Title

A normalized-cut algorithm for hierarchical vector field data segmentation

Authors

Chen, Jiann-Liang
Bai, Zhaojun
Hamann, Bernd
et al.

Publication Date

2003-01-13

A Normalized-cut Algorithm for Hierarchical Vector Field Data Segmentation

Jiann-Liang Chen[a], Zhaojun Bai[b], Bernd Hamann[a,c] and Terry J. Ligocki[c]

[a]Center for Image Processing and Integrated Computing (CIPIC), Department of Computer Science, University of California, One Shields Avenue, Davis, CA 95616
[b]Department of Computer Science, University of California, Davis, CA 95616
[c]National Energy Research Scientific Computing Center (NESRC), Lawrence Berkeley National Laboratory, Berkeley, CA 94720

ABSTRACT

In the context of vector field data visualization, it is often desirable to construct a hierarchical data representation. One possibility to construct a hierarchy is based on clustering vectors using certain similarity criteria. We combine two fundamental approaches to cluster vectors and construct hierarchical vector field representations. For clustering, a locally constructed linear least-squares approximation is incorporated into a similarity measure that considers both Euclidean distance between point pairs (for which dependent vector data are given) and difference in vector values. A modified normalized cut (NC) method is used to obtain a near-optimal clustering of a given discrete vector field data set. To obtain a hierarchical representation, the NC method is applied recursively after the construction of coarse-level clusters. We have applied our NC-based segmentation method to simple, analytically defined vector fields as well as discrete vector field data generated by turbulent flow simulation. Our test results indicate that our proposed adaptation of the original NC method is a promising method as it leads to segmentation results that capture the qualitative and topological nature of vector field data.

Keywords: Normalized Cut, Vector Field, Segmentation, Clustering, Least Squares, Visualization, Approximation

1. INTRODUCTION

Visualizing vector field data is challenging due to the size and the complexity of the data sets produced by today's complex numerical simulations. Many hierarchy construction techniques have been proposed to make it possible to deal with the complexity and size of vector field data sets. Different criteria for combining and approximating vectors have been introduced to achieve the goal of faithfully preserving the topology of an original vector field.[2, 3, 5, 6, 11, 12] Utilizing the fact that in a vector field every vector can be associated with a certain type of critical point, a method based on an image segmentation algorithm, called "normalized cut" (NC)[9] was proposed.[1] The goal of our method is to cluster vectors together that are associated with the same critical point, and therefore leads to a "natural cluster." Under the concept of hierarchical clustering, a cluster on the finest level of representation is defined as the set of all vectors that can be expressed or related to the same critical point. The size of a cluster, in a geometrical sense, represents the "influence region" of the associated critical point.

Further author information: (Send correspondence to J.L.C.)
J.L.C.: E-mail: chenjj@cs.ucdavis.edu
Z. B.: E-mail: bai@cs.ucdavis.edu, Department of Computer Science, University of California, One Shields Avenue, Davis, CA 95616
B. H.: E-mail: hamann@cs.ucdavis.edu Center for Image Processing and Integrated Computing, Department of Computer Science, University of California, One Shields Avenue, Davis, CA 95616
T.J.L.: E-mail: tjligocki@lbl.gov, National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, Mail Stop 50A-1148, One Cyclotron Road, Berkeley, CA 94720

We have extended the original NC method and demonstrated its ability to extract global and dominant structures in a vector field by identifying types and sizes of clusters. The ability of this method to approximate vectors within a cluster fairly accurately is described and demonstrated in the following sections. A brief description of the concepts of clustering, similarity measure, and an extended NC method is presented in Section 2. Cluster refinement strategies and hierarchy construction techniques for the extended NC method are discussed in Section 3. The applications of our methods to analytically defined and real data sets are presented in Section 4. Concluding remarks and a discussion of future work are presented in Section 5.

2. THE NC METHOD AND VECTOR FIELD SEGMENTATION

2.1. Similarity Measure and Clusters

A linear 2D vector field can be expressed as

$$\begin{aligned} \mathbf{v}(x,y) &= \begin{bmatrix} v_1(x,y) \\ v_2(x,y) \end{bmatrix} = \begin{bmatrix} a_{1,1}x + a_{1,2}y + b_1 \\ a_{2,1}x + a_{2,2}y + b_2 \end{bmatrix} \\ &= \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ &= \mathbf{A}\mathbf{x} + \mathbf{b}. \end{aligned} \tag{1}$$

A critical point (x_c, y_c) is a point where the vector field is zero, i.e, $\mathbf{v}(x_c, y_c) = \mathbf{0}$.

The interesting features of a 2D vector field are its critical points, or derived features related to critical points. Critical points can be classified according to the behavior of nearby vector field data, subject to a certain local polynomial approximation. Several techniques have been developed to understand the nature of 2D critical points and to classify them. A summary of classification techniques is presented in Ref. 1.

A similarity measure between vector should consider the fact that a vector quantity has direction and magnitude. Another desirable feature of a similarity measure is a smoothly varying behavior. Considering these requirements, for vectors $\mathbf{v}_i = \begin{bmatrix} v_{i,x} & v_{i,y} \end{bmatrix}^T$ located at positions $(x_i, y_i)^T$ and $\mathbf{v}_j = \begin{bmatrix} v_{j,x} & v_{j,y} \end{bmatrix}^T$ located at positions $(x_j, y_j)^T$, we define the similarity measure

$$w(\mathbf{v}_i, \mathbf{v}_j) = \alpha \cdot e^{-\mathrm{dist}(\mathbf{v}_i, \mathbf{v}_j)} + (1 - \alpha) \cdot e^{-\mathrm{diff}(\mathbf{v}_i, \mathbf{v}_j)}, \tag{2}$$

where

$$\mathrm{dist}(\mathbf{v}_i, \mathbf{v}_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{3}$$

and

$$\mathrm{diff}(\mathbf{v}_i, \mathbf{v}_j) = (v_{i,x} - v_{j,x})^2 + (v_{i,y} - v_{j,y})^2. \tag{4}$$

