Flow Visualization

Tutorial on Information Theory in Visualization

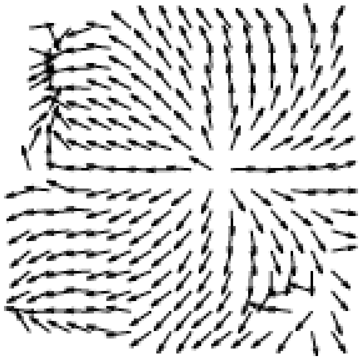
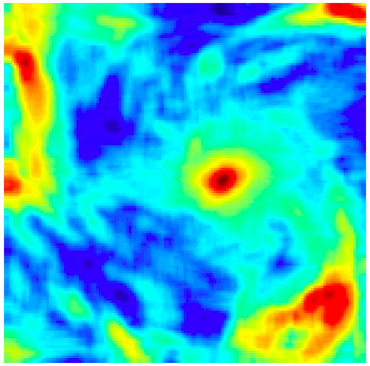
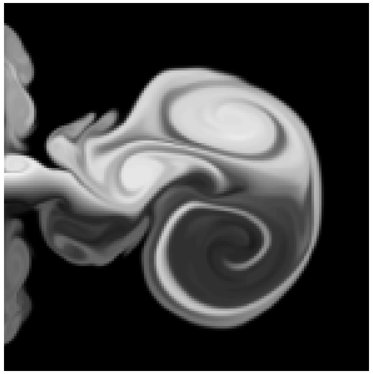
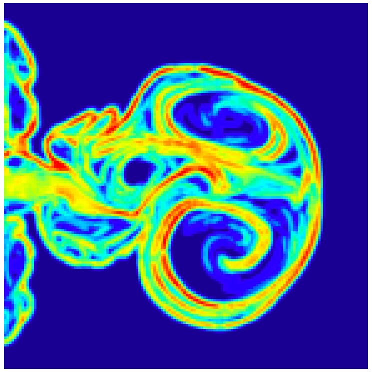
Han-Wei Shen

The Ohio State University

EUROGRAPHICS 2016

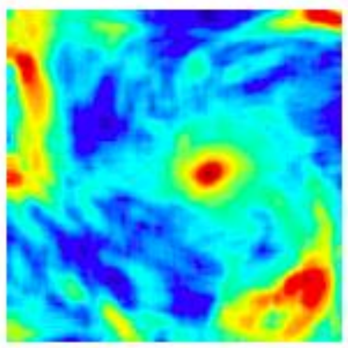
Entropy for Scientific Data

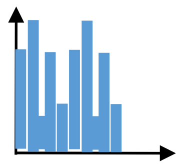
- A data set can be considered as a random variable
- Each data point can be considered as an outcome of the random variable
- We can estimate the information content for the whole data set or for local regions

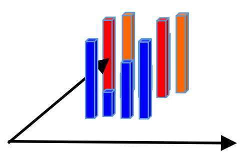
Distributions from Scientific Data

Scalar Distributions



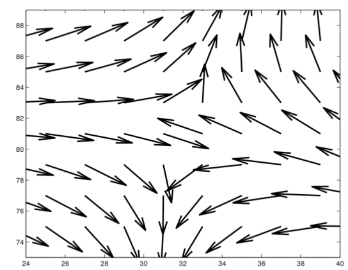


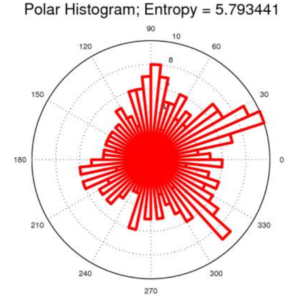
Uni-variate



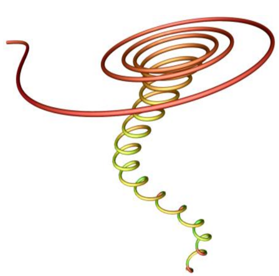
Multi-variate

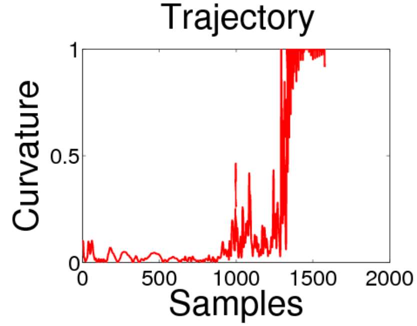
Vector Distributions



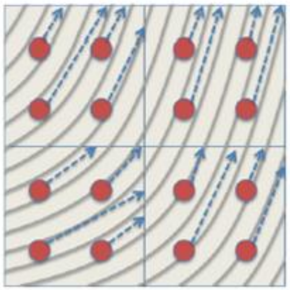


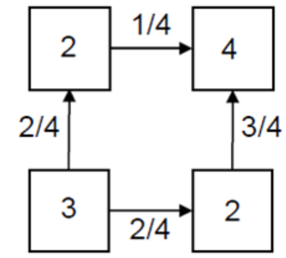
Feature Distributions





State Transitions





Data Sets with Multiple Variables

- Assuming your data set contains two variables X and Y
- You want to know the relationship between X and Y
- You can calculate the conditional entropy, mutual information, etc between these two variables
- Some of the metrics can be used as the ‘information distance’ between two variables

Entropy for Multiple Variables

- Joint Entropy

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y)$$

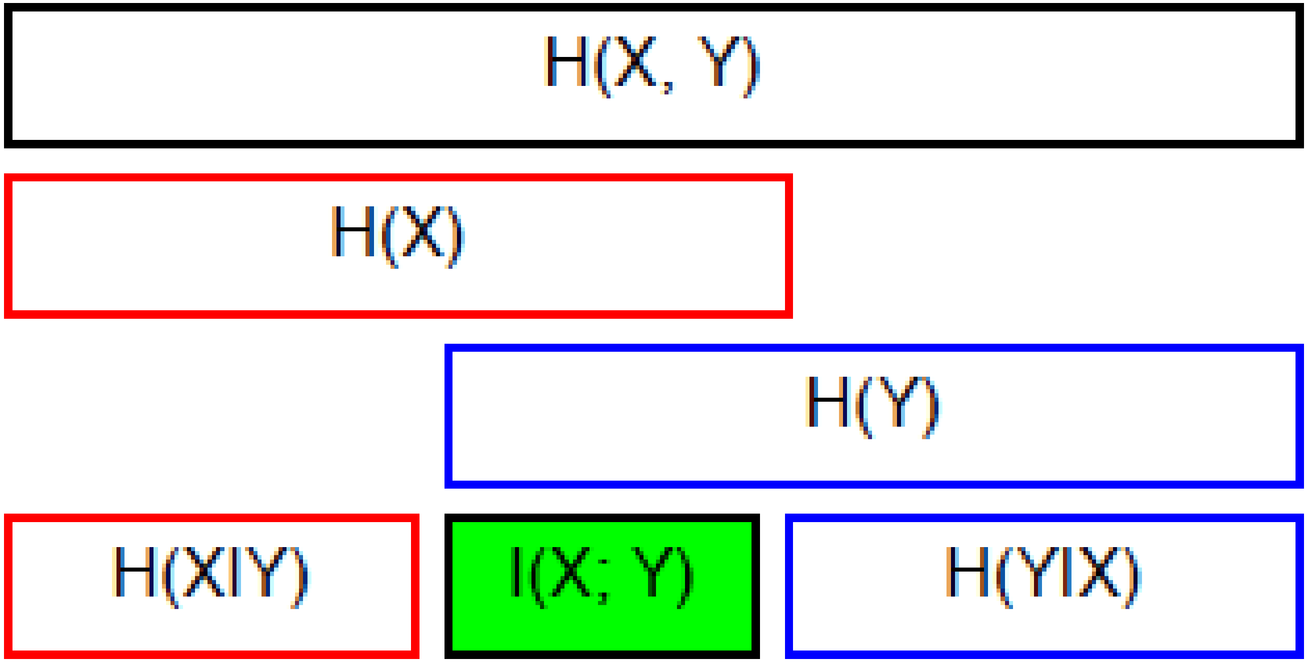
- Conditional Entropy

$$H(X|Y) = \sum_{y\in\mathcal{Y}} p(y)H(X|Y=y) = -\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} p(x,y)\log p(x|y)$$

- Mutual Information

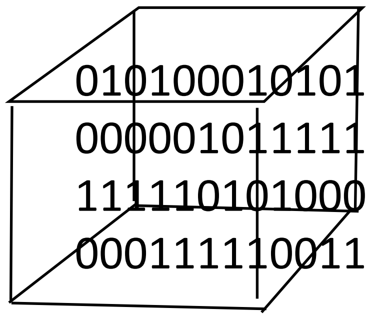
$$I(X;Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Relations of Entropy Measures

