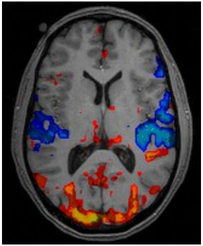
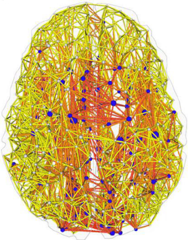
Topological Analysis of Modern Biomedical Data

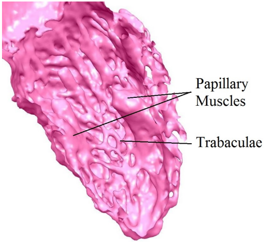
Chao Chen

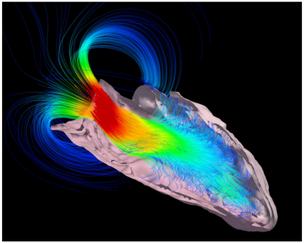
Rutgers University
City University of New York (CUNY)

July 2015

Modern: Huge Amount, High Quality, High Complexity



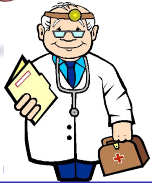


- Climate data, Social science, Genomic data, etc.

Modern Data

- Learning: e.g. better diagnosis/prediagnosis
- Exploration: discover new knowledge



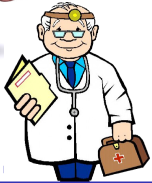
Modern Data

- Learning: e.g. better diagnosis/prediagnosis
- Exploration: discover new knowledge

Structure is the key!

- Key questions:
  - Enable domain experts to effectively explore the data, data-driven
  - Communicate learning results effectively, new hypothesis



Topology Data Analysis

Structures

- Clusters, linear (e.g. PCA), nonlinear (e.g. manifold), etc.
- Topology: global, multiscale, nonparametric, robust to noise

Topology Data Analysis

Structures

- Clusters, linear (e.g. PCA), nonlinear (e.g. manifold), etc.
- Topology: global, multiscale, nonparametric, robust to noise

Interdisciplinary

- Computer science, mathematics, statistics, etc.
- Within computer science: theory, machine learning
- Application: vision, natural language processing, biomedicine, etc.
- Industry: Ayasdi ($95 million)

Outline

Topology 101

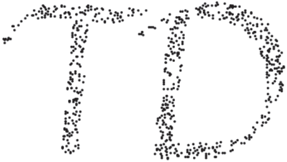
Topology: the science of structures.

- The most global structural information.
- Forgetting local deformation.

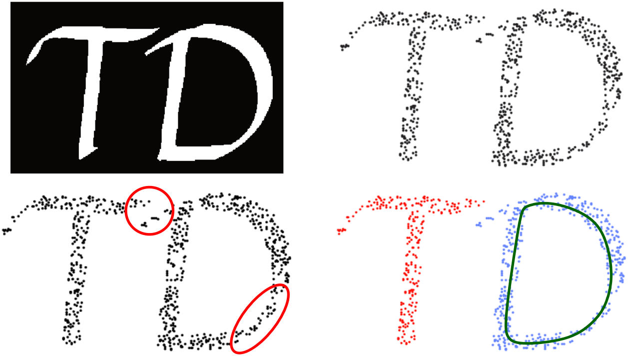
0 dim: components | 1 dim: loops | 2 dim: voids

Topological Structures of Data (TDA)

- For a dataset, what are the components and loops of the data?
- TDA: detect these structures in a robust way.

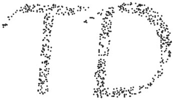


- For a dataset, what are the components and loops of the data?
- TDA: detect these structures in a robust way.



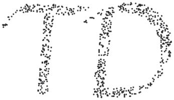
Persistent Homology

- Input: a density/image function, f
- Output:
  topological structures & their persistence

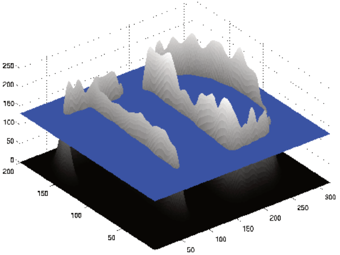


Persistent Homology

- Input: a density/image function, f
- Output:
  topological structures & their persistence