The parameter α is also required to vary between 0 and 1. A small value of α emphasizes the difference in direction and magnitude while a large value of α places more weight on distance between data locations. Equation (3) is the Euclidean distance function, and Equation (4) is a measure for direction and magnitude. Other difference measures could also be used as long as they lead the desired characteristics of a clustering. For example, we could also use the measure

$$w(\mathbf{v}_i, \mathbf{v}_j) = e^{-\left(\frac{\mathrm{dist}(\mathbf{v}_i, \mathbf{v}_j)}{\alpha}\right)^2} \cdot e^{-\left(\frac{\mathrm{diff}(\mathbf{v}_i, \mathbf{v}_j)}{\beta}\right)^2}. \tag{5}$$

To simplify the use of Equation (5), the parameter β could be set to 1 and α could vary in the range $(0, \infty)$. With this setting, an α value smaller than 1 would place more emphasis on the vector norm. An α value larger than 1 shifts emphasis to Euclidean point distance.

2.2. The Extended NC Method

The similarity measure should not be applied to the original vectors data directly, since two vectors associated with the same critical point can have opposite directions and highly different magnitudes. Instead, similarity is determined by how well two vectors can be represented by a linear least-squares approximation. The process of measuring similarity and constructing a partition of vector field data can be described as follows:

1. Construct the weight (or association) matrix $\mathbf{W}$:

   (a) For each vector $\mathbf{v}_i = \begin{bmatrix} v_{i,x} & v_{i,y} \end{bmatrix}^T$ at location $(x_i, y_i)^T$, randomly pick m neighboring vector data within a circle of radius r centered at $(x_i, y_i)^T$. The neighboring vectors can be represented as $\mathbf{v}_k = \begin{bmatrix} v_{k,x} & v_{k,y} \end{bmatrix}^T$, where $k = 1, 2, 3, \ldots, m$. The locations of these neighboring vectors are $\mathbf{x}_k = (x_k, y_k)$.

   (b) Considering positional and vector data, solve the following linear least-squares problem:

$$\begin{bmatrix} x_i & 0 & y_i & 0 & 1 & 0 \\ 0 & x_i & 0 & y_i & 0 & 1 \\ x_1 & 0 & y_1 & 0 & 1 & 0 \\ 0 & x_1 & 0 & y_1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_m & 0 & y_m & 0 & 1 & 0 \\ 0 & x_m & 0 & y_m & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ a_{1,2} \\ a_{2,2} \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} v_{i,x} \\ v_{i,y} \\ v_{1,x} \\ v_{1,y} \\ \vdots \\ v_{m,x} \\ v_{m,y} \end{bmatrix}. \tag{6}$$

   (c) Evaluate the resulting linear least-squares approximation at points $\mathbf{x}_n$, i.e., at the locations of the chosen neighboring vectors, leading to

$$\hat{\mathbf{v}}_k = \begin{bmatrix} \hat{v}_{k,x} & \hat{v}_{k,y} \end{bmatrix}^T = \mathbf{A} \cdot \mathbf{x}_k + \mathbf{b}.$$

   (d) Compute the similarity $\mathbf{w}(\mathbf{v}_i, \mathbf{v}_k)$ between $\mathbf{v}_i$ and $\mathbf{v}_k$ by applying Equation (2) to $\mathbf{v}_i$ and $\hat{\mathbf{v}}_k$, and place it as the (i, k) and (k, i) elements of the weight matrix $\mathbf{W}$.

2. Construct a vector $\mathbf{d}$ as follows:

$$\mathbf{d}(i) = \sum_{j=1}^{N} \mathbf{W}(i, j),$$

   where N is the total number of initial vectors.

3. Compute the eigenvector associated with the second-smallest eigenvalue of the matrix $\mathbf{D}^{-\frac{1}{2}} \cdot (\mathbf{D} - \mathbf{W}) \cdot \mathbf{D}^{-\frac{1}{2}}$, where $\mathbf{D}$ is a diagonal matrix having $\mathbf{d}$ as its diagonal elements.

4. Partition the vector field:

   (a) Use the signs of the eigenvector components as indicators to partition the vector data set. Vectors associated with the same sign are placed in the same cluster. The resulting partition is the first-level clustering result.

   (b) Finer partitioning is achieved by executing the described procedure again on each cluster or by using the eigenvectors associated with the third- and fourth-smallest eigenvalues as indicator vectors.[9]

Once a cluster has been determined, the vectors $\mathbf{v}_m$ in the cluster can be used to compute the linear least-squares approximation, defined by $\mathbf{A}$ and $\mathbf{b}$ as in Equation (1). Let $\hat{\mathbf{v}}_m$ be the approximated vectors, the quality of the approximation is measured by the mean-squared error (MSE) defined as

$$\text{MSE}(\mathbf{v}_m, \hat{\mathbf{v}}_m) = \frac{1}{M} \sum_{i=1}^{M} \left[(\mathbf{v}_{i,x} - \hat{\mathbf{v}}_{i,x})^2 + (\mathbf{v}_{i,y} - \hat{\mathbf{v}}_{i,y})^2 \right], \tag{7}$$

where M is the number of vectors in the cluster.