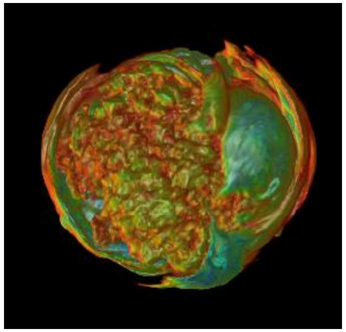


Evaluating Visualization

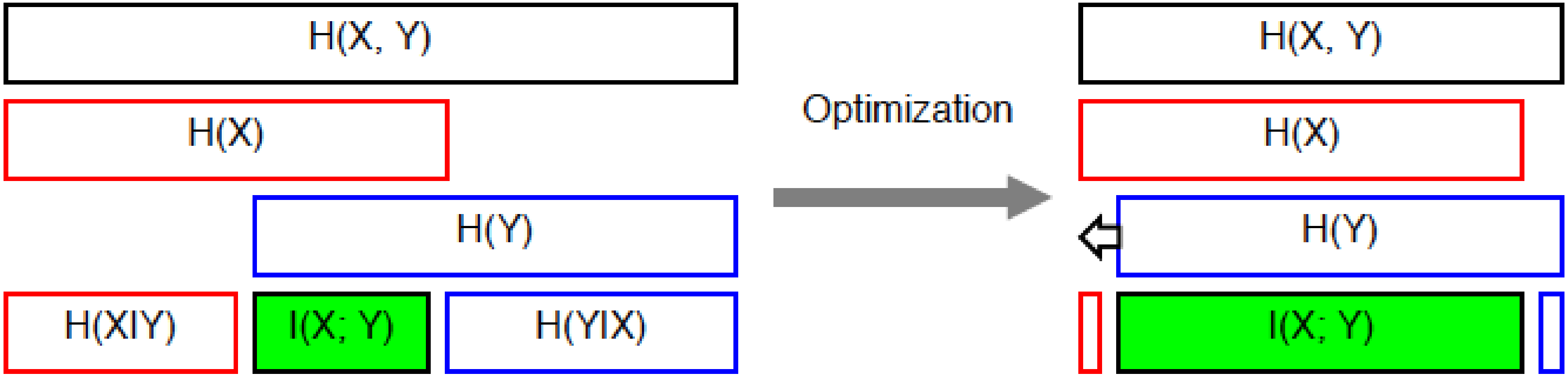
X



Visualization pipeline



Y

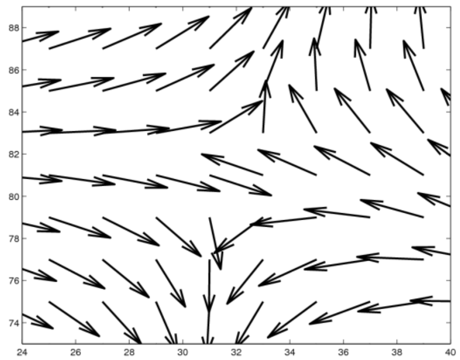


$$H(x) = -\sum_{i=1}^{n} p_i \log p_i$$

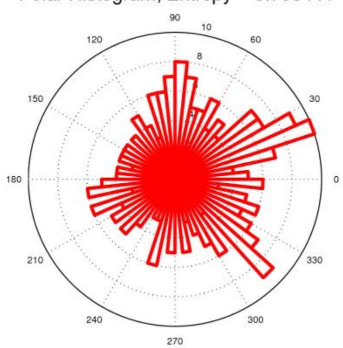
$$I(X;Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Vector Field Analysis

- Concept
  - Treat the vector field as a data source that generates vector orientation as outcome
  - The more diverse the vector orientations, the more information is contained in the vector field
- Measurement
  - Estimate the distribution of the vector orientation
  - Compute the entropy of this distribution as the measurement