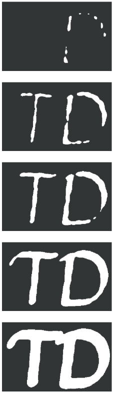
- Def: given threshold t, the superlevel set $f^{-1}[t, +\infty) := \{x | f(x) \geq t\}$

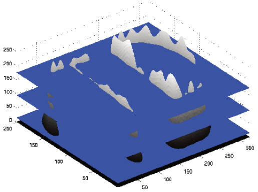


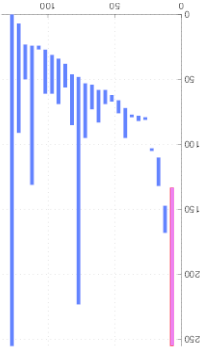
Persistent Homology (continued)

- the true structures are hidden in superlevel sets
- consider the whole stack of superlevel sets
- identify structures that often appear (high persistence)
- Output: persistence diagram – dots representing all structures

Demo



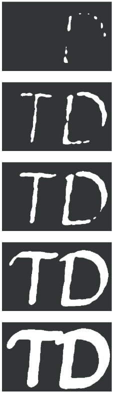


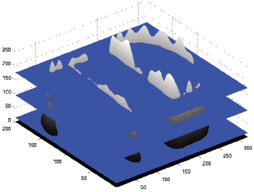


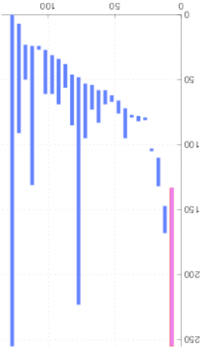
Persistent Homology (continued)

- the true structures are hidden in superlevel sets
- consider the whole stack of superlevel sets
- identify structures that often appear (high persistence)
- Output: persistence diagram – dots representing all structures

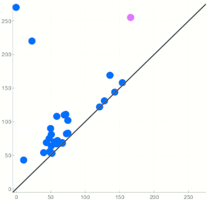
Demo







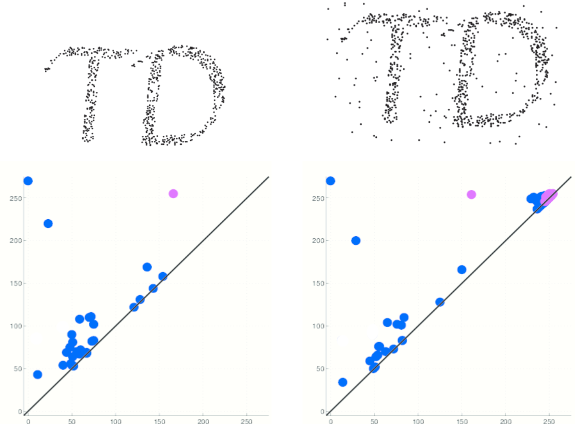
Diagram



Theoretical Guarantee: Stability

The Stability Theorem [Cohen-Steiner *et al.* DCG 2007]

$$\|diagram(f) - diagram(g)\| \leq \|f - g\|_\infty$$

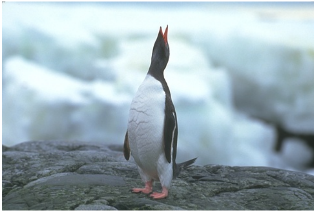


Persistence ≈ robustness to noise

Outline

Segmentation

Important first step for most analysis



- Segmentation (binary): find a labeling $y \colon V \to \{0, 1\}$
  - $y_i = 1$ object, $y_i = 0$ background
  - The *object*: $FG(y) = y^{-1}(1) = \{i | y_i = 1\}$

Energy-Based Segmentation

- How: $y^* = \operatorname{argmin}_y E(y)$
- Energy: Markov Random Field (MRF), active contour, etc.

The Energy

$$E(y) = E_{data}(y) + \omega E_{model}(y)$$

Energy-Based Segmentation

- How: $y^* = \text{argmin}_y E(y)$
- Energy: Markov Random Field (MRF), active contour, etc.

The Energy

$$E(y) = E_{data}(y) + \omega E_{model}(y)$$

- Data term:
  - likelihood, local observation
- Model term:
  - shorter boundary
- Algorithm: mincut-maxflow
  - Augmented path: Dinic 1970, $O(|V|^2 \cdot |E|)$.
  - Boykov and Kolmogorov, PAMI 2004.