An analytically defined vector field with a saddle point and a repelling focus is shown in Fig. 1 along with the initial partition and the approximated vectors for the two clusters obtained. The initial partition is obtained with our extended NC method by using the similarity measure defined by Equation (2). Cluster 1 contains 198 vectors, and the critical point inside is identified as a saddle point. Cluster 2 contains 202 vectors, and it is identified as a repelling focus. The MSE is 0.008414 for cluster 1, and 0.018914 for cluster 2. The linear least-squares approximation representation for cluster 1 is given by

$$\mathbf{A}_1 = \begin{bmatrix} 0.0000 & -1.0000 \\ -1.9779 & 1.0207 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_1 = \begin{bmatrix} 0.0000 \\ -1.0207 \end{bmatrix}.$$

The linear least-squares approximation for cluster 2 is given by

$$\mathbf{A}_2 = \begin{bmatrix} -0.0000 & -1.0000 \\ 1.8927 & 0.9591 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_2 = \begin{bmatrix} 0.0000 \\ -0.9658 \end{bmatrix}.$$

The same vector field with a resolution of 50-by-50 is shown in Fig. 2. The MSE value for the two clusters are 0.122503 and 0.104881, respectively. A rotated saddle-focus vector field is shown in Fig. 3. The MSE for the two clusters are 0.006641 and 0.003203, respectively. In both examples, the types of critical points are correctly identified as saddle point and repelling focus.

Fig. 4 and Fig. 5 show the effect of using different α values for our extended NC method. The parameter setting used in Fig. 4 is $r = 6$ and $m = 6$, while the examples in Fig. 5 use the parameter setting $r = 5$ and $m = 14$. The partitions with α set to $0.0, 0.1, 0.5$ and 1.0 are shown in both figures. It is interesting to observe that the values of r and m seem to have a greater effect on the quality of the partition than the value of α.

3. CLUSTER REFINEMENT AND HIERARCHY CONSTRUCTION

3.1. Cluster Refinement

The MSE is a quantitative measure for the quality of a partition. Therefore, the strategy for refining clusters should be aimed at the reduction of the MSE. We have derived a cluster refinement strategy based on the goal of reducing the MSE values associated with clusters. The refinement strategy consists of these steps:

1. Store the membership information and the approximation matrices for both clusters.
2. Find the vector in cluster 1 with largest squared-error (SE) value. Move it to cluster 2 and compute a new linear least-squares approximation and new MSE values for both clusters. If the sum of the MSE values of the two clusters is smaller than the sum of the original cluster MSE values, go back to step 1; otherwise, continue with step 3.
3. Store the membership information and the approximation matrices for both clusters.
4. Find the vector in cluster 2 with largest SE value. Move that vector to cluster 1 and compute a new linear least-squares approximation and new MSE values for both clusters. If the sum of the MSE values of the two clusters is smaller than the sum of the original cluster MSE values, go back to step 3; otherwise, continue with step 5.
5. If any vector in cluster 2 has been moved to cluster 1 under step 4, go back to step 1; otherwise, stop.

The linear least-squares approximation and the MSE values must be re-computed for both clusters each time a vector is moved.

Our experiments indicate that the refinement strategy works reasonably well for vector fields with two critical points. Parts (f), (g) and (h) of Fig. 1 show the refined partition of the saddle-focus vector field and the approximated vectors using the refinement strategy. The MSE values for both clusters are 0.000000 after refinement. Both clusters contain 200 vectors after refinement. The types and the locations of the critical points are accurately identified and approximated.

Another example obtained with the refinement strategy is shown in part (d) of Fig. 3. The MSE value is 0.000000 after refinement, and both clusters contain 200 vectors. The types and locations of the critical points are also accurately identified and approximated.

3.2. Hierarchy Construction

To obtain a finer partition, the eigenvector associated with the third-smallest eigenvalue can be used as indicator vector to split the two refined clusters obtained from the first-level segmentation into four clusters. A refinement process can then be applied to the four clusters. This refinement process is similar to the one described in Section 3.1 that aims at reducing the MSE values of clusters by moving the vector with the largest SE to the neighboring cluster.

Experiments with a four-focus vector field are shown in Figs. 6 and 7. Part (a) of Fig. 6 shows the initial partition with MSE values of 1.044294 and 1.044294 for cluster 1 and cluster 2, respectively. Both of the clusters are incorrectly identified as saddle points. After the application of the refinement strategy, the MSE values of the clusters are reduced to 0.797128 and 0.786701. The types of clusters are both identified as attracting nodes after the refinement. Part (c) of Fig. 6 shows the initial second-level cut together with the refined first-level cut. The first cluster (in the lower-right corner) has an MSE value of 0.319273. The second cluster (in the upper-right corner) has an MSE value of 0.373211. The third cluster (in the lower-left corner) has an MSE value of 0.379029. The fourth cluster (in the upper-left corner) has an MSE value of 0.463941. All four clusters are correctly identified as attracting foci. But the approximated locations of the critical points are incorrect. After the second-level refinement process, as shown in part (d) of the same figure, the locations of the critical points are all correctly approximated. The MSE value of every cluster after the refinement is 0.000000. Each cluster contains exactly 100 vectors after refinement.

Fig. 7 shows an example where the strategy of performing a second-level cut after refining first-level partition fails. Only one of the four clusters is correctly identified as attracting focus, while others are being identified as attracting nodes or saddle points. Based on the results shown in this figure, we derive another approach to perform the hierarchical clustering. This procedure is described as follows: First, perform multiple level of clustering. Second, once the partitioning process stops, apply the refinement process on each cluster. The results of applying procedure are shown in Figs. 8 and 9 and they will be discussed in the next section.

4. RESULTS

4.1. Two-dimensional Analytically Defined Vector Fields

Fig. 8 shows an example of performing the refinement only after two levels of partition. Part (c) of Fig. 8 shows the four clusters are correctly identified along with the locations of the approximated critical points. The MSE value of every cluster after in part (c) is 0.000000.

Fig. 9 shows another example of same process used for the vector field shown in Fig. 8. As shown in part (b) of Fig. 9, only two locations of the critical points are correctly identified though three of the clusters are classified correctly as attracting foci. Cluster 1 (in the upper-right corner) has a MSE value of 0.426785 and contains 117 vectors. Cluster 2 (in the upper-left corner) has a MSE value of 0.000000 and contains only 57 vectors. Cluster 3 (in the lower-right corner) has a MSE value of 0.367830 and contains 132 vectors. Cluster 4 (in the lower-left corner) has a MSE value of 0.000000 and contains 94 vectors.

