). All three measures agree that case \( D \) exhibits the least similar pair, and most observers would agree with this assessment. But while the three indices all agree that case \( E \) is the most similar pair, our guess is that most observers would disagree with this result, and instead choose pair \( C \) as the most similar. We think the explanation for the apparent discrepancy between human observation and mathematical assessment lies with the properties of the ellipses. Case \( E \) features two ellipses with the same means, but different shapes and orientations, while Case \( C \) has a pair of ellipses with the same shape and orientation, but different means. This suggests that the three models weight central tendency more heavily than humans do. The important point is that all three measures agree—here. But we shall see below that they are quite different on more complex sets of ellipses.

There may be (perhaps often will be) many ellipse pairs, which have very nearly the same or even equal similarity values that have very different spatial configurations using any of these measures, even when \( p=2 \). (We know this to be the case for the focal similarity.) To see that it can also happen for the other measures, consider \( s_{cn}=abc \), the product of three numbers \( a, b, \) and \( c \) all of which lie in \([0,1]\). Suppose \( s_{cn}=0.5 \) and \( a=1 \). Then \( bc \) must equal 0.5 with \( b \) and \( c \) in \([0,1]\), but are otherwise unconstrained, so there are many ellipse pairs – all different from one another – that result in this single value. While this observation seems to deflate the value of measuring similarity by any of these functions, in practice we will rarely, if ever, encounter a “tie”, wherein two quite different ellipse pairs yield the same value of any of these measures.

\[
\begin{array}{c|c|c|c|c|c}
A & B & C & D & E \\
\hline
s_{cn}(E_i, E_j) & 0.018 & 0.001 & 0.024 & 0.000 & 0.031 \\
s_{fs}(E_i, E_j) & 0.148 & 0.122 & 0.16 & 0.07 & 0.355 \\
s_{te}(E_i, E_j) & 0.460 & 0.225 & 0.469 & 0.00 & 0.737 \\
\end{array}
\]

Table 1

Some examples of ellipsoidal similarity.

Fig. 5. Several cases of different focal distance components of \( d_{ct} \) at (25).

Fig. 6. Similarity coefficients for the five ellipse pairs in Table 1.
6. Tendency assessment with VAT and iVAT

Our aim is to use similarity and dissimilarity measures and their iVAT images to find clusters in sets of ellipsoids. Before considering this specific problem, we introduce some concepts from clustering theory that are needed to proceed with our objectives. Clustering is the problem of partitioning a set of unlabeled objects \( O = \{ o_1, \ldots, o_n \} \) into groups of similar objects [17,18–20]. The field comprises three canonical problems: (CPs). (CP1) is assessment: prior to finding any clusters, we ask—are there clusters in \( O \); if so, how many? (CP2) is clustering: what are the clusters in \( O \)? (CP3) is validation: are the found clusters “good” in any useful or meaningful way? When \( o_0 \in O \) is represented by \( x_0 \in \mathbb{R}^d \), \( x_0 \) is an object data representation of \( O \). The \( k \)th component of \( x_0 \) is the \( k \)th feature (e.g., height, hair color, number of legs, etc.) of \( o_0 \). When relational values between pairs of objects are available, we have relational data. Any relation \( \rho \) on \( O \times O \) is representable by a square matrix \( R_{n \times n} = [ r_{ij} ] \), where \( r_{ij} = \rho(o_i, o_j) \) is the relationship between \( o_i \) and \( o_j \). \( 1 \leq i, j \leq n \). \( X \) can be converted into dissimilarity data \( D_{n \times n} = [ d_{ij} ] = [ \| x_i - x_j \| ] \) using any norm on \( \mathbb{R}^d \). Similarity data \( S_{n \times n} \) are always convertible to dissimilarity data \( D \) using simple transformations such as \( D = [1-S] \). This is the method we use for our examples; see [21] for other methods.

Our approach to clustering ellipsoids is based on visual assessment. Visual methods vary greatly in complexity and computational cost from simple techniques such as histograms and box-and-whisker plots to those implemented in larger interactive software systems such as the IBM Open Visualization Data ExplorerTM (http://www.research.ibm.com/dx/). A classic reference for the principles of effective visual display is [22]. Many useful data mining and visualization methods are covered in [23,24]; a nice reference for the practical application of some of the known techniques, and [25] contains some classic approaches for visual analysis in multidimensional object vector data.