Problem: Segmentation with Topological Constraints

Pro: Smoothing effect

Con: Shrinking effect: length vs. area

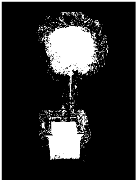
$\omega = 0.0$ $\omega = 0.2$ $\omega = 0.8$ $\omega = 1.4$

Problem: Segmentation with Topological Constraints

Pro: Smoothing effect

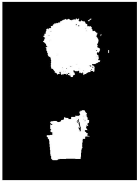
Con: Shrinking effect: length vs. area

$\omega = 0.0$

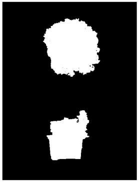


$\omega = 0.2$

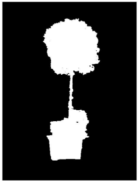
$\omega = 0.8$



$\omega = 1.4$



Question: incorporate the prior information of connectedness?



Algorithm

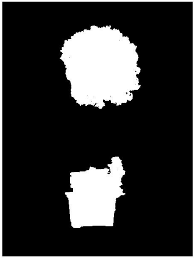
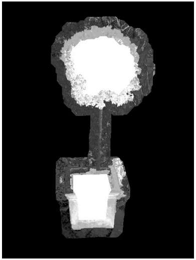
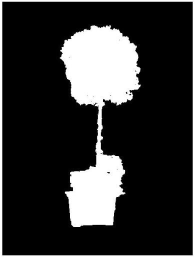
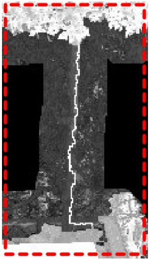
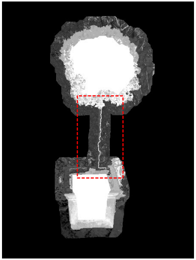




Image → GC → Unaries

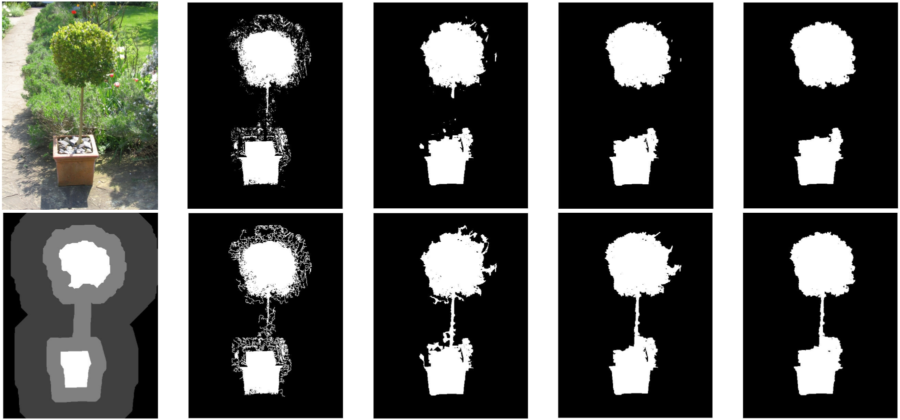
TopoCut ← Zoom In ← Adjusted Unaries

Result 1

- Data: GrabCut (Top: Baseline; Bottom: Our Results)