Same procedure is applied to a three-focus vector field as shown in Fig. 10 and Fig. 11. The refined partition shown in Fig. 10 is a perfect cut. The MSE value for every cluster is 0.000000 and the types and the location of the critical points are all identified correctly. But the results shown in Fig. 11 indicate that the only cluster 2 (on the left-hand side) contains the majority of the correct vectors as indicated by its relatively small MSE value of 0.012896. The other three clusters have high MSE values and incorrectly approximated locations of critical points.

An observation can be made from the results. If the cluster obtained from the initial cut does not contain enough vectors that are associated with same critical point, then the linear least-squares approximation will not be good. The SE and the MSE values will not provide enough useful information for the refinement strategy to work with. A large SE value computed from the least-squares approximation does not give correct indication on which vector is mis-clustered.

Another consideration for hierarchy construction is the need for merging neighboring clusters that are associated with the same critical point. As shown in Fig. 10, the location of cluster 1's critical (asterisk) and the location of cluster 3's critical point (cross) are inside the same cell. This information along with the fact that both clusters are identified as attracting focus show that these two clusters should be merged together to form a single cluster.

4.2. Two-dimensional Numerical Simulated Vector Fields

The extended NC method is applied on a complete turbulent flow data set.[8] This data set contains 703 critical points. A complete partition of this data set will required $log_2(703)$ applications of the current version of extended NC method. In order to have a meaningful complete partition of this data set, a robust refinement strategy is needed. Fig. 12 shows a two-level partition of the data set without any using any refinement strategy.

A partition of a smaller portion of the turbulent flow data set are shown in Fig. 13. Part (a) of Fig. 13 shows the un-refined two-level cut. The first cluster (in the lower-right corner) has an MSE value of 0.080653 and contains 95 vectors. The second cluster (in the lower-left corner) has an MSE value of 0.102449 and contains 85 vectors. The third cluster (in the upper-right corner) has an MSE value of 0.190171 and contains 92 vectors. The fourth cluster (in the upper-left corner) has an MSE value of 0.038333 and contains 88 vectors. After the second-level refinement process, as shown in part (b) of the same figure, the first cluster has an MSE value of 0.080617 and contains 96 vectors. The second cluster has an MSE value of 0.101152 and contains 87 vectors. The third cluster has an MSE value of 0.186285 and contains 98 vectors. The fourth cluster has an MSE value of 0.000385 and contains 79 vectors. The first cluster is identified as attracting focus, while other clusters are identified as saddle points. An observations can be made from the results. Although the MSE values for these clusters are small, further partitions of the clusters are still needed. A relative error measure could be used to aid the qualitative analysis of the clusters. This will be the subject of further study.

5. CONCLUSIONS AND FUTURE WORK

In preparation of achieving the goal of constructing a hierarchical representation for vector fields, we have introduced the utilization of the normalized cut method along with the proposed similarity measurement. A cluster refinement is presented. Experimental results demonstrate the potentials of the approach. Many issues are still remained to be resolved in order to realize the ultimate goal of applying the proposed method to 3D time-varying vector fields. Some of these issues are:

1. Cluster refinement and hierarchy construction: To obtain the desired hierarchical representation of a vector field, the extended NC method must incorporate a robust cluster refinement strategy in order to obtain high quality (in terms of MSE or other criteria) fine clusters. One alternative to the previous refinement method can be summarize as follows:

   (a) Store the membership information and the approximation matrices for both clusters.

   (b) Find the vector with largest SE value among all vectors in the clusters. Move that vector to the other cluster, compute a new linear least-squares approximation and new MSE values for both clusters. If the sum of the MSE values of the two clusters is smaller than the sum of the MSE values of the original clusters, go back to step 1; otherwise, stop.

   Another possible refinement solution is the use of Linkage Refinement scheme.[4]

2. Handling 3D data sets: The feature we would like to identify in a 3D vector field data set is the vortex core. The first step is to obtain the capability to identify vortex cores.[7,10] The next step is to incorporate the vortex core identification technique into the similarity measurement. Alternatively, one can also perform the extraction of the vortex cores prior the clustering process, then use the extracted volumes as the building blocks in the hierarchy construction process.

ACKNOWLEDGMENTS

We express our gratitude to the members of CIPIC's Visualization and Graphics Research Group at the University of California, Davis, and to the members of the Applied Numerical Algorithms Group(ANAG) and the Visualization Group at the Lawrence Berkeley National Laboratory for their valuable suggestions and support. We would also like to express our gratitude to Gerik Scheuermann and Wolfgang Kollmann for providing the turbulent vector field data set.

This work was supported by the Director, Office of Science, Office of Advanced Scientific Computing Research, Mathematical, Information and Computational Sciences Division, U.S. Department of Energy under contract No. DE-AC03-76SF00098.

This work was supported by the Lawrence Berkeley National Laboratory; the National Science Foundation under contract ACI 9624034 (CAREER Award) and ACI 022909, through the Large Scientific and Software Data Set Visualization (LSSDSV) program under contract ACI 9982251, and through the National Partnership for Advanced Computational Infrastructure (NPACI). We also acknowledge the support of ALSTOM Schilling Robotics and SGI.

