Vector Field Segmentation With Normalized Cut

Jiann-Liang Chen[1,3,*] Zhaojun Bai[1,*] Bernd Hamann[1,3,*] Terry J. Ligocki[2,†]

[1]Center for Image Processing and Integrated Computing (CIPIC), Department of Computer Science
University of California, Davis
[2]Applied Numerical Algorithms Group, Computational Research Division
Lawrence Berkeley National Laboratory
[3]National Energy Research Scientific Computing Center (NESRC)
Lawrence Berkeley National Laboratory

ABSTRACT

Given a set of points with associated velocity vectors, the goal is to partition a given data set into segments that capture the structure of the underlying two-dimensional (2D) field as defined by its critical points. The presented method models a vector field data as an undirected, weighted graph with the weights computed by a similarity measure that considers locally constructed linear least-squares approximations. A normalized-cut technique is used to partition a given data set into segments such that the vectors in each segment are highly correlated. For very large vector field data sets, we discuss a multilevel method that incorporates coarsening and refinement operations.

1 INTRODUCTION

A common problem in visualizing large vector field data is caused by the large numbers of critical points present in the datasets. The large number of critical points implies a complex topology that is difficult for a user to understand. First reported in [1], We present a method that transforms the 2D vector field visualization problem into an eigenproblem and a maximal matching problem. Assuming that our data is represented in a piecewise-linear fashion, our method models a discrete 2D vector field as a graph, where edges are defined by relationships of the data. Clusters are extracted from the dataset by using an image segmentation algorithm called *normalized cut* (NC) [4] that utilizes the second smallest eigenvector of the graph's associated normalized Laplacian matrix. To improve the quality of clusters, a variant of the *Kernighan-Lin (KL) refinement algorithm* [3] is applied to clusters after segmentation. On the finest level of data representation, a cluster is the set of all originally provided vectors associated with the same critical point that can be expressed or approximated well by a linear polynomial associated with that critical point.

For vector field data with a large number of vectors and critical points, we use multilevel approach that combines a graph coarsening scheme with the NC method. The NC method is applied to the coarsest graph with the smallest number of vertices. Deatil is inserted back into the resulting partition until the original graph is partitioned. A refinement scheme is applied to the graph at every level to improve the quality of partitions.

*{chenjj, bai, hamann}@cs.ucdavis.edu
†tjligocki@lbl.gov

2 ALGORITHM

2.1 Feature Identification

A piecewise linear vector field defined in the xy-plane 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 $\mathbf{v}(x_c, y_c) = \mathbf{0}$. Our objective is to cluster vectors associated with the same critical point together. Essentially, these vectors should be reproducible by the same $\mathbf{A}$ and $\mathbf{b}$ matrices.

2.2 Pairwise Similarity Measures

Considering vector data $\mathbf{v}_i = (v_{i,x}, v_{i,y})^T$ located at position $(x_i, y_i)^T$, $\mathbf{v}_j = (v_{j,x}, v_{j,y})^T$ located at position $(x_j, y_j)^T$, and $\mathbf{v}_k = (v_{k,x}, v_{k,y})^T$ located at position $(x_k, y_k)^T$ let $\hat{\mathbf{v}}_j$ be the approximated vector at location $(x_j, y_j)^T$. We define the similarity between data i and j as

$$w(i,j) = \alpha \cdot e^{-\mathit{dist}(i,j)} + (1-\alpha) \cdot e^{-\mathit{diff}(j)}. \tag{2}$$

The parameter α is set to vary between zero and one. A small value of α emphasizes difference in direction and magnitude, while a large value of α places more weight on distance between data locations. Euclidean distance can be used in the "*dist*" term in (2), i.e.,

$$\mathit{dist}(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{3}$$

The "*diff*" term can be defined as relative L_2-norm:

$$\mathit{diff}(j) = \frac{\sqrt{(v_{j,x} - \hat{v}_{j,x})^2 + (v_{j,y} - \hat{v}_{j,y})^2}}{\sqrt{v_{j,x}^2 + v_{j,y}^2}}. \tag{4}$$

2.3 Vector Field Segmentation

To model a discretized vector field $\mathbf{v}(x,y)$ represented by M vectors, we first construct an undirected, weighted graph $\mathbf{G} = (\mathbf{V}, \mathbf{W})$. The vertices in $\mathbf{V}$ represent the data points, while the weighted edges in $\mathbf{W}$ denote similarities. The goal of this method is to minimize the "disassociation" between two disjoint subsets A

and B, where $A \cup B = \mathbf{V}$ and $A \cap B = \emptyset$. Disassociation is defined as

$$NCut(A,B) = \frac{cut(A,B)}{assoc(A,\mathbf{V})} + \frac{cut(A,B)}{assoc(B,\mathbf{V})}, \tag{5}$$

where $cut(A,B) = \sum_{u \in A, v \in B} \mathbf{W}(u,v)$, $assoc(A,\mathbf{V}) = \sum_{u \in A, t \in V} \mathbf{W}(u,t)$, and $assoc(B,\mathbf{V}) = \sum_{u \in B, t \in V} \mathbf{W}(u,t)$.

