Stony Brook University

SUNY

Analytical 3D Digital Pathology

Fusheng Wang
Department of Biomedical Informatics
Department of Computer Science

In collaboration with Jun Kong, Emory University

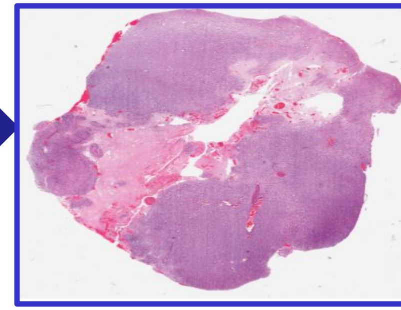
May 26, 2017

Digital Pathology

Stony Brook University

Glass Slides

Scanning



Whole Slide Images

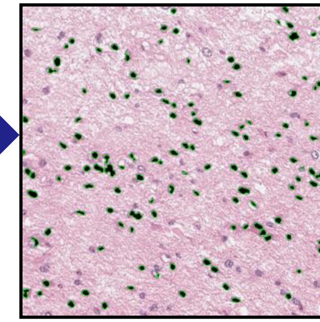


Image Analysis

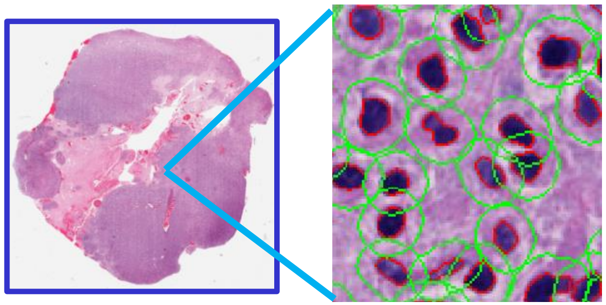
- Digital pathology slides provide rich information about morphological and functional characteristics of biological systems, have tremendous potential for understanding diseases and supporting diagnosis
- FDA recently approved review and interpretation of digital surgical pathology slides

Analytical Digital Pathology for Scientific Research: Examples

Stony Brook University

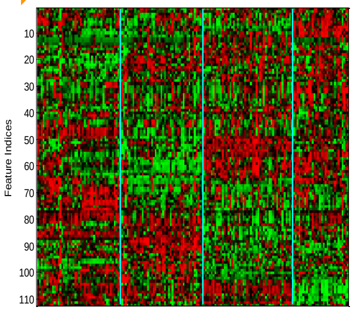
- Do nuclear morphometry features carry prognostic value?

Nuclei/Cell Segmentation

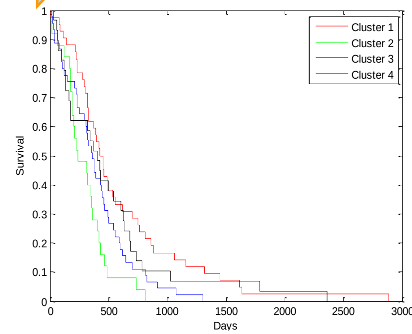


Nuclear Features

Aggregation/ Clustering

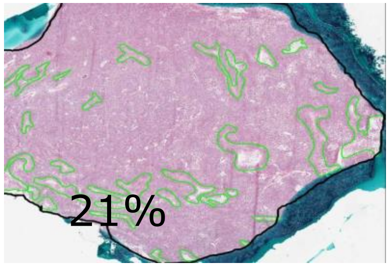


Survival Analysis

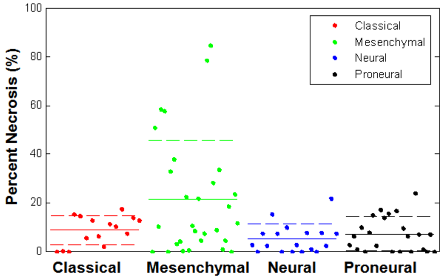


- Do pathologic hallmarks (e.g. necrosis) affect the transcriptional class enrichment?