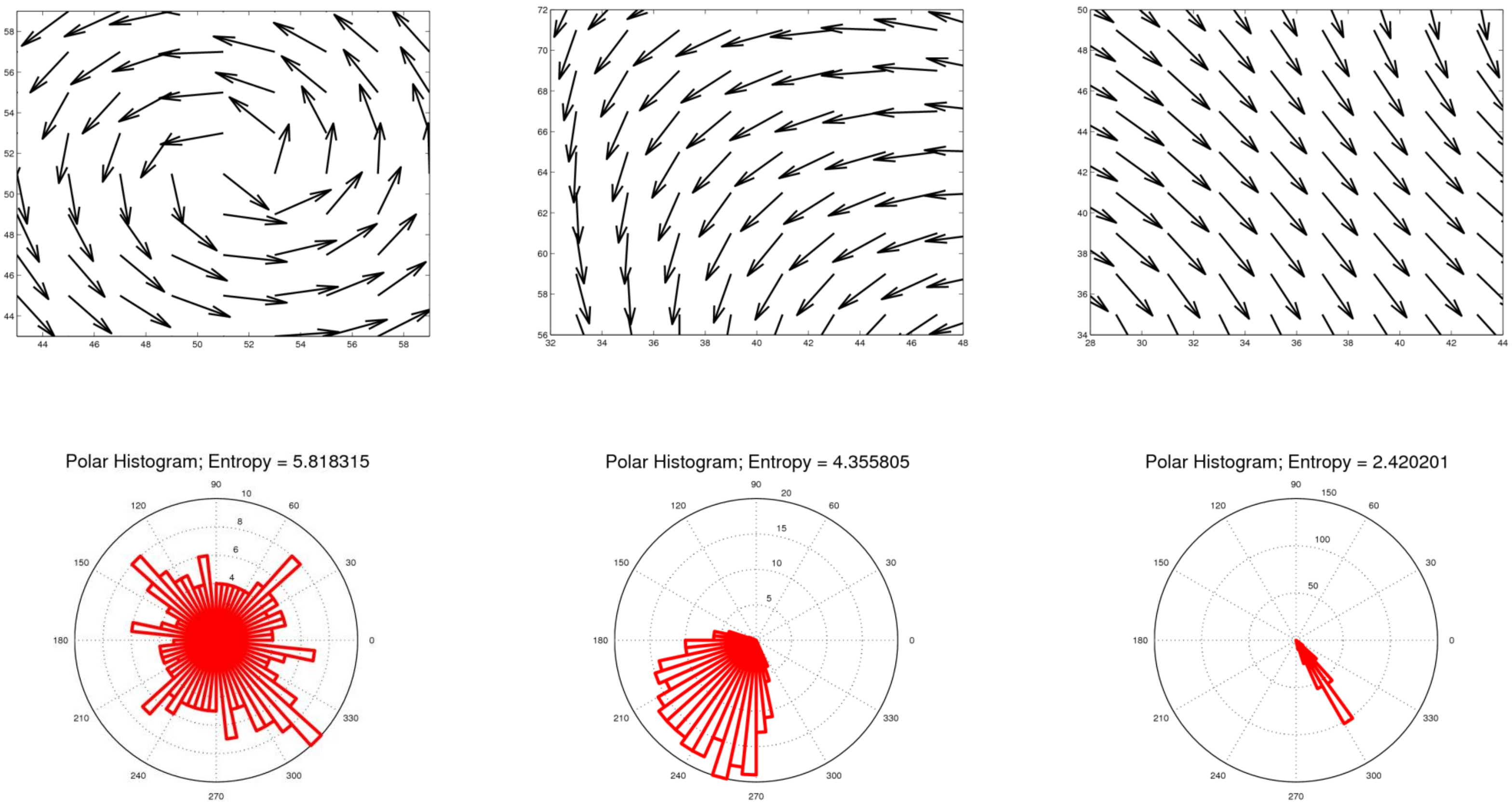


Vector field



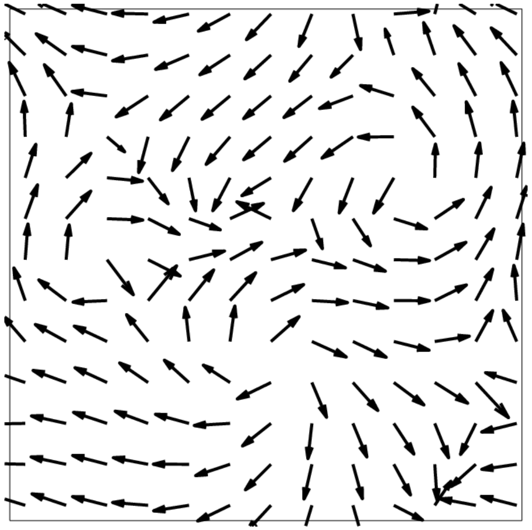
Polar Histogram

Information in Vector Fields

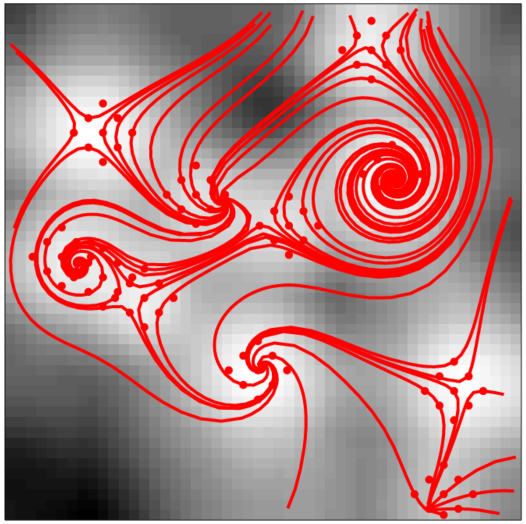


Entropy Field and Seeding

Measure the entropy around each point's neighborhood



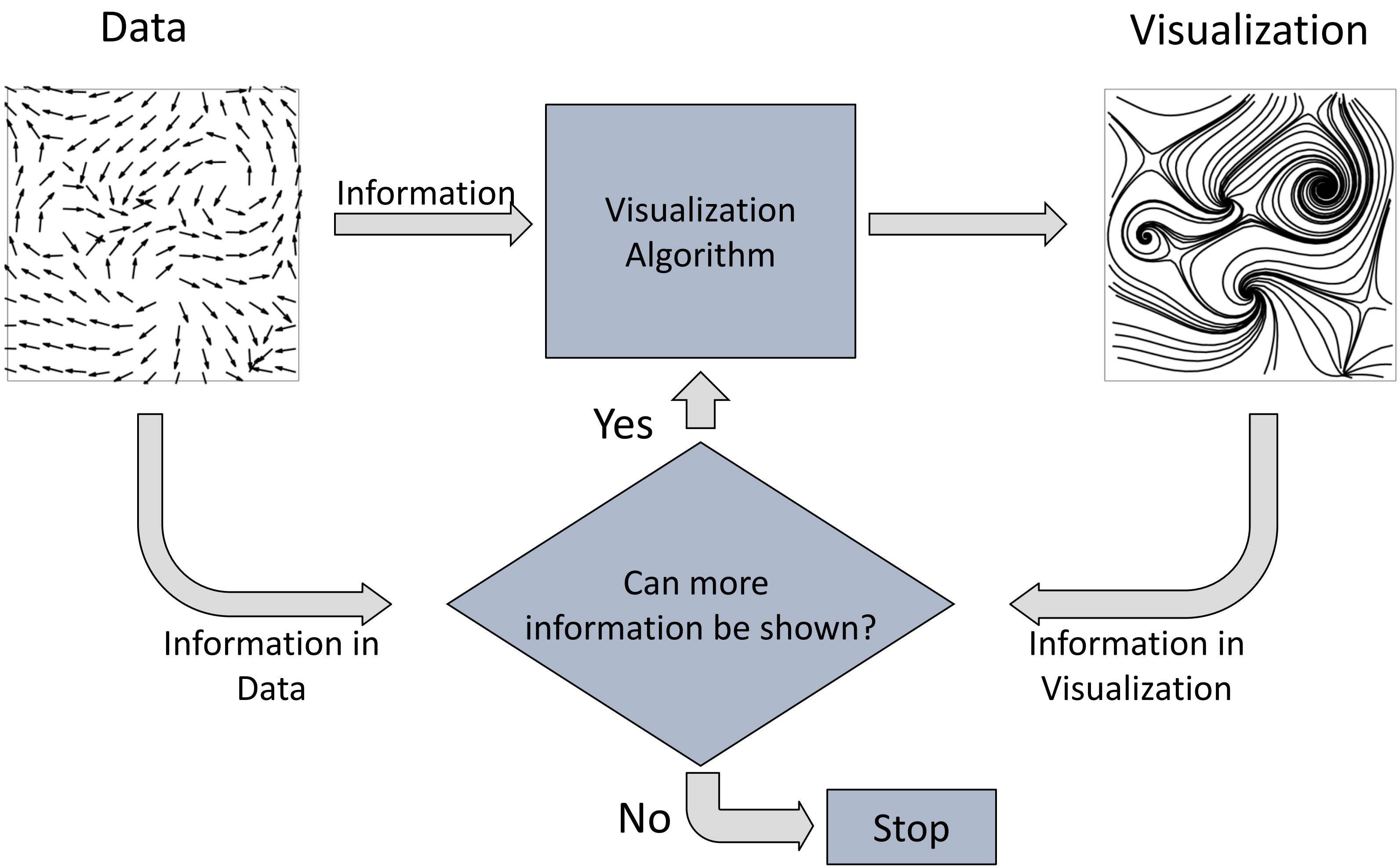
Vector Field



Entropy field: higher value means more information in the corresponding region

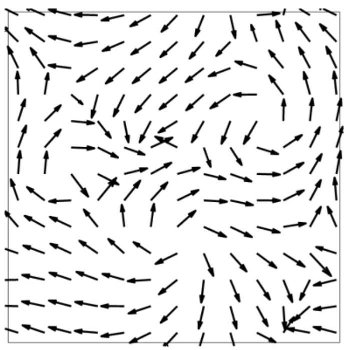
Entropy-based seeding: Places streamlines on the region with high entropy

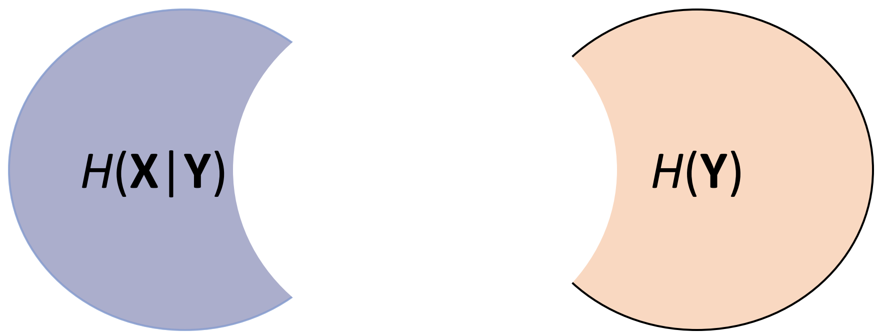
Evaluation of Visualization



Information Comparison between Data/Visualization

Vector Field **X**





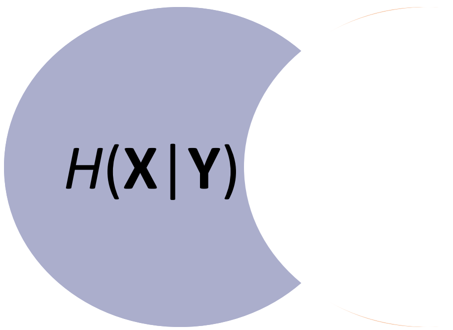
Streamlines **Y**

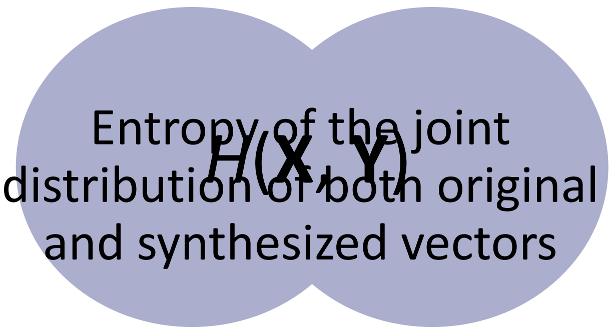
Conditional entropy $H(\mathbf{X}|\mathbf{Y})$:
The information in **X** not represented by **Y**

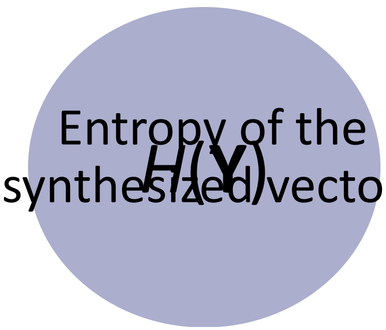
An effective visualization should represent most information in the data, i.e. $H(\mathbf{X}|\mathbf{Y})$ should be small

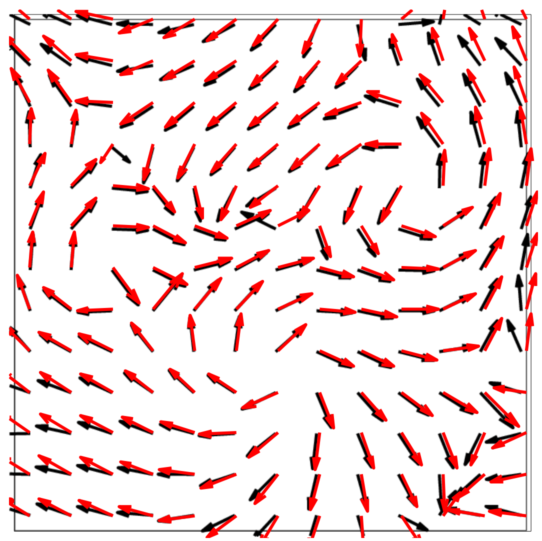
Conditional Entropy and Joint Entropy

Conditional Entropy of both **X** given **Y** = Joint Entropy of both **X** and **Y** – Entropy of **Y**