REFERENCES

1. J.-L. Chen, Z. Bai, and T. Ligocki, "Segmentation of piecewise linear vector fields," submitted to *Geometric Modeling for Scientific Visualization*, G. Brunnett, B. Hamann, and H. Müller, eds., Springer-Verlag, 2003.
2. W. de Leeuw, R. Van Liere, "Visualization of global flow structures using multiple levels of topology," in *Data Visualization*, E. Gröller, H. Löffelmann, W. Ribarsky, eds., pp. 45-52, Springer-Verlag, Vienna/New York, 1999.
3. W. de Leeuw and R. Van Liere, "Collapsing flow topology using area metrics," in *Proc. Visualization '99*, D. Ebert, M. Gross, and B. Hamann, eds., pp. 349-354, 1999.
4. C. Ding, X. He, H. Zha, M. Gu, and H. Simon, *Spetrcal Min-max Cut for Graph Partitioning and Data Clustering*, Tech. Report LBNL-47848, Natonal Engery Research Center, Scientific Computing Division, Lawrence Berkeley National Laboratory, 2001.
5. H. Garcke, T. Preusser, M. Rumpf, A. Telea, U. Weikard, and J. J. Van Wijk, "A continuous clustering method for vector vields," in *Proc. Visualization 2000*, T. Ertl, B. Hamann, and A. Varshney, eds., pp. 351-358, 2000.
6. H. Garcke, T. Preusser, M. Rumpf, A. Telea, U. Weikard, and J. J. Van Wijk, "A phase field model for continuous clustering on vector fields," *IEEE Trans. Visualization and Computer Graphics*, **7**, pp. 230-241, 2001.
7. D. N. Kenwright, and R. Haimes, R. "Automatic vortex core detection", *IEEE Computer Graphics and Applications*, **18**, pp. 70-74, 1998.
8. G. Scheuermann, B. Hamann, K. I. Joy, and W. Kollmann, "Visualizing local vector field topology," *J. Electronic Imaging*, **9**, pp. 356-367, 2000.
9. J. Shi, and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **22**, pp. 888-905, 2000.
10. D. Sujudi, and R. Haimes,*Identification of Swirling Flow in 3-D Vector Fields*, Tech. Report, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 1995.
11. A. Telea, and J. J. Van Wijk,"Simplified representation of vector fields," *Proc. Visualization '99*, D. Ebert, M. Gross, and B. Hamann, ed., pp. 35-42, 1999.
12. X. Tricoche, G. Scheuermann, and H. Hagen, "A topology simplification method for 2D vector fields," *Proc. Visualization 2000*, T. Ertl, B. Hamann, and A. Varshney, eds., pp. 359-366, 2000.

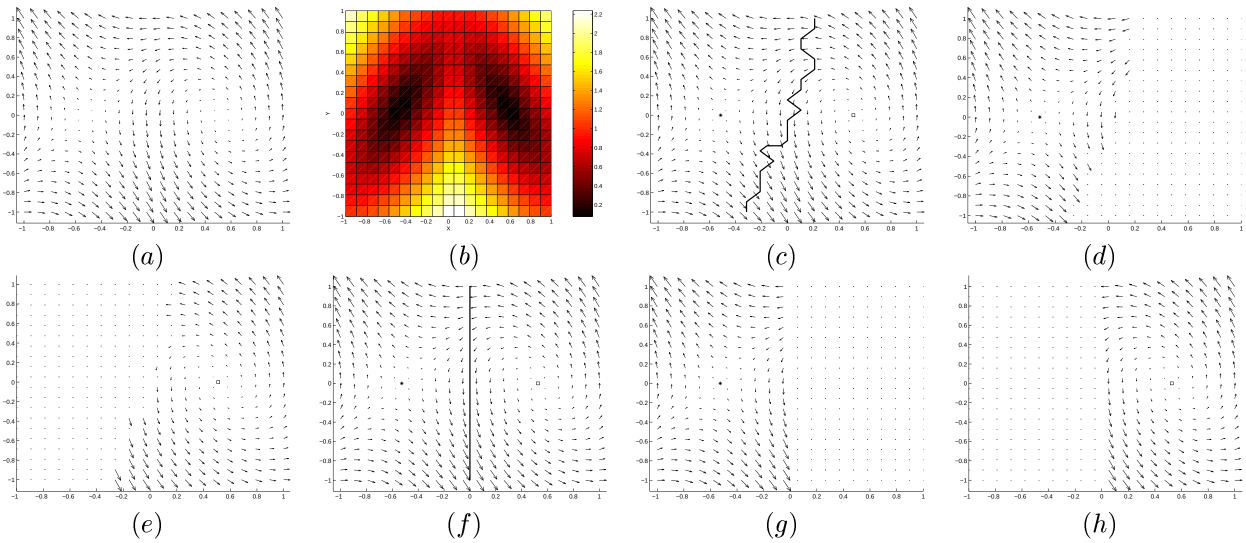


Figure 1. (a) Saddle-focus vector field of resolution 20-by-20. (b) The magnitude of the vector field. (c) The initial partition of the vector field with parameter values $r = 5$, $m = 14$, and $\alpha = 0.1$. The approximated locations for the critical points inside cluster 1 and 2 are represented by an asterisk and a square box, respectively. (d) The approximated vectors in cluster 1. (e) The approximated vectors in cluster 2. (f) The refined partition of the vector field. (g) The approximated vectors in cluster 1 based on the refined partition. (h) The approximated vectors in cluster 2 based on the refined partition.

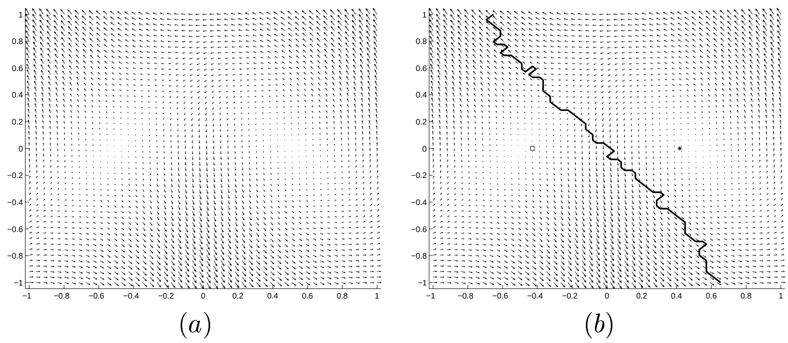


Figure 2. (a) Saddle-focus vector field of resolution 50-by-50. (b) The initial partition of the vector field using parameter values $r = 5$, $m = 14$, and $\alpha = 0.1$.