It was shown that the minimization of $NCut(A,B)$ is an NP-complete problem [4]. However, an approximate solution is obatined by funding the second smallest eigenvector of the normalized Laplacian matrix, defined as $\mathbf{D}^{-\frac{1}{2}} \cdot (\mathbf{D} - \mathbf{W}) \cdot \mathbf{D}^{-\frac{1}{2}}$, where

$$\mathbf{D}(i,j) = \begin{cases} 0 & , \quad i \neq j. \\ \sum_{k=1}^{M} \mathbf{W}(i,k) & , \quad i = j. \end{cases} \tag{6}$$

2.4 Cluster Refinement

Let $\mathbf{v}_i = (v_{i,x}, v_{i,y})^T, i = 1, 2, 3, \ldots, N_c$ be the vectors associated with a cluster and $\hat{\mathbf{v}}_i = (\hat{v}_{i,x}, \hat{v}_{i,y})^T$ be the approximated vectors, the quality of the cluster is defined by the mean-squared error (MSE):

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

The basic idea of the refinement scheme is to move a datum with the largest squared-error value to another cluster and check whether the MSE decreased. This process is repeated until no further improvement is possible.

2.5 Multilevel Graph Coarsening

Graph coarsening technique is used to accelerate the segmentation process for vector field data with a large number of vectors and critical points. A simplified representation can be reduced further until a desired level of complexity is achieved. Eventually, a spectral graph partitioning technique can be applied to the coarsest level of representation. A key aspect of this multilevel approach is the use of the solution obtained from the coarser levels for the construction of the finer levels. The results obtained for the coarser levels can be viewed as partial solutions to the problem at the finer level. NC method is applied to the coarsest graph with the smallest number of vertices. The resulting partition is "projected back" to finer levels until the original graph is partitioned. A KL–type algorithm is applied to the graph at every level to improve the quality of partitions. We use heavy-edge matching to construct the coarse graphs [2].

3 RESULTS

An analytically defined vector field with a saddle point and a repelling focus along the resulting segmentation is shown in Fig. 1. The NC method was also applied recursively to a portion of a numerically simulated dataset describing turbulent flow. A hierarchical representation of the vector field was obtained by recursively applying the NC method. The similarity measure defined by (2) was used along with the "*dist*" term defined by (3) and the "*diff*" term defined by (4). Level one contains the original vector field; level two contains the two clusters extracted from level one. The original vector field and some of its streamlines are shown in parts (a) and (b) of Fig. (2); parts (c) and (d) show some of the streamlines obtained from the approximated vector fields using two and 28 clusters, respectively.

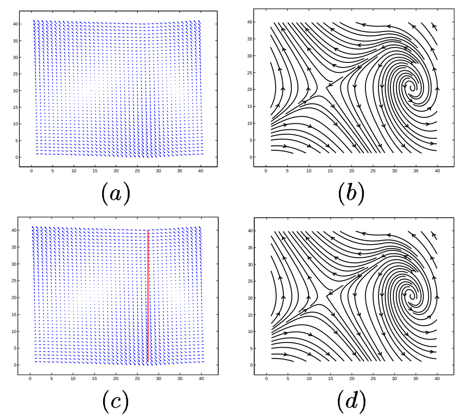

Figure 1: (a) Saddle-focus vector field of resolution 40-by-40. (b) Streamlines computed from the original vector field (a). (c) Refined partition obtained. (d) Streamlines computed from the approximated vector field after refinement.

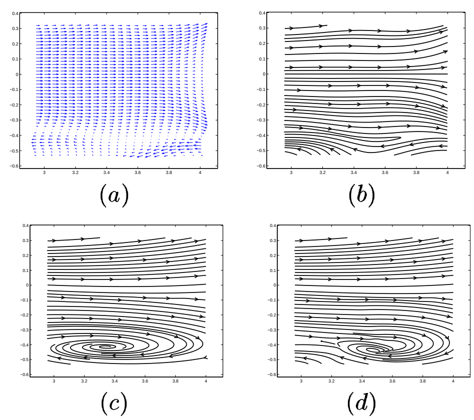

Figure 2: (a) Portion of a turbulent flow field dataset. (b) Streamlines computed from original vector field data shown in (a). (c) Streamlines computed from approximated vector field with two clusters used. (d) Streamlines computed from approximated vector field with 28 clusters used.

REFERENCES

[1] J.-L. Chen, Z. Bai, B. Hamann, and T. J. Ligocki. A normalized-cut algorithm for hierarchical vector field data segmentation. In *Visualization and Data Analysis, 2003*, pages 79–90, 2003.

[2] G. Karypis and V. Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing*, 20(1):359–392, 1998.

[3] B. W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *The Bell System Technical Journal*, 49(2):291–307, 1970.

[4] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.