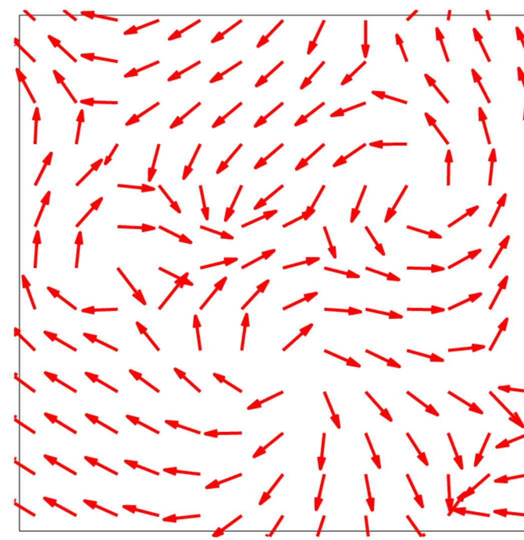








Input vector field

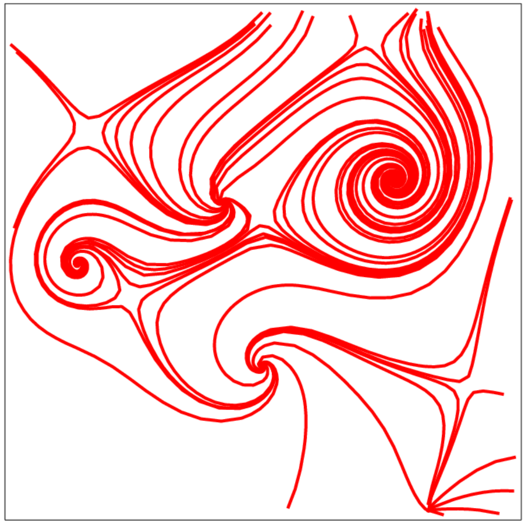


Vector field from the streamlines

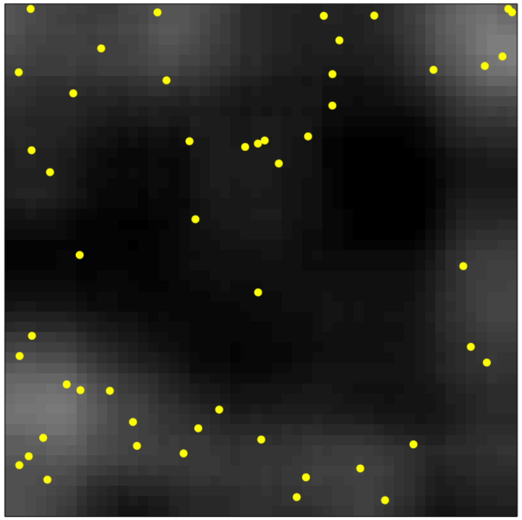
streamlines

Conditional Entropy Field and Seeding

Measure the under-represented information in local regions



Streamlines

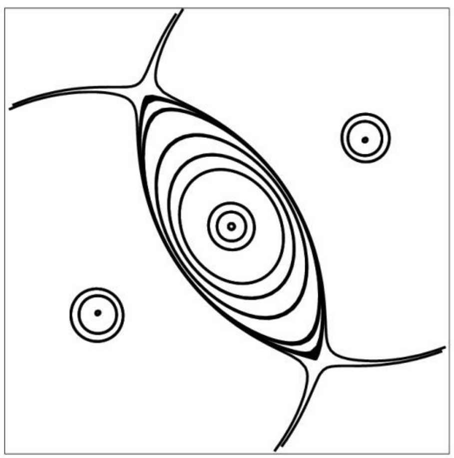


Conditional entropy field

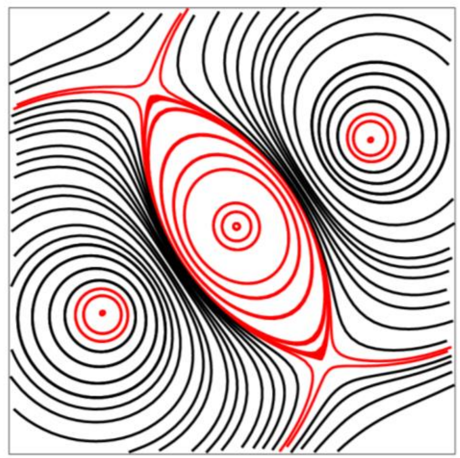
Conditional-entropy-based seeding: Place more seeds on regions with higher under-represented information

Result

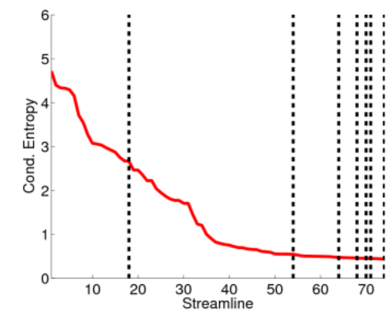
1st iteration: Entropy-based seeding



2nd iteration: Cond.-entropy-based seeding



Conditional entropy



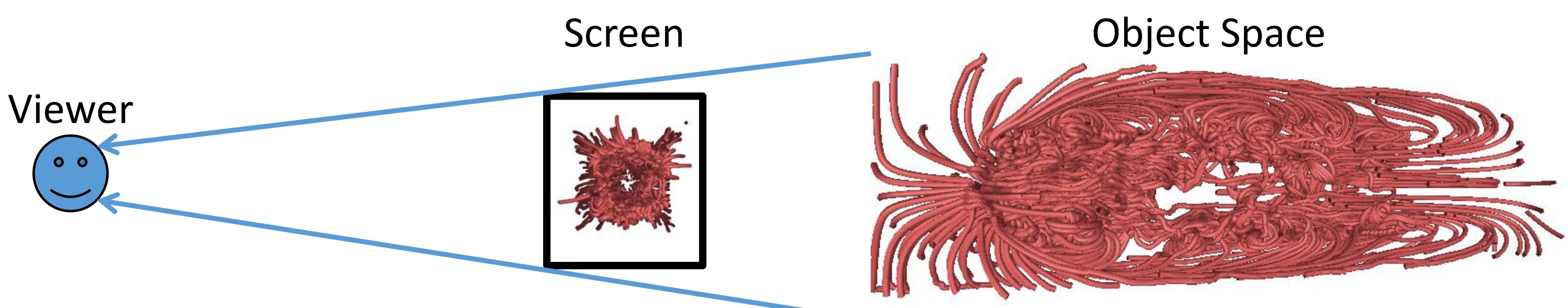
When conditional entropy converges

View-dependent Flow Visualization

- Goal: create a clear view of important features in 3D flow fields by streamline placement
- Issue: occlusion among the flow features
- Approaches
  - Evaluate flow field in screen space by information theory
  - Place streamline to highlight salient flow features with less occlusion

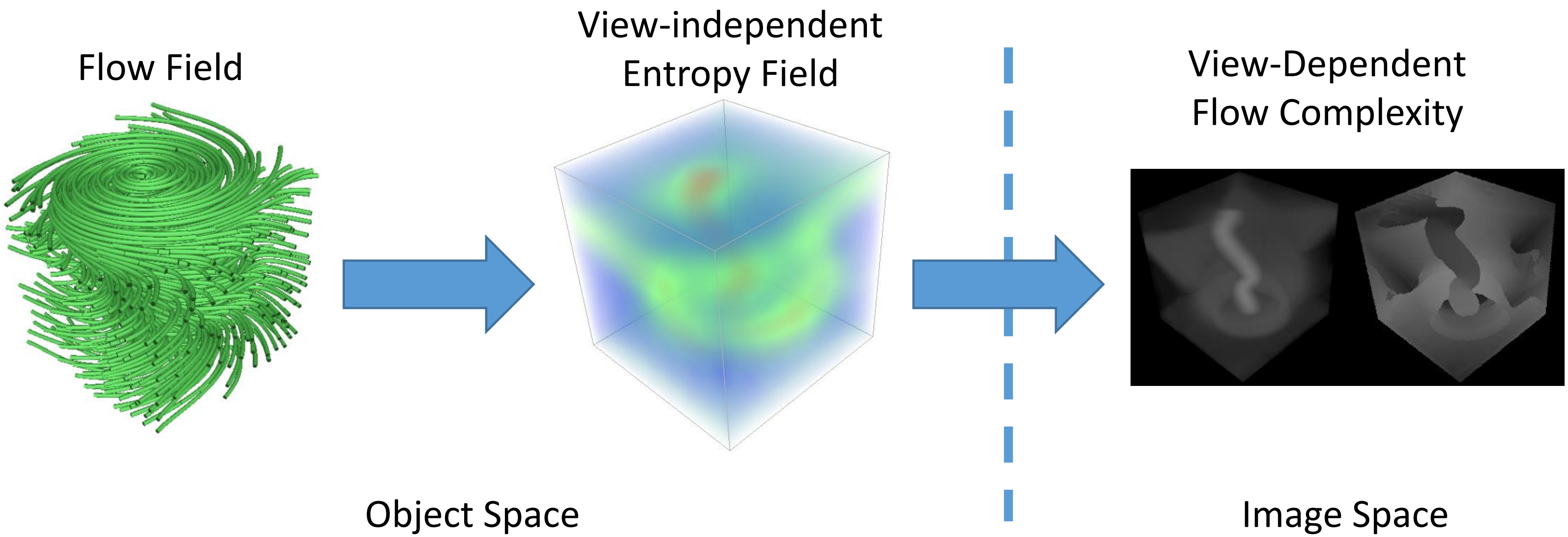
Image-Space Flow Complexity

- Goal
  - Measure the flow complexity on the screen
  - Not trivial because multiple flow features can overlap on the screen
- Approach: consider the most complex flow features visible from the given viewpoint



If the salient flow features are self occluded, only a subset of the them are visible

Flow Complexity Evaluation

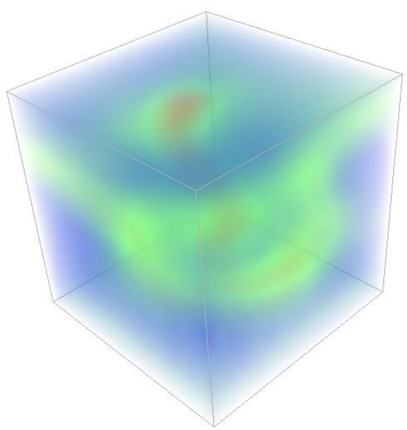


Maximal Entropy Projection (MEP)

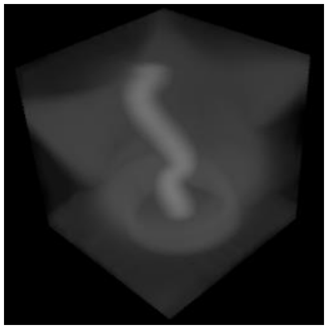
MEP: Project the entropy field to the screen via Maximal Intensity Projection (MIP)

- Sample the maximal entropy visible to each pixel
- Store the sampled entropy and depth in the MEP Framebuffer