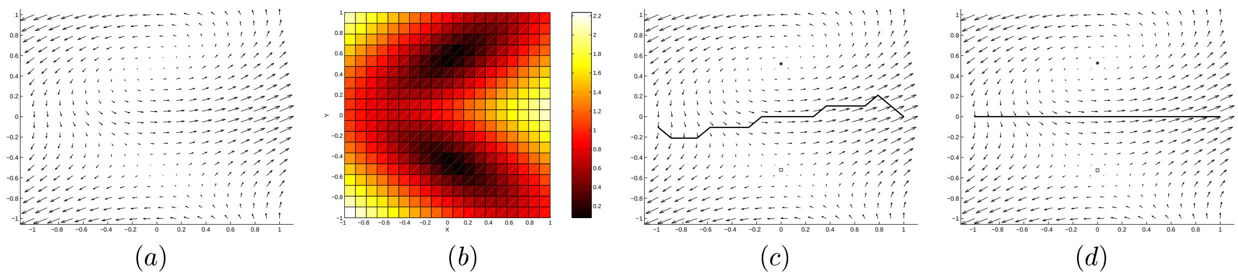


Figure 3. (a) Saddle-focus vector field of resolution 20-by-20. (b) The magnitude of the vector field. (c) The initial partition of the vector field using parameter values $r = 5$, $m = 14$, and $\alpha = 0.1$. The approximated locations for the critical points inside clusters 1 and 2 are represented by a asterisk and a square box, respectively. (d) The refined partition of the vector field.

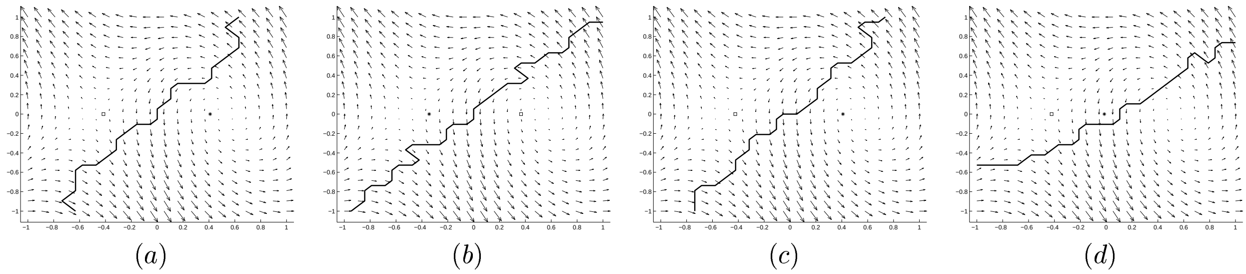


Figure 4. Partition of saddle-focus vector field of resolution 20-by-20 using parameter values $r = 3$, $m = 6$, and (a) $\alpha = 0.0$, MSE values are 0.159024 and 0.136324. (b) $\alpha = 0.1$, MSE values are 0.257995 and 0.240696. (c) $\alpha = 0.5$, MSE values are 0.164455 and 0.145078. (d) $\alpha = 1.0$, MSE values are 0.364313 and 0.332760.

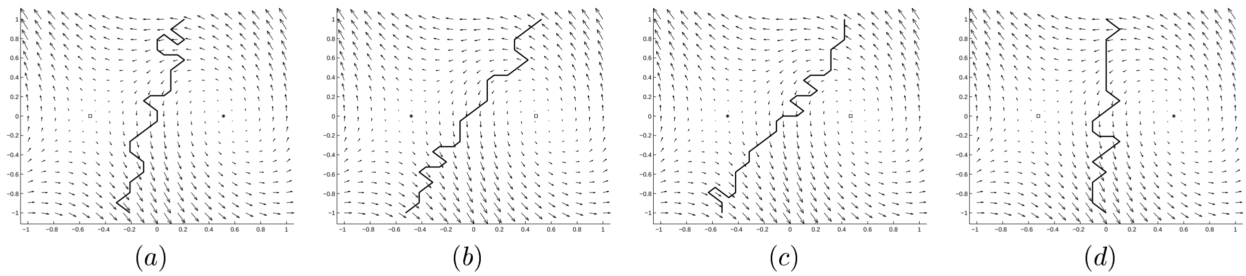


Figure 5. Partition of saddle-focus vector field of resolution 20-by-20 using parameter values $r = 5$, $m = 14$ and (a) $\alpha = 0.0$, MSE values are 0.014938 and 0.006757. (b) $\alpha = 0.1$, MSE values are 0.061088 and 0.065333. (c) $\alpha = 0.5$, MSE values are 0.052077 and 0.090109. (d) $\alpha = 1.0$. MSE values are 0.001250 and 0.000697.