The visual representation of structure in unlabeled dissimilarity data has a long history. Tryon [26] paved the way for this branch of clustering when he introduced visual assessment and aggregation of hand-rendered profile graphs for all three problems in 1939. Cattell [27] first depicted clusters in pairwise dissimilarity data about the objects in \( O \) as an \( n \times n \) image. Important advances in visual clustering include Sneath [28], Floodgate and Hayes [29], Ling [30], and the VAT/iVAT/coVAT/ iVAT papers [31–35]. The common denominator in all these methods is 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. We use recursive iVAT to produce RDIs in the sequel.

VAT reordered an input dissimilarity matrix \( D \to D^* \) and displays a grayscale image \( I(D^*) \) whose \( j \)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. Off the diagonal, the 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 “small” values, whose diagonal is superimposed on the diagonal of the image matrix. The iVAT method [34] transforms \( D \to D^* \) using a path-based distance and then VAT is applied to \( D^* \) to get \( D^{**} \), resulting in an iVAT image \( I(D^{**}) \).

Construcing the iVAT matrix \( D^{**} \) as in [34] can be computationally expensive (\( O(n^3) \)). The recursive computation of \( D^{**} \) given here and in [35] does not alter the VAT order of \( D^* \) and is \( O(n^2) \). Recursive iVAT [35] builds the matrix \( D^{**} \) more efficiently than iVAT by first applying VAT to \((D) \to D^* \), and then recursively using \( D^* \) to build \( D^{**} \). The main limitation of iVAT is size; hardware and software limits \( D_{n \times n} \) to about \( n = O(10^4) \), but for our application, VAT is well within its working capacities. 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 result of applying VAT to \( D_{n \times n} \) is \( D_n^{**} \); and the displayed output is the VAT image \( I(D_{n \times n}) \). The VAT reordering for \( D^* \) is stored in array \( P = (P(1), \ldots, P(n)) \).

VAT/iVAT: visual assessment of tendency [31,34,35]

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 \( (i, j) = \arg \max \{ D_{pq} \} \); set \( P(1) = i \); \( P(2) = j \).

Step 2: For \( t = 2, \ldots, n \); select \( (i, j) = \arg \min \{ D_{pq} \} \); \( P(t) = j \);

\( I = \{ i, j \} \); and \( J = K - \{ i, j \} \).

Step 3: Form the ordered dissimilarity matrices \( I(VAT): D^*; \ y^* = \min(D^*) \) for \( 1 \leq i, j \leq n \); \( I(VAT): D^{**} = [0]_{n \times n} \) for \( r = 2; \ldots; n \) do

\[ j = \arg \min \{ D^*_k \} \]

\[ k = 1, \ldots, r-1 \]

\[ D^{**} = D^*_r, \quad c = j \]

\[ D^{**} = \max(D^*_k, D^{**}_{c,j}), \quad c = 1, \ldots, r-1, \ c \neq j \]

For \( 2 \leq j \leq n \) do

Step 4: Display \( I(D^*) \) and \( I(D^{**}) \), scaled so that \( \max \{ d_{ij}^* \}, \max \{ d_{ij}^{**} \} = \text{white} \) and \( 0 = \text{black} \).

Fig. 7(a) is a scatterplot of a “boxes and stripes” cursor data set similar to that used in [34,35]. This two-dimensional data has two round clusters, two rectangular clusters, and one elongated stripe cluster. Most would agree that there are \( c = 5 \) clusters in this data. These object data were converted to \( D = [d_{ij}] = [\| x_i - x_j \|] \) using the Euclidean norm.

The \( c = 5 \) visually apparent clusters in Fig. 7(c) are quite clearly suggested by the 5 distinct diagonal blocks in Fig. 7(c), \( I(D^*) \), which is the VAT RDI of the data. Compare this to view (b), which is the VAT image \( I(D^*) \) of these data. \( I(D^*) \) presents some evidence supporting the view that this data contains four clusters (the four clouds are seen in both (b) and (c)), but it misses the stripe cluster. Interpretation of substructure in the data suggested by \( I(D^{**}) \) is a significant improvement. Next we turn to the use of iVAT for assessment of clusters in sets of ellipsoids.