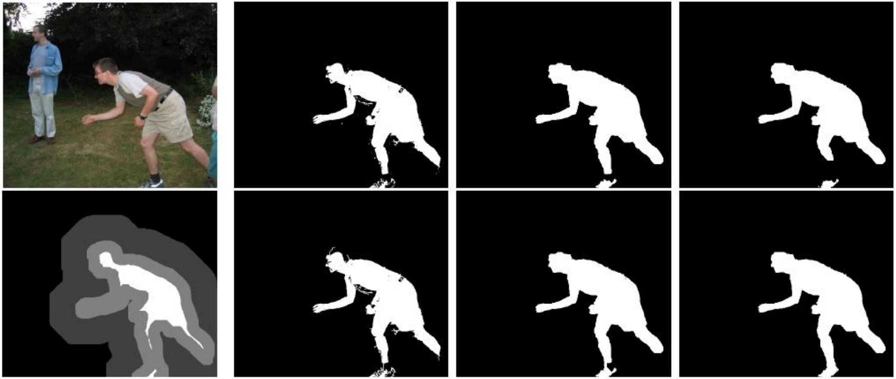
$$\omega = 0.0, 0.4, 0.8, 1.4$$



Result 2

- Top: Baseline; Bottom: Our Results

$$\omega = 0.0, 0.4, 0.8$$



Result 3

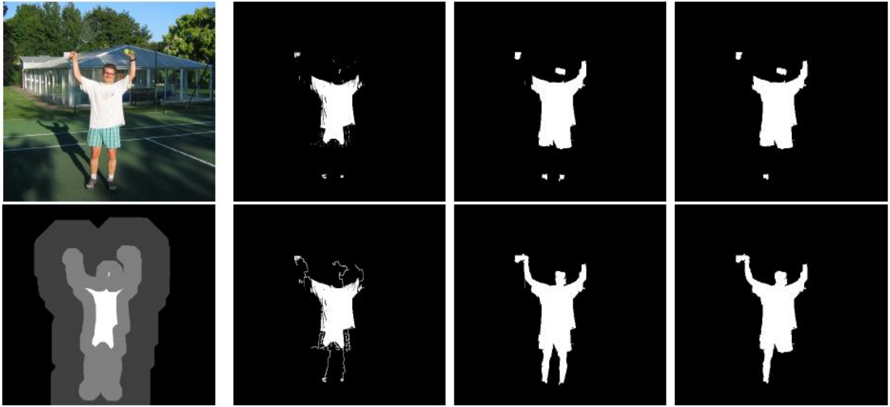
- Top: Baseline; Bottom: Our Results

$$\omega = 0.0, 0.4, 0.8$$

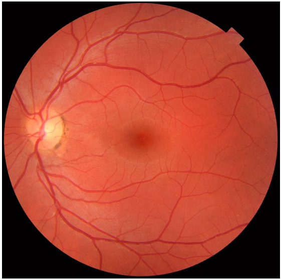
Result 4

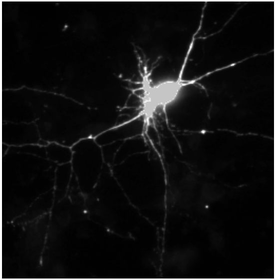
- Top: Baseline; Bottom: Our Results

$$\omega = 0.0, 0.4, 0.8$$



Integrate with Topological Prior

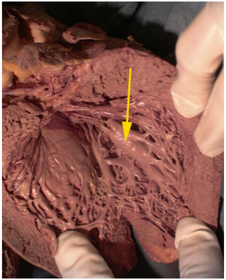


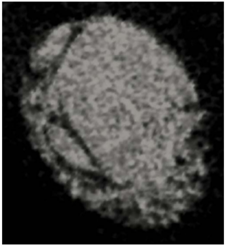


[**Chen** *et al*. CVPR 2011]

High-Order Topological Structures

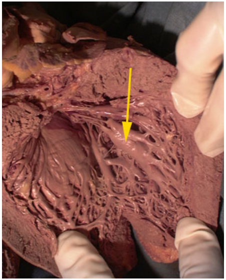
- cardiac data (**Demo** Visible Heart)

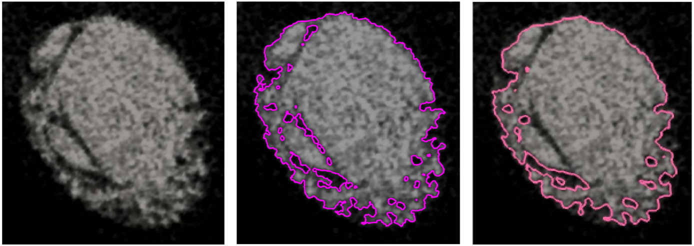




High-Order Topological Structures

- cardiac data (**Demo** Visible Heart)