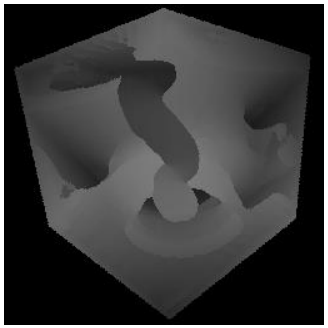
Entropy Field



MEP Framebuffer



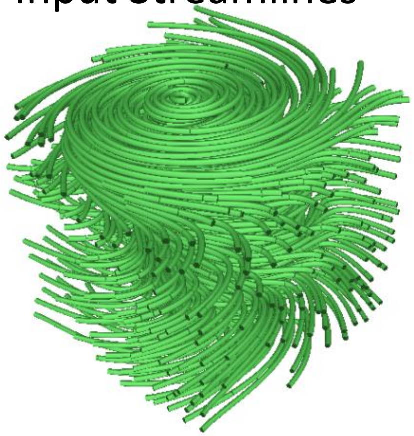
Entropy



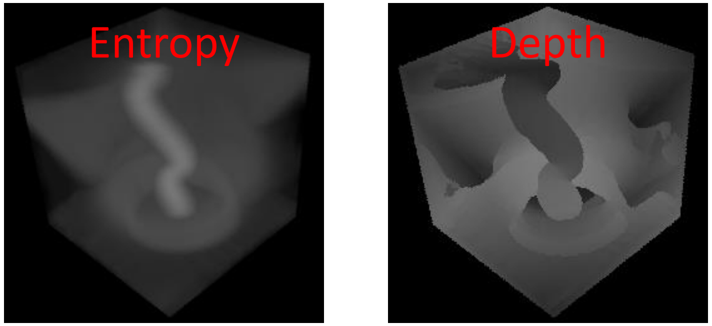
Depth

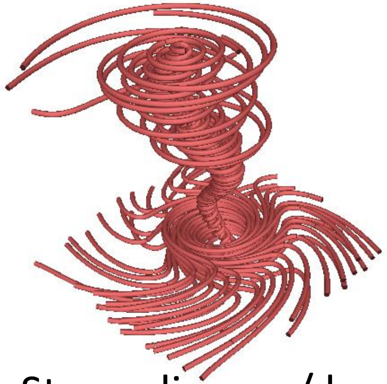
Streamline Evaluation

Input Streamlines

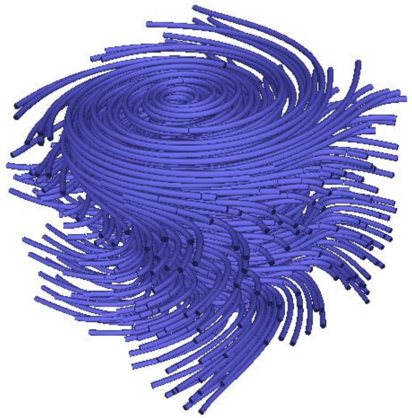


MEP Framebuffer





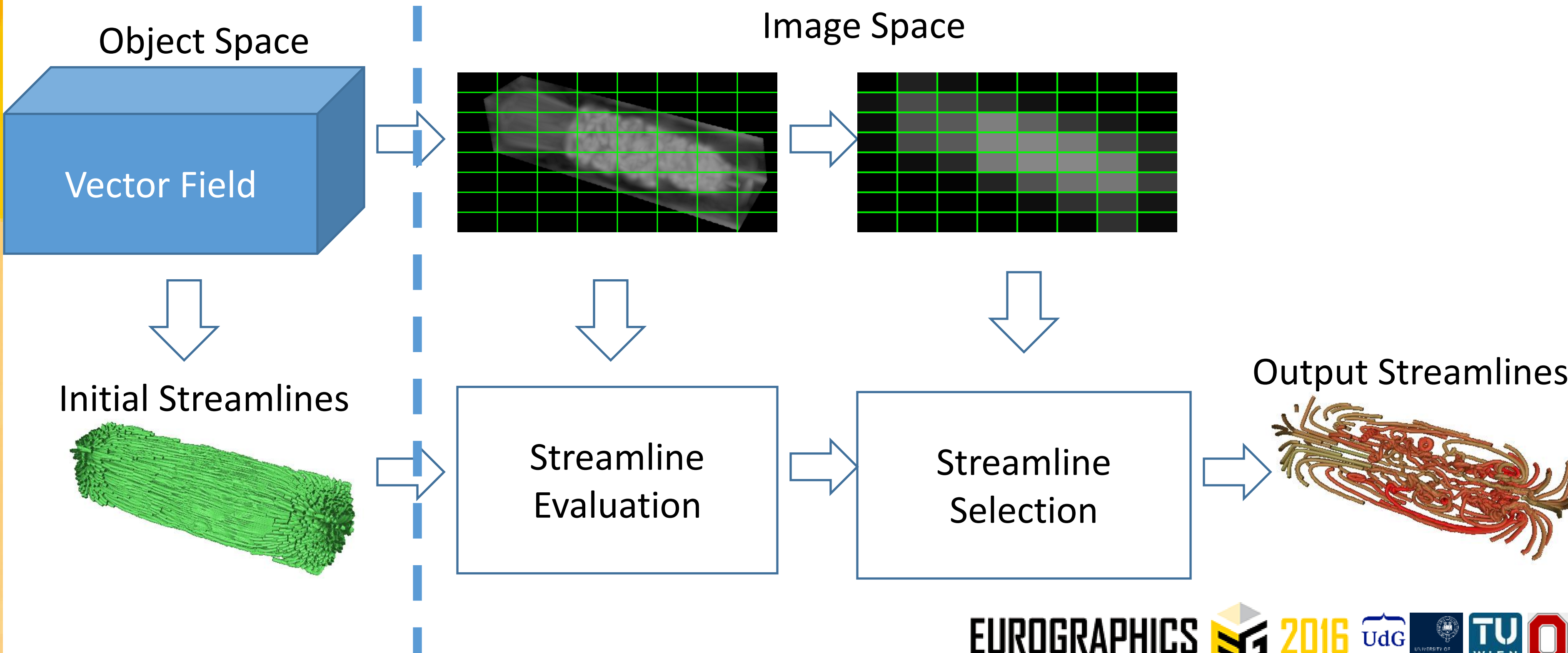
Streamlines w/ less occlusion to the MEP Framebuffer



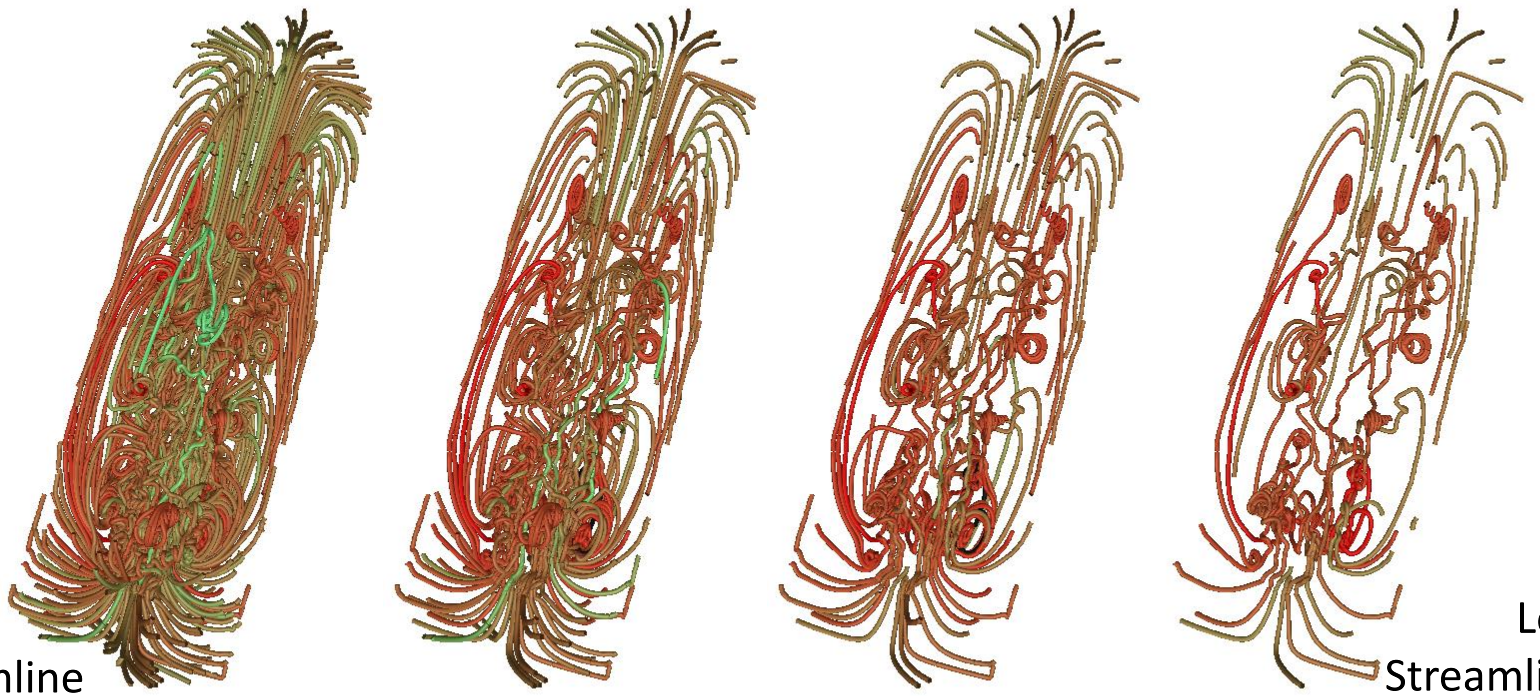
Streamlines that occluded to the MEP Framebuffer