7. Tendency assessment for sets of ellipsoids

Let \( E \) denote \( n \) ellipsoids in \( p \)-space, \( E = \{ E_1, E_2, \ldots, E_n \} \). For \( E_i, E_j \in E \times E \), compute \( s_{ij} = \delta(E_i, E_j) \) with any of our three measures of similarity, and array these \( n^2 \) values as the \( n \times n \) similarity relation matrix \( S = [s_{ij}] \). The transformation \( D = [d_{ij}] = [1 - s_{ij}] \) yields a dissimilarity relation on \( E \times E \). (Actually, we need not do this for the focal measure, as it is, by definition, a dissimilarity measure already.) Applying the iVAT algorithm to \( D \) will yield an RDI that can be used to assess clustering tendencies of the ellipsoids in \( E \times E \). We illustrate the use of iVAT images for this purpose with three data sets: E30, E40, and E54. The first two data sets are synthetic (i.e., we constructed these sets of ellipses to form the clusters they appear to have). E54 is the real IRBL data set shown in Fig. 2. The synthetic ellipses are generated with \( t = 1 \), so normalization is not needed, and the sample-based ellipses cannot be normalized because \( t \) is unknown.
Example 2. Fig. 8(a) is the data set $E_{30}$, which has, by design, a set of three well-separated clusters, 10 ellipses each. Failure of one of the measures, algorithms or validity indicators on this idealized data will raise a caution flag about its utility for this problem.

The iVAT image in Fig. 8(c) has 3 primary, very dark blocks along its diagonal, each block of size $10 \times 10$, strongly suggesting that there are $c = 3$ clusters in $E_{30}$. The images in (b) and (c) are much less conclusive. The energy image at (b) has three primary dark blocks, but clearly indicates substructure within each primary cluster. Reading down the diagonal from the top, there are 2, 5, and 3 substructural blocks in (b). The compound image is even less conclusive: its primary structure suggests perhaps 17 clusters. We will be alert for further evidence that these latter two measures are less reliable than focal distance.

Example 3. This example is based on the set $E_{40}$ is shown in Fig. 9a. This is a much more challenging test for our measures. The lower left cluster contains 30 ellipses roughly centered at (12, 15). 15 of these ellipses have a horizontal major axis while the other
15 have a vertical major axis. The other cluster, roughly centered at (25, 30), has 10 ellipses that are fairly circular. Many observers see three clusters because of the orientation of the two sets of 15, but other observers see primary cluster structure at $c=2$ because the set of 30 have a strong and similar central tendency. Does the visually apparent structure of $E_{40}$ suggest $c=2$? Or is $c=3$ a better choice?

All three iVAT images suggest that we take $c=2$ for the primary structure in the data. The 30 × 30 block in the focal distance image (9c) is (albeit faintly) subdivisible into 2 15 × 15 blocks—that is, $c=3$ is suggested as a secondary interpretation of this data by focal distance. The other two images are quite fragmented beyond their primary implication that $c=2$. We think that $c=2$ and $c=3$ are both acceptable interpretations of $E_{40}$.

**Example 4.** Fig. 10(a) is a repeat of Fig. 2. Recall that these 54 ellipses represent data collected at the 54 nodes in the IBRL network, and that the “horizontal” ellipse that is visually apparent in this data is atypical node 17. Looking further at these ellipses, you may notice that the ones with axes tilted at roughly $-45^\circ$ have means that are considerably displaced along this direction, and quite a few of them are much shorter than others. So, many of these ellipses may be more dissimilar to their similarly oriented neighbors than to ellipse 17. Nonetheless, the preferred value for this real data is $c=2$, since ellipse 17 is a known second order anomaly in this WSN. Thus, we hope to deduce from visual assessment that these data contain $c=2$ clusters of ellipses.

