Fast Uncertainty-driven Large-scale Volume Feature Extraction on Desktop PCs

Jinrong Xie, Franz Sauer, Kwan-Liu Ma
Introduction: Feature Extraction

• Extensively used in visualization and analysis
  – Explore physical quantities
  – Distinguish/isolate patterns
  – Data reduction
• Large-scale datasets?
Introduction: Large-scale Data

• Drawbacks of distributed computing environments
  – Low interactivity
  – Compete over resources
  – Limited control over tools

• Can we leverage desktop PCs in new ways to handle large-scale data?
Introduction: Overview

- Balance between **accuracy** and **performance**
- Hybrid technique
  - Multiresolution advantages of supervoxels
  - Uncertainty-based refinement
  - GPU accelerated SLIC
Background: SLIC

- Simple Linear Iterative Clustering (SLIC)*
- K-means based
- Limited search area
- Predefined cluster centers
  - Pixel values
  - Location to center

*Achanta et al. 2011
Methods: Supervoxel Generation

• Similar to down-sampling
• Compute relevant information
  – Average intensity
  – Standard deviation
• Use GPU acceleration
  – Supervoxel independence
  – Out-of-core if necessary

\{\text{ave.}, \text{std. dev.}\}
Methods: Supervoxel Clustering

• Extend SLIC to 3D volumes
• Seed initial cluster centers
• User controlled weights
  – Scalar data values
  – Euclidean distance
• Use GPU acceleration

\[ d_{i,k} = w_1 \| c_k - v_i \|_2 + w_2 \| I_k - I_i \|_2 \]
Methods: Hierarchical Merging

• SLIC tends to over segment the volume
  – Number of clusters is fixed
  – Redundant clusters in homogenous regions

• Need to merge clusters

• Need to determine topological connectivity
Methods: Hierarchical Merging

- Running template
- Construct a graph
  - Nodes $\rightarrow$ clusters; edges $\rightarrow$ neighbors
  - Adjacency matrix

![Diagram showing hierarchical merging with nodes and edges connected in a graph, and an adjacency matrix with numbers indicating connections.](attachment:diagram.png)
Methods: Hierarchical Merging

• Merge nodes in graph if they are too similar
Methods: Uncertainty-based Refinement

- Only refine supervoxels with high uncertainty
- Fetch necessary raw data
- Re-cluster new voxels
- Performance vs. accuracy
Results: Combustion Dataset

- S3D, Sandia National Labs
- Highly turbulent and complex features
- 704 x 540 x 550
- Mixture fraction variable

Mixture Fraction Ratio
oxidizer fuel
Results: Combustion Dataset
Results: Combustion Dataset

- Investigate uncertainty-based refinement
- Achieve results that closely match full resolution volume
Results: Ocean Dataset

- POP model, NOAA Geophysical Fluid Dynamics Lab
- Turbulent mixing
- 3600 x 2400 x 42
- Flow velocity mag.
Results: Flow Dataset

- Flow simulation, Argonne National Laboratory
- Columnar and stratified structures of Boussinesq flows
- 4096 x 4096 x 4096
- Vorticity variable
Results: Performance Tests

• Combustion Dataset (512 clusters)

Optimum:
SV Size: $17^3$
Total Time: 187 ms
Results: Performance Tests

• Ocean Dataset (400 clusters)

Optimum:
SV Size: $19^3$
Total Time: 2299 ms
Results: Performance Tests

• Combustion and ocean easily manageable
• Flow dataset?
  – Hundreds of GB per timestep
  – Large I/O cost
• In situ supervoxel generation
Results: Performance Tests

- Emulate in situ case (Titan)

Constant number of nodes (256)
Results: Performance Tests

- Flow Dataset (1000 clusters)

Optimum:
SV Size: $10^3$
Total Time: 9159 ms
Conclusion

• Develop a hybrid feature extraction technique
  – Multiresolution advantages of supervoxels
  – Uncertainty-based refinement
  – GPU accelerated SLIC

• Control over performance and accuracy

• Can interactively handle large datasets on a desktop PC
Acknowledgements

• Data sets:
  – Sandia National Laboratory
  – National Oceanic and Atmospheric Administration
  – Argonne National Laboratory

• Funding sources:
  – U.S. National Science Foundation via grants NSF DRL-1323214 and NSF IIS-1320229
  – U.S. Department of Energy via grants DE-SC0005334, DE-FC02-12ER26072, and DE-SC0012610
Thank You

Questions?