MEP-based Streamline Placement

- Highlight salient flow features
- Reduce occlusion to these features



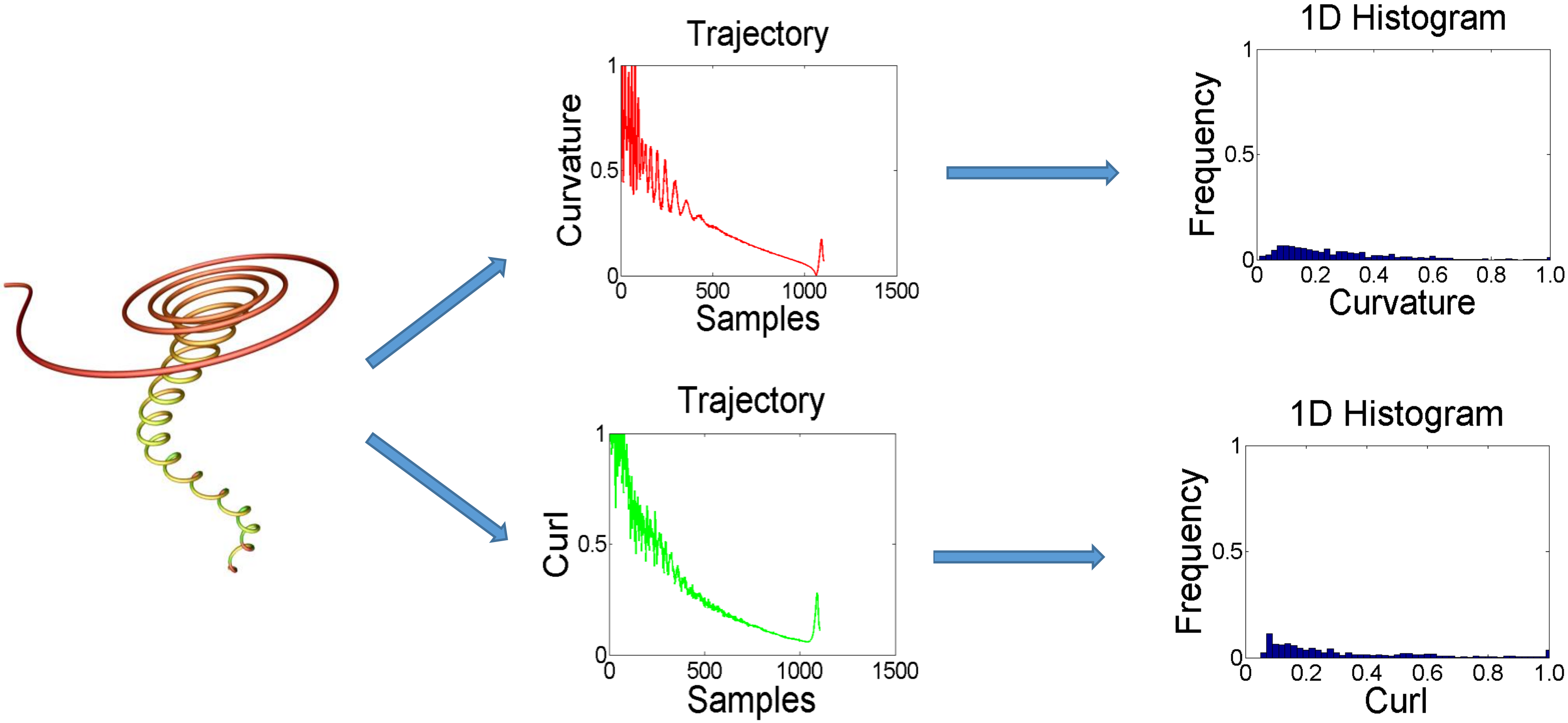
MEP-based Streamline Placement



High Streamline Density → Low Streamline Density

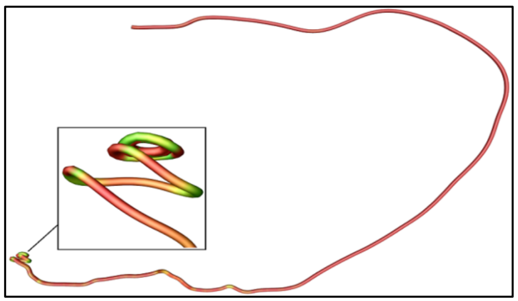
Streamline Statistical Feature Descriptors

- Each streamline is represented as one or more distributions of feature measures such as curvature, curl and torsion

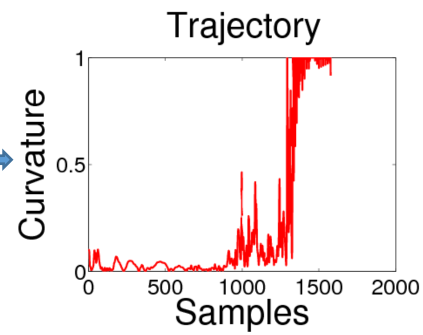


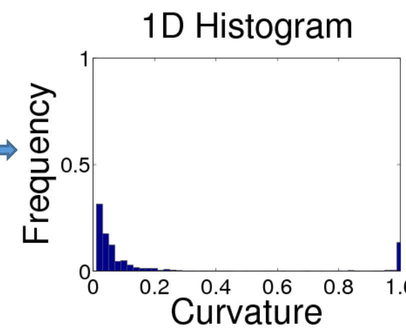
Streamline Statistical Feature Descriptors

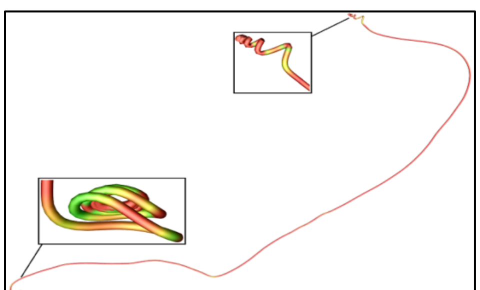
- Problem of 1D histograms
  - The order of features is not preserved in the final histogram



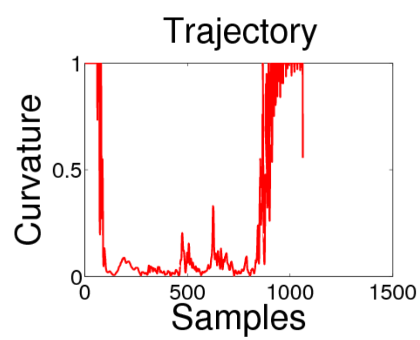
A streamline with
only one high curvature zone

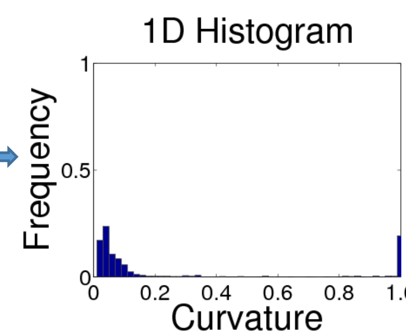






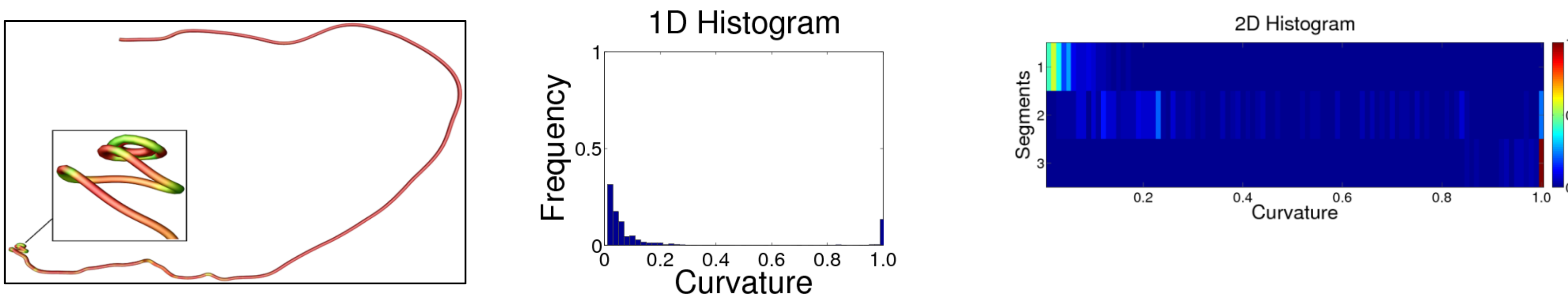
A streamline with
two high curvature zone

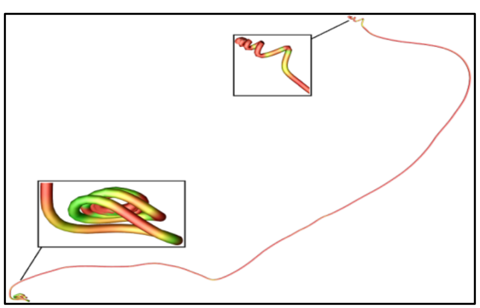


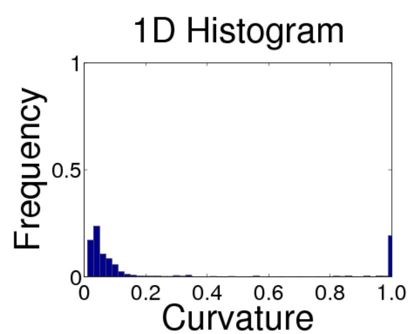


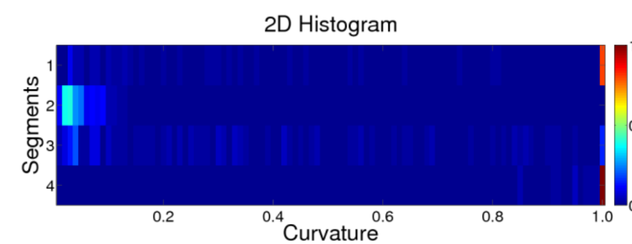
Streamline Statistical Feature Descriptors