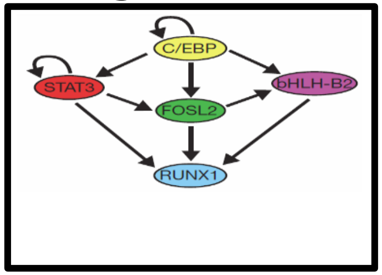
Necrosis Detection



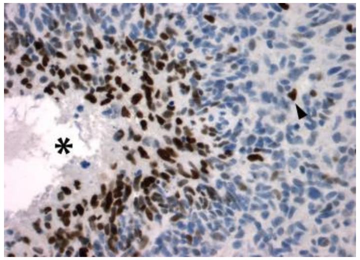
Necrosis perc. vs. trans. class



Master Regulators

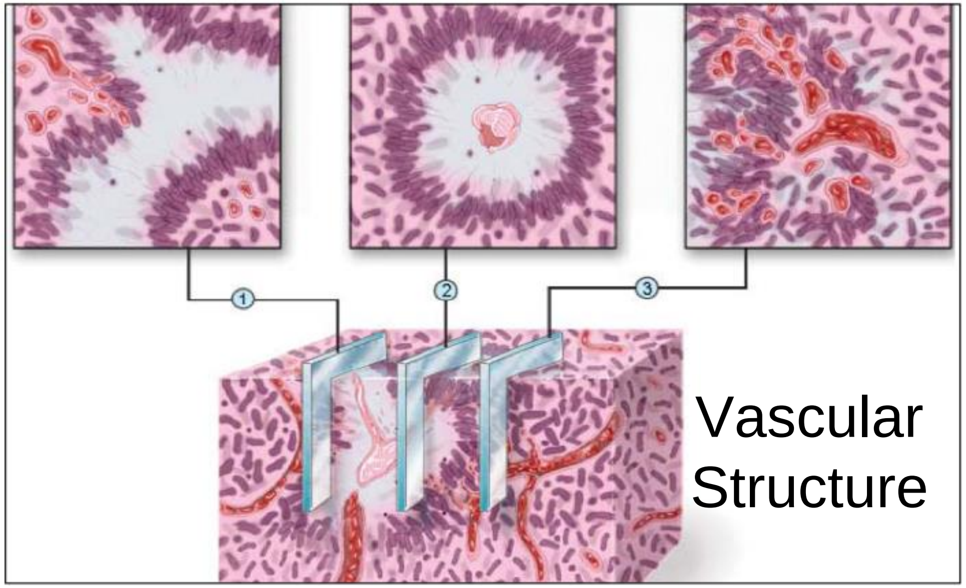


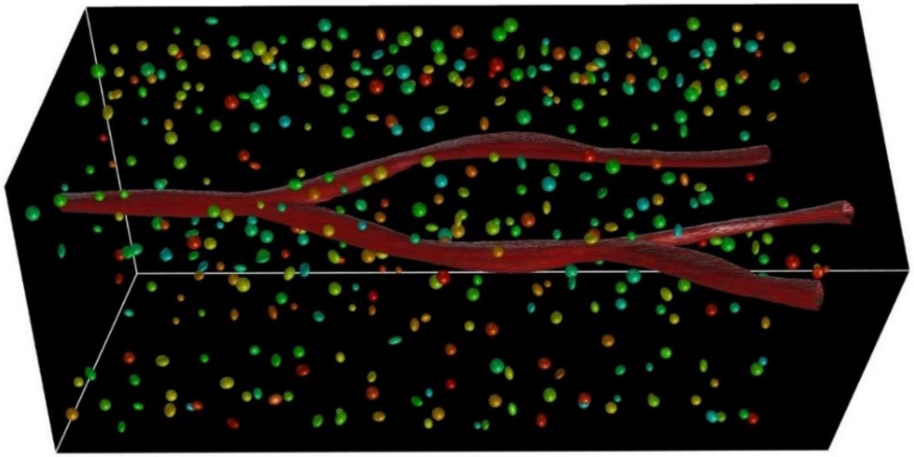
CEBP-β is hypoxia



3D Digital Pathology

- 2D projection depends on random locations and angles of the cutting planes, leading to misrepresented spatial relationships and inaccurate morphological features
- Recently, 3D digital pathology is made possible through slicing tissues into serial thin sections, with significant potential to enhance digital pathology through high throughput 3D image analysis and spatial analytics