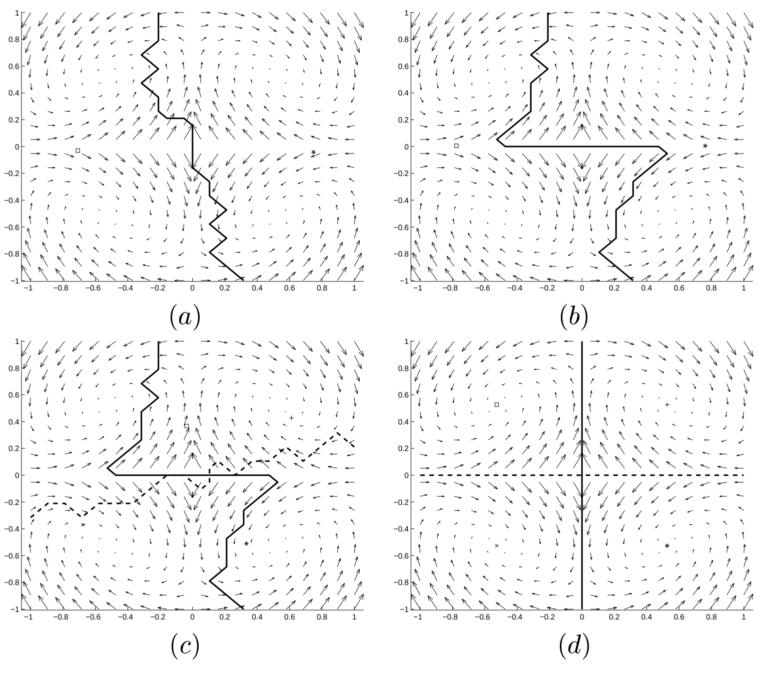


Figure 6. (a) Initial partition of a four-focus vector field using parameter values $r = 4$, $m = 10$, and $\alpha = 0.1$. (b) First-level partition with refinement. (c) Refined first-level partition and initial second-level partition (shown by dashed line). (d) Refined first-level partition and refined second-level partition. The types and the locations of the critical points are all identified correctly. The MSE value for every cluster is 0.000000.

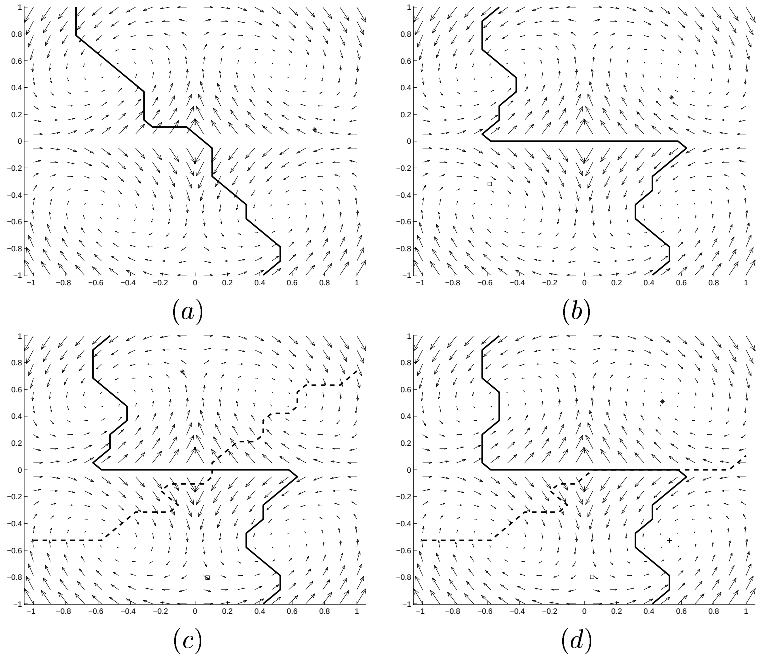


Figure 7. (a) Initial partition of a four-focus vector field with parameter values $r = 4$, $m = 10$, and $\alpha = 0.1$. (b) First-level partition with refinement. (c) First-level partition and initial second-level partition (shown by dashed line). (d) Refined first-level partition and refined second-level partition. Only one type and one location of the critical points are identified correctly.

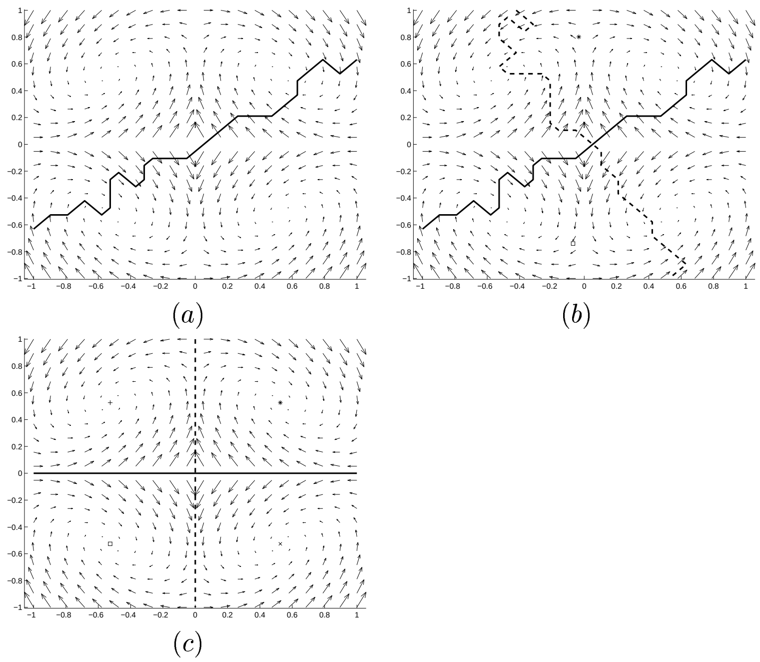
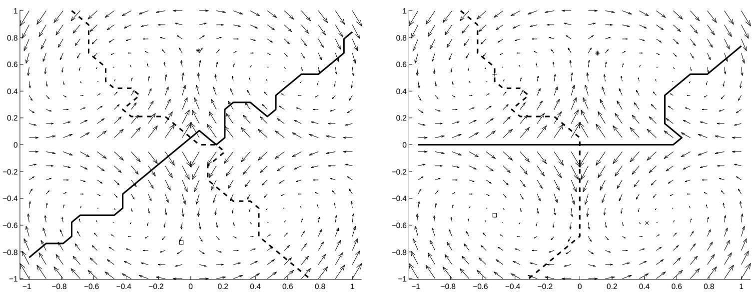
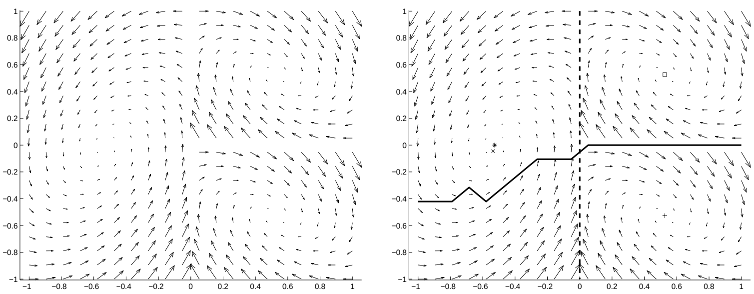


Figure 8. (a) Initial partition of a four-focus vector field using parameter values $r = 4$, $m = 10$, and $\alpha = 0.1$. (b) Initial first-level partition and initial second-level partition (shown by dashed line). (d) Refined partition of the vector field. The types and locations of the critical points are all identified correctly. The MSE value for every cluster is 0.000000.