- Solution: 2D Histograms
  - Decompose the streamline into a fixed number of segments
  - Create 1D histogram of appropriate quantity for each segment
  - Stack the 1D histograms to form a 2D histogram which preserve the order between segments





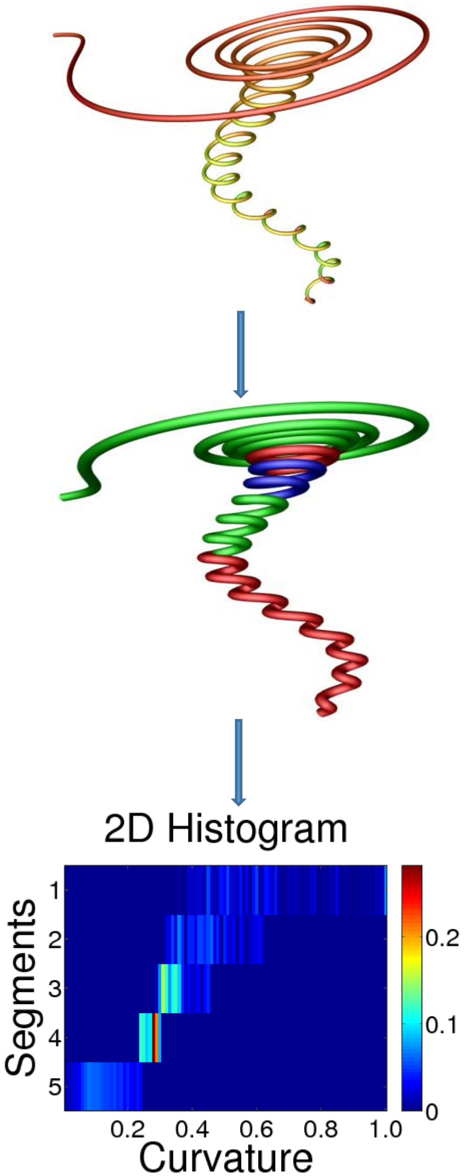




EUROGRAPHICS 2016 UdG UNIVERSITY OF OXFORD TU WIEN

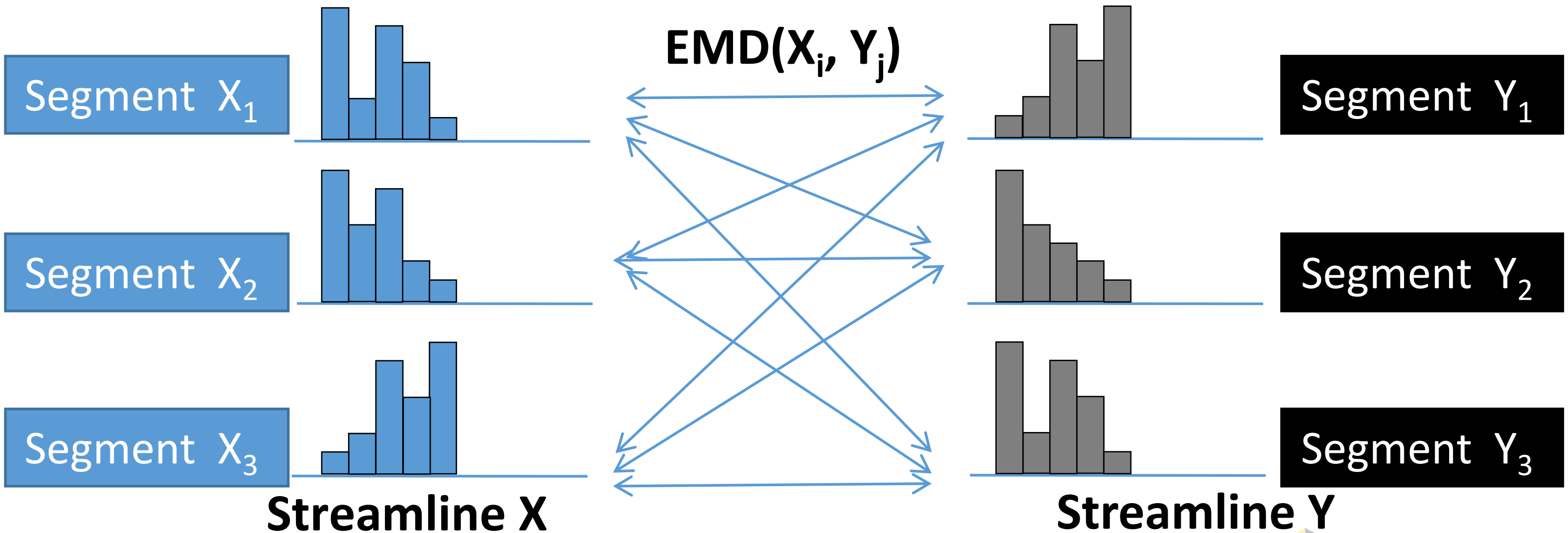
Streamline Decomposition

- An iterative segmentation algorithm
- Recursively divide into segments until:
  - ➢The difference in the 1D histograms between two halves is smaller than a threshold
  - ➢Streamline segment is too short to be further segmented



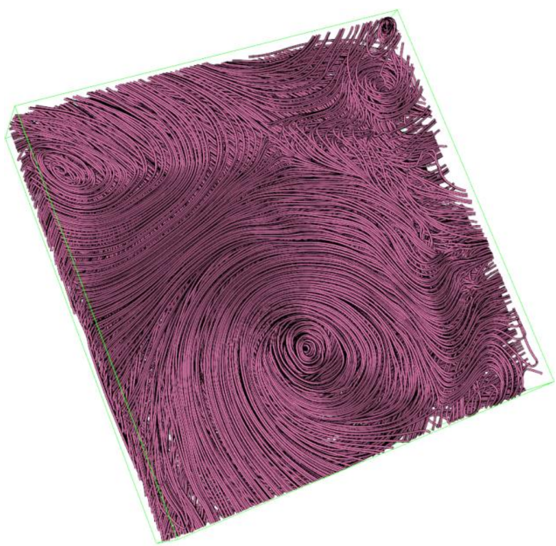
Measure Similarity Between Two Streamlines

- Compute similarity between the 2D histograms of two streamlines
  - As two streamline have different number of segments,
    - Apply **Dynamic Time Warping (DTW)** to find an optimal mapping between segments
  - For each pair of segments,
    - Use **Earth Mover's Distance** to measure the distance of their 1D histograms

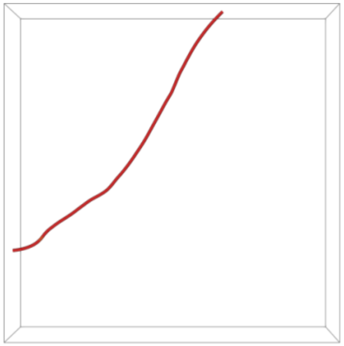


Similarity-based Streamline Query
(Hurricane Isabel Data Set)

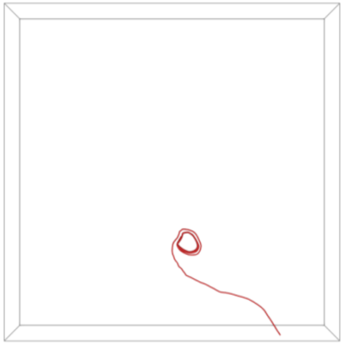
- Streamlines having similar features as the one selected by the user are displayed to highlight features in the data
- Histograms based on Curvature and Torsion are used to answer query in this particular case



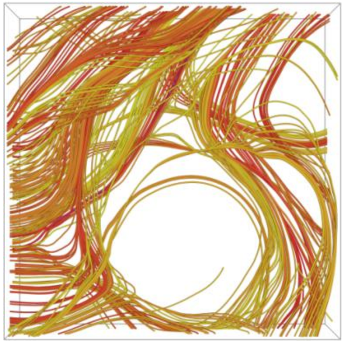
Hurricane Isabel



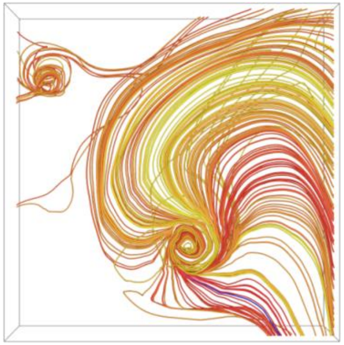
User selected target



User selected target



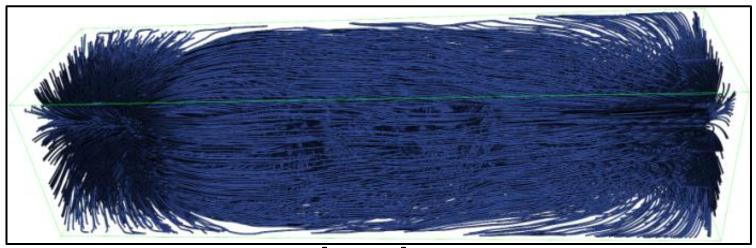
Top 400 matches



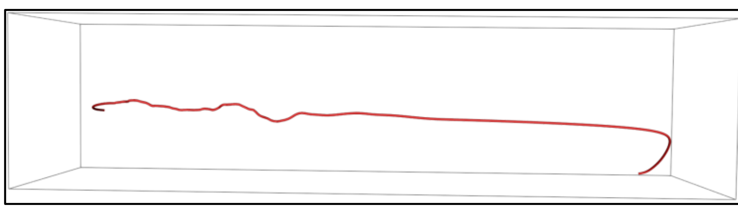
Top 200 matches

Similarity-based Streamline Query
(Solar Plume Data Set)