What do the iVAT images tell us about $E_{54}$? We see exactly the structure we are hoping for in Fig. 10(c)—node 17 corresponds to the single dark pixel in the bottom right corner of the focal distance image, the remaining 53 ellipses being the $53 \times 53$ block. Moreover, this image reveals a number of substructures within the primary cluster of 53 ellipses, as is borne out by visual examination of the data set. Neither the transformation energy nor the compound normal images in views 10(b) and (d) present a very clear picture of structure in $E_{54}$. The overall results of Examples 2, 3, and 4 suggest that focal similarity $s_{fs}$ provides the “best” iVAT images: they fully agree with the visually apparent clusters in all three data sets. Now we turn to detecting the clusters suggested by these images.

### 8. Finding clusters in sets of ellipsoids

Looking for clusters in $E$ raises two questions. First, before clustering, we must ask how many clusters to look for? Second, after clustering, how much credence shall we put on the “optimal” partition of the data? The iVAT images of Section 7 offer visual suggestions for value(s) of $c$ in each of our three test sets prior to clustering. There are many, many other ways to estimate $c$ prior to clustering. The second pre-clustering approach tested here is based on the eigenvalues of $D$. Ferenc proved in [39] that a nonsymmetric $n \times n$ matrix consisting of $c$ (dark) blocks has $c$ large eigenvalues of order $c$ while the other characteristic values
remain of order $\sqrt{n}$ as $n$ tends to infinity. Fallah et al. [40] recently showed that Ferenc’s theorem could be used as a pre-clustering assessment method to help when choosing the best SL clusters, by looking for a “big jump” in a plot of the square roots of the ordered eigenvalues (OEVs) of $D$. (Square roots just improve the visual interpretation of where the big jump occurs.) Note that the theory underlying this strategy is not tied to any clustering algorithm.

Fig. 11 shows plots of the first 15 OEVs of each $D$ for each of our three data sets.

According to the theory, we expect to find $c$ big eigenvalues, a knee at the big jump, and then a leveling off in the graph for the smaller eigenvalues. The focal similarity graphs have pretty well defined breaks that suggest $c=3$ for all three data sets. It is much harder to see the indicated choices from the graphs of the other two measures, as they are much flatter. The most inconclusive graph is the compound graph for $E54$, which does not really suggest that there are $c$ dominant eigenvalues.

Table 2 lists the pre-clustering best guesses for $c$ for each of our three test sets using the iVAT and OEV methods. Bold values are...
incorrect estimates. The column headed “c*” shows the visually apparent correct number of clusters in each data set. The two methods agree with c* in 3/9 cases. iVAT indicates the correct c* for all three data sets using either focal distance or the energy similarity. The compound measure does not provide good estimates of c* using either the visual (iVAT) or analytic (OEVs) methods.

There are also many ways to look for the clusters in D suggested by iVAT images. Here we discuss two related approaches. One is the well-known single linkage (SL) algorithm [16] and the other is the CLODD (clustering in ordered dissimilarity data) algorithm [37]. CLODD processes dark block images made by reordering D, while SL processes D directly. These two algorithms are known to produce the same clusters from D under some – but not all – circumstances. Interested readers may consult [38] for a discussion of the theory underlying this relationship.

The classic method for choosing the “optimal” number of clusters found by SL is to look for a “big jump” in the graph of SL merger distances, and back up one step. The heuristic justifying this procedure is that the biggest merger distance indicates the maximum resistance for merger, so the clusters just ahead of this merger are the most desirable. Fig. 12 shows the SL merger distance graphs for c=15 to c=2 (merging, we plot higher c’s to the left). Table 3 shows the values of c identified as optimal by this method, headed as SLmd (single linkage merger distance). CLODD has an internal measure of validity—viz., its objective function values. Each validation method has successes and failures (as do all validation indices). The failures are again shown as bold values. We see that the

Table 3

<table>
<thead>
<tr>
<th>Data</th>
<th>c*</th>
<th>Focal</th>
<th>Energy</th>
<th>Compound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SLmd</td>
<td>CLODD</td>
<td>SLmd</td>
<td>CLODD</td>
</tr>
<tr>
<td>E30</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>E40</td>
<td>2/3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>E54</td>
<td>2</td>
<td>2</td>
<td>10</td>
<td>51</td>
</tr>
</tbody>
</table>

Fig. 12. Plots of SL merger distances for each D and each data set.