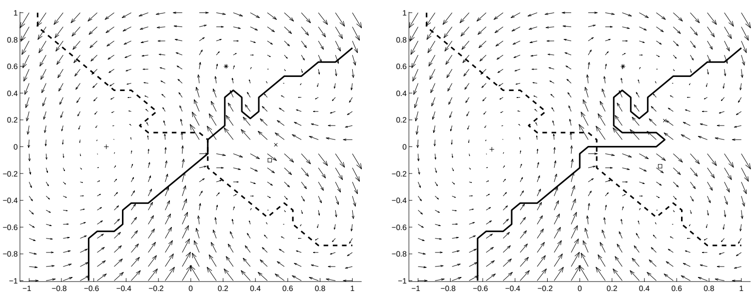
(a) (b)

Figure 9. (a) Initial partition of a four-focus vector field using parameter values $r = 4$, $m = 10$, and $\alpha = 0.1$. (b) Refined partition of the vector field. Only three of the types and two of the locations of the critical points are identified correctly.



(a) (b)

Figure 10. (a) A three-focus vector field. (b) Refined partition of the vector field. The types and locations of the critical points are all identified correctly. The MSE value for every cluster is 0.000000.



(a) (b)

Figure 11. (a) Initial partition of a three-focus vector field. (b) Refined partition of the vector field. Only one cluster is identified correctly.

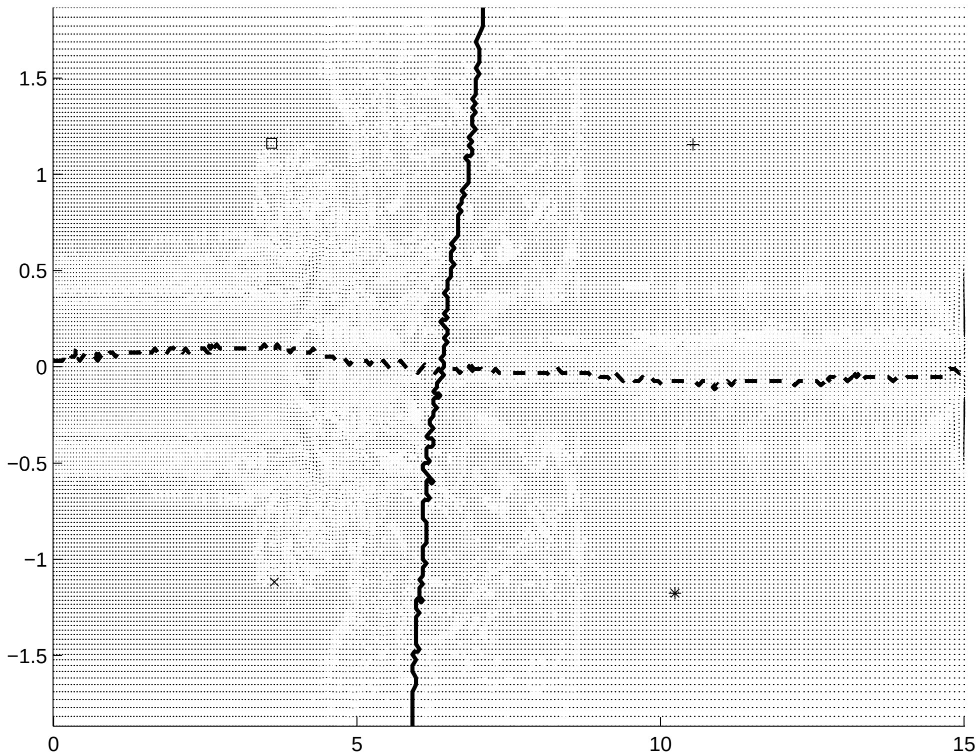
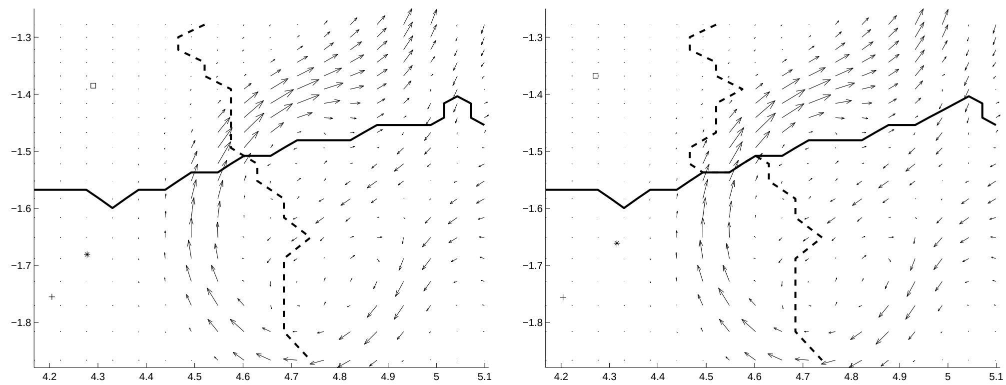


Figure 12. Two-level partition of the turbulent flow data set without using any refinement strategy.



(a) (b)

Figure 13. (a) A small portion of the turbulent flow data set. (b) A refined two-level partition of the vector field.