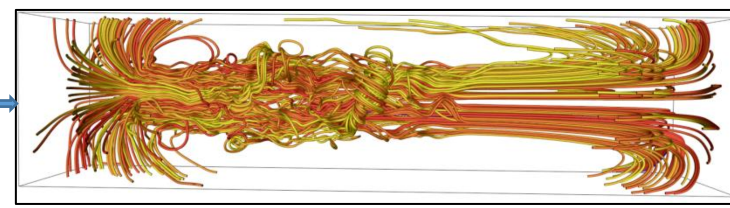
- Query response using curvature and torsion based histograms



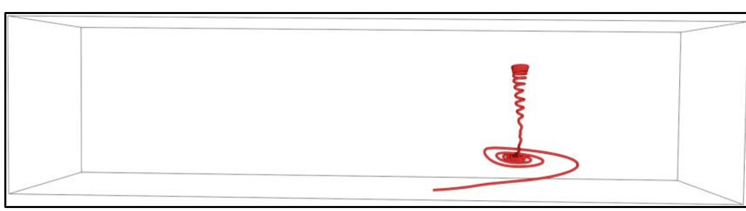
Solar Plume



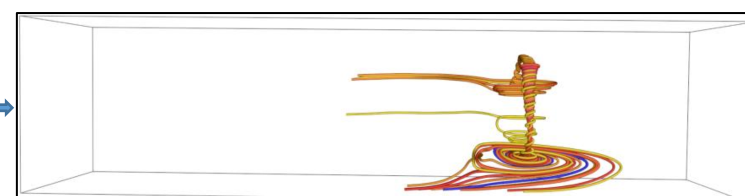
User selected streamline



Top 200 matches



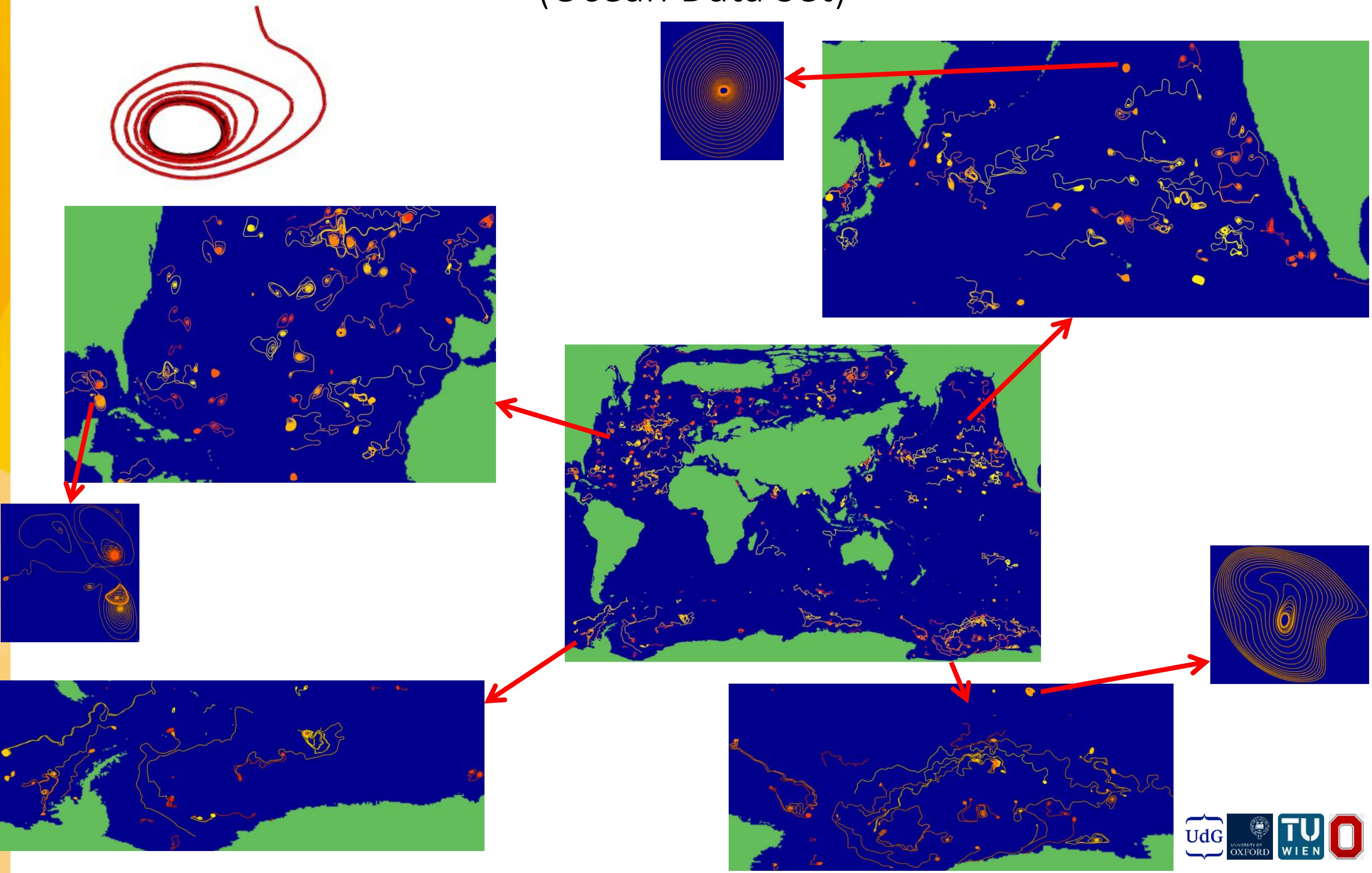
User selected streamline



Top 20 matches

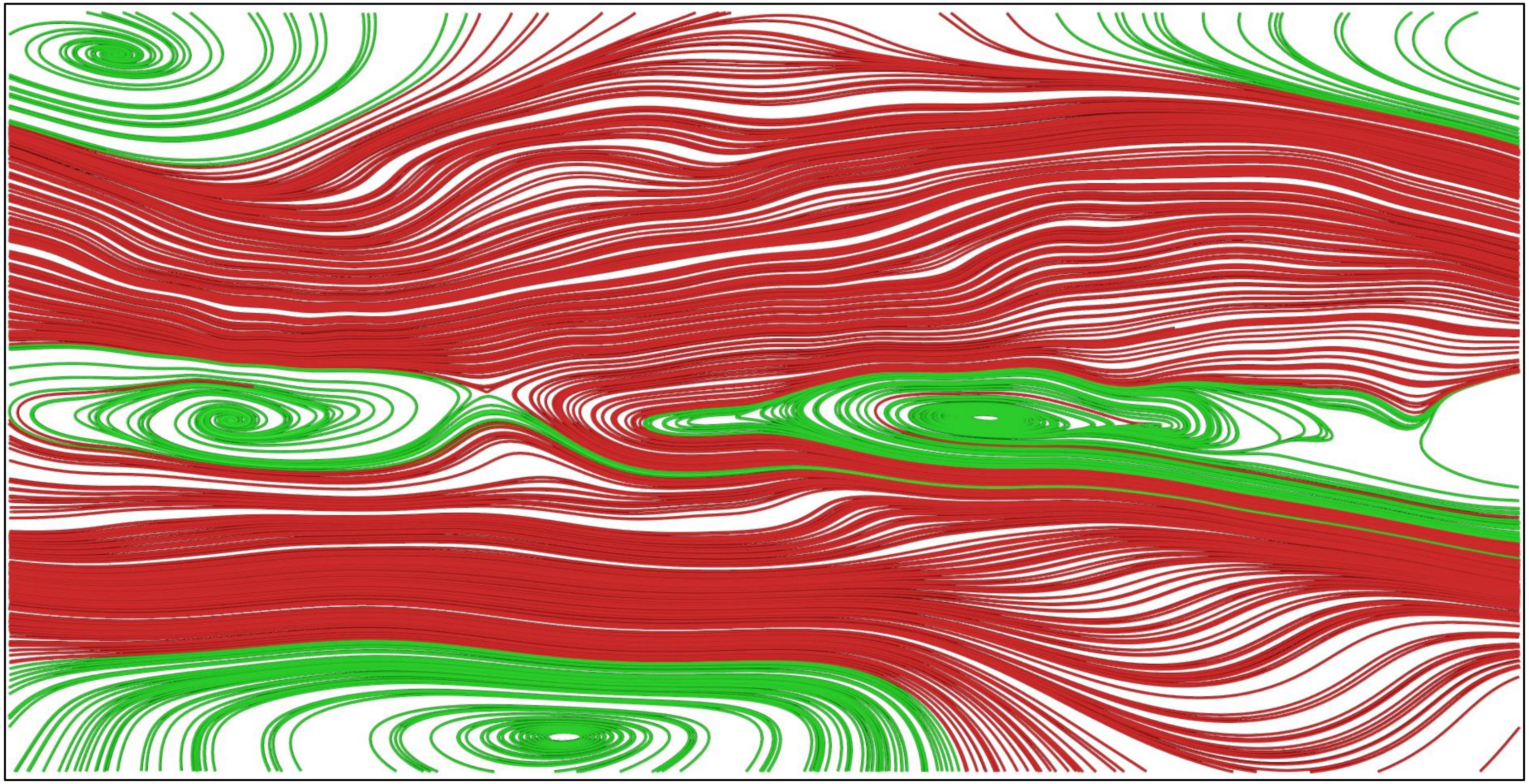
Similarity-based Streamline Query

(Ocean Data Set)



Streamline Clustering

- Clusters are formed based on curvature distribution
- Vortices and linear regions are in two different clusters



2D Ocean Wind dataset