Fig. 13. CLODD objective function and partition extracted from the energy transform iVAT image of E30.
preferred SL and CLODD partitions agree with \( c^* \) in all three tries only for the focal distance measure. Most importantly, the optimal SL partitions of all three data sets agree with visual assessment of the input data when focal similarity is used to build the input dissimilarity data, and this method is the only one that recovers the apparently correct answer for the WSN data \( E54 \).

Fig. 14. Optimal single linkage partitions of the data sets based on the focal similarity measure.
CLODD processes reordered images directly. CLODD is run on $D$ for $c=2, 3, \ldots, c_{\text{max}}$, and the optimal partition of $c$ is taken as the one that maximizes the CLODD objective function. Table 3 shows the values that CLODD chooses as optimal based on its objective function. CLODD recovers the desired clusters in 3/9 tries, so it seems less effective than SL for this application. Bear in mind, however, that CLODD extracts clusters from the iVAT image $I(D')$ by identifying the best fit of its objective function to the dark block structure in the image. Thus, when an iVAT image seems to have imbedded substructure in its primary dark blocks, CLODD will extract a partition corresponding to this visual structure. Fig. 13 is an example of CLODD applied to the iVAT image of $E30$ based on the energy transform similarity. The left view shows a plot of the CLODD objective function, which maximizes at $c=6$. The right view is the optimal partition extracted by CLODD at this value of $c$. The dotted red lines indicate the boundaries of the partition. We point out that, unlike SL, which has no “tuning parameters”, CLODD has several user-defined parameters that can be adjusted to alter model performance.

Finally, Fig. 14 displays the three data sets and the optimal single linkage partitions of each. As you can see, there seems to be perfect agreement between the clusters obtained by single linkage and the visually apparent clusters in the data. The corresponding partitions based on the energy transform were the same as those for $E30$ and $E40$, but the optimal solution at $c=2$ for $E54$ identified the smallest of the 53 aligned ellipses in the data as the singleton. Optimal SL partitions based on the compound similarity matched the ones in Fig. 11 only on $E30$. Fig. 14 also displays the single linkage partition at $c=2$ for $E40$. This is the primary partition that is suggested by the iVAT image of $E40$ in view 9(c). This partition is also the one found at $c=2$ by applying single linkage to the other two matrices of dissimilarity, so this interpretation of the data is compelling. The results shown in Fig. 14 corroborate our earlier assertion that focal distance is the most reliable of the three measures. Using the focal distance dissimilarity matrix results in the correct partitioning of the input data by single linkage clustering in all three tests.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Build 3 measures</th>
<th>Pre-clustering assessment (to estimate $c$)</th>
<th>Find partitions $U$ of ellipse data</th>
<th>Post-clustering validation (choose best $U$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic data $E30$ and $E40$</td>
<td>Focal distance</td>
<td>OEVs of $D$</td>
<td>Single linkage applied to $D$</td>
<td>Biggest jump in SL merger distance</td>
</tr>
<tr>
<td>Real IBRL data $E54$</td>
<td>Transformation energy</td>
<td>$iVAT$ image $I(D')$</td>
<td>CLODD applied to $I(D')$</td>
<td>Maximum value of CLODD function</td>
</tr>
<tr>
<td>Compound normalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4

The procedures used in this paper.

References


Masud Moshtaghi received his B.Sc. degree in 2006 in computer science, and his M.S. in software engineering in 2008 from the University of Tehran. He has been with the University of Melbourne from March 2009. His research interests include pattern recognition, artificial intelligence for network security, data mining, and wireless sensor networks.

Timothy C. Havens received his M.S. degree in electrical engineering from Michigan Tech University in 2000. After that, he was employed at MIT Lincoln Laboratory where he specialized in the simulation and modelling of directed energy and global positioning systems. In 2006, he began work on his Ph.D. degree in electrical and computer engineering at the University of Missouri. His interests include clustering in real data and ontologies, fuzzy logic, and bioinformatics but, by night, he is a jazz bassist.


Laurence Park received his B.E. (Hons.) and B.Sc. degrees from the University of Melbourne, Australia in 2000 and Ph.D. degree from the University of Melbourne in 2004. He joined the Computer Science Department at the University of Melbourne as a Research Fellow in 2004, and was promoted to Senior Research Fellow in 2008. Laurence joined the School of Computing and Mathematics at the University of Western Sydney as a Lecturer in Computational Mathematics and Statistics in 2009, where he is currently investigating methods of large scale data mining and machine learning. During this time, Laurence has been made an Honorary Senior Fellow of the University of Melbourne.