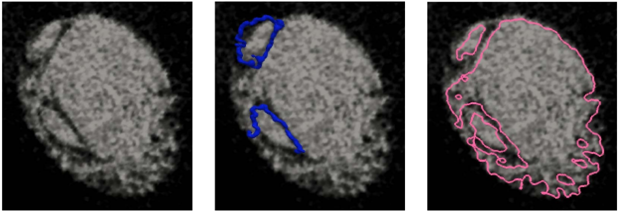


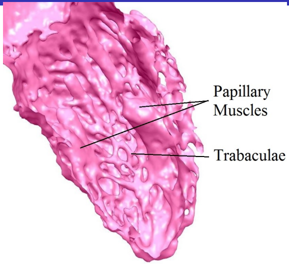


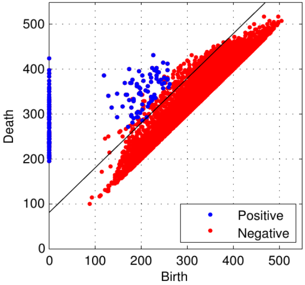
Data Term $\quad + $ Model Term

High-Order Topological Structures

Recovering missing trabeculae:



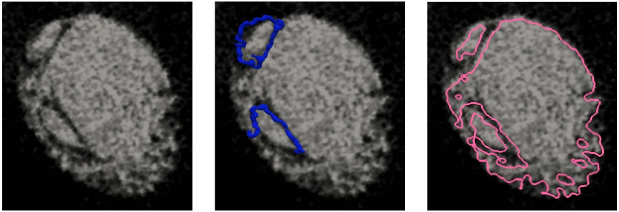


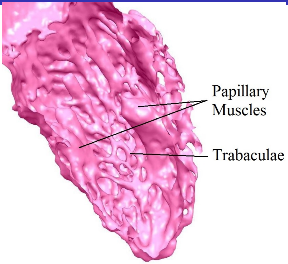


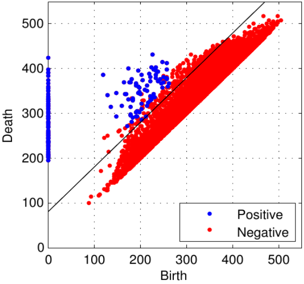
[Gao, **Chen** , *et al.* IPMI 2013]

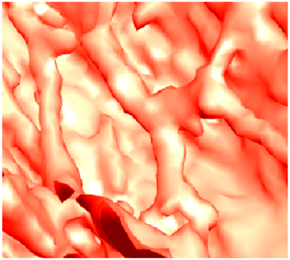
High-Order Topological Structures

Recovering missing trabeculae:

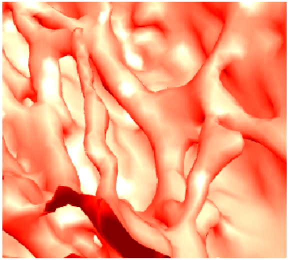








Before



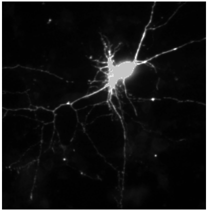
Fixed

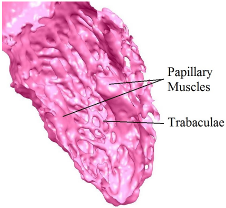
[Gao, **Chen** , *et al.* IPMI 2013]

From Structures to Features

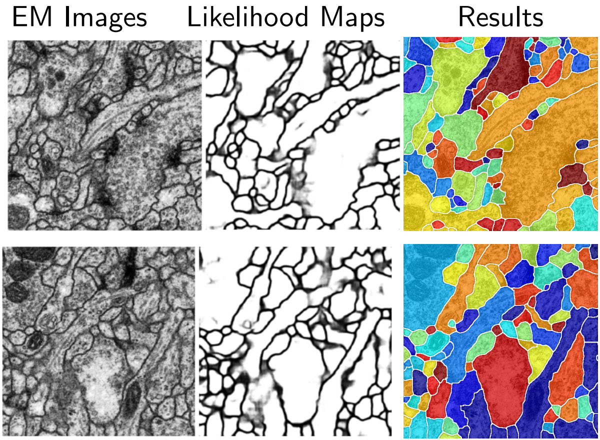
Persistent Homology:
a robust and novel way to extract structures of data

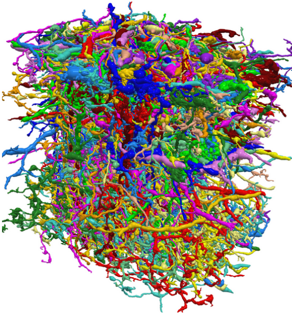
- Global structures, complementing existing tools
- Quantitative: presistence = how likely a structure is true





