Edit Distances for Comparing Merge Trees

Raghavendra Sridharamurthy, Adhitya Kamakshidasan, Vijay Natarajan

Visualization and Graphics Lab, Department of Computer Science and Automation,
Indian Institute of Science, Bangalore.

http://vgl.csa.iisc.ac.in; {raghavendrag, adhitya, vijayn}@iisc.ac.in

Problem Statement

Design a distance measure to compare merge trees [CSA 2003].
- Prove theoretical guarantees.
- Provide efficient implementation.
- Applications to time-varying data.
- Applications to feature tracking.

Motivation

Applications
- Topological shape matching.
- Symmetry and similarity detection in scalar fields.
- Feature tracking in time-varying data.
- Comparison between simulated and measured data.

Why merge trees?
- Features in real data are either at loci minibasalima.
- Simple to implement.
- Easy mapping between regions in the domain and tree components.
- Well defined simplification strategy.

Challenges
- Efficiency: Theoretical vs Practical.
- Noise: Small perturbations in the field results in significant changes in the tree structure.
- Guarantees and Properties: Hard to prove
  - Metric properties
  - Stability
  - Discrimination

Background

Tree Edit Distance based Measure

Modified gap model
Categorise the set of edit operations using the properties of merge trees.

1. Permissible set
2. Non-permissible set

Modified cost model
1. Relabel cost $r(i,j)$: Absolute difference in function values.
2. Gap cost $g(i)$: Persistence represented by the pairing.

Tree gap models
- Complete subtree gap
- General subtree gap

Distance
Distance is given by minimum over all such sets of edit operations.

Measure
- $R$ is set of all relabels.
- $G$ is set of all gaps.
- Cost $C = \sum_{(i,j) \in R} r(i,j) + \sum_{g \in G} g(i)$
- Distance $D = \min(C)$ over all allowed edit operations.

Results and Future Work

Periodicity in time-varying data

Data: Bénard von Kármán vortex street, 2D flow around a cylinder; [400 × 50], 1001 timesteps;
Source: Weinkauf [2010].
- Features: Local maxima capture the vortex centres.
- Experiment: Study periodic vortex shedding, with known periodicity of 75.
- Key result: We use our distance measure $D$ and identify periodicity of 74-75.

Detecting symmetry/asymmetry

Data: Synthetic data, contains both regions of symmetry and asymmetry.
- Features: Merge trees of the regions $(i, ..., g)$.
- Experiment: Find whether $D$ is effective in capturing the symmetry/asymmetry.
- Result: $D = 0$ for symmetric regions (for example $D(c,d) = 0$), $D > 0$ in other cases (for example $D(c,b) = 0.53$) which is consistent with the premise of data synthesis.

Future Work
- Prove theoretical properties/guarantees.
- Introduce spatial overlap to enhance discrimination.
- Improve the efficiency, both in theory and in practice.

Acknowledgements
We thank the members of VGL for suggestions and active discussions. We acknowledge support from the Department of Science and Technology (DST/INT/USA-0015-16) and from the JSPS Advanced Technology Programme, Indian Institute of Science (JST/RGC/031014).