Christopher Leckie is an Associate Professor and Deputy-Head of the Department of Computer Science and Software Engineering at the University of Melbourne in Australia. A/Prof. Chris Leckie has over two decades of research experience in artificial intelligence (AI), especially for problems in telecommunication networking, such as data mining. He has made significant theoretical and practical contributions in efficiently analyzing large volumes of data in resource-constrained environments, such as wireless sensor networks.

Sutharshan Rajasegarar received his B.Sc. Engineering degree in Electronic and Telecommunication Engineering (with first class honours) in 2002, from the University of Moratuwa, Sri Lanka, and his Ph.D. in 2009 from the University of Melbourne, Australia. He is currently a Research Fellow with the Department of Electrical and Electronic Engineering, The University of Melbourne, Australia. His research interests include wireless sensor networks, anomaly/detector detection, machine learning, pattern recognition, signal processing, and wireless communication.

James M. Keller received his Ph.D. in Mathematics in 1978. He holds the University of Missouri Curators’ Professorship in the Electrical and Computer Engineering and Computer Science Departments on the Columbia campus. He is also the R. L. Tatum Professor in the College of Engineering. His research interests center on computational intelligence: fuzzy set theory and fuzzy logic, neural networks, and evolutionary computation with a focus on problems in computer vision, pattern recognition, and information fusion including bioinformatics, spatial reasoning in robotics, geospatial intelligence, sensor and information analysis in technology for eldercare, and landmine detection. His industrial and government funding sources include the Electronics and Space Corporation, Union Electric, Geo-Centers, National Science Foundation, the Administration on Aging, The National Institutes of Health, NASA/JSC, the Air Force Office of Scientific Research, the Army Research Office, the Office of Naval Research, the National Geospatial Intelligence Agency, the Leonard Wood Institute, and the Army Night Vision and Electronic Sensors Directorate. Professor Keller has coauthored over 350 technical publications. Jim is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) for whom he has presented live and video tutorials on fuzzy logic in computer vision, is an International Fuzzy Systems Association (IFSA) Fellow, an IEEE Computational Intelligence Society Distinguished Lecturer, a national
lecturer for the Association for Computing Machinery (ACM) from 1993 to 2007, and a past President of the North American Fuzzy Information Processing Society (NAFIPS). He received the 2007 Fuzzy Systems Pioneer Award from the IEEE Computational Intelligence Society. He finished a full six year term as Editor-in-Chief of the IEEE Transactions on Fuzzy Systems, is an Associate Editor of the International Journal of Approximate Reasoning, and is on the editorial board of Pattern Analysis and Applications, Fuzzy Sets and Systems, International Journal of Fuzzy Systems, and the Journal of Intelligent and Fuzzy Systems. Jim was the Vice President for Publications of the IEEE Computational Intelligence Society from 2005 to 2008, and is currently an elected Adcom member. He was the conference chair of the 1991 NAFIPS Workshop, program co-chair of the 1996 NAFIPS meeting, program co-chair of the 1997 IEEE International Conference on Neural Networks, and the program chair of the 1998 IEEE International Conference on Fuzzy Systems. He was the general chair for the 2003 IEEE International Conference on Fuzzy Systems.

Marimuthu Palaniswami received his M.E. from the Indian Institute of Science, India, M.Eng.Sc. from the University of Melbourne and Ph.D. from the University of Newcastle, Australia before rejoining the University of Melbourne. He has published over 340 refereed research papers. He currently leads one of the largest funded ARC Research Network on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) programme—that is structured it run as a network centre of excellence with complementary funding for fundamental research, test beds, international linkages and industry linkages. His leadership includes as an external reviewer to an international research centre, a selection panel member for senior appointments/promotions, grants panel member for NSF, advisory board member for European FP6 grant centre, steering committee member for NCRIS GBROOS and SEMAT, and board member for IT and SCADA companies. His research interests include SVMs, Sensors and Sensor Networks, Machine Learning, Neural Network, Pattern Recognition, Signal Processing and Control.