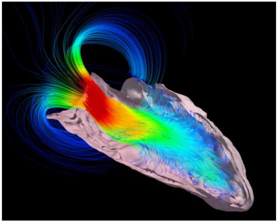
Electron Microscopy Images of Fly/Mouse Brains

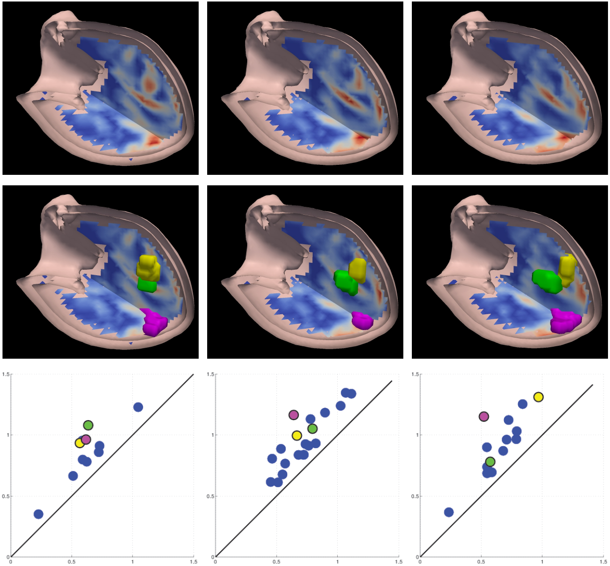




[Uzunbaș, **Chen** and Metaxas, MICCAI'14; MedIA'15]

Cardiac Flow Analysis (Vorticity)



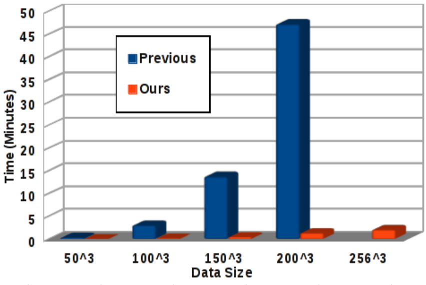


[Kulp, **Chen** *et al.* ISBI 2015]

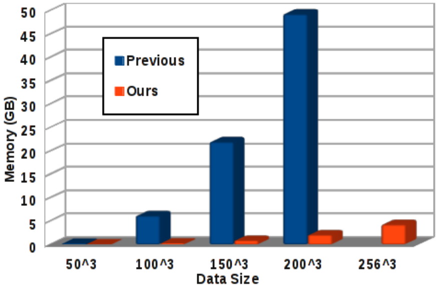
Outline

1. Background: Topology and Persistent Homology
2. Applications: Structures and Features
3. Behind the Applications

- Algorithm
  - [Chen and Kerber SOCG 2011(Top 5); CGTA 2013]
  - [Chen and Freedman SODA'10; CGTA'11; DCG'11]
- Practice: speed and memory efficiency



Time

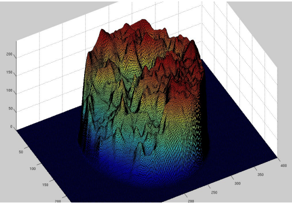


Memory

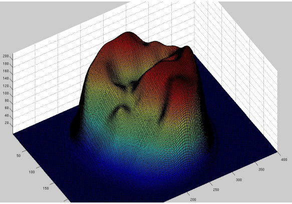
[Chen and Kerber, EuroCG 2011]
[Wagner, Chen and Vucini, TopoInVis 2011 (best paper runner-up)]

Persistence v.s. multiscale

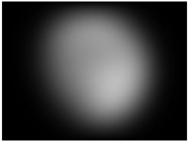
$t = 5$

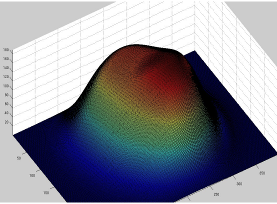


$t = 100$



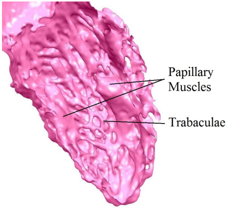
$t = 900$

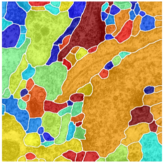
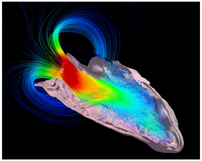




[**Chen** and Edelsbrunner, ICCV 2011]

The End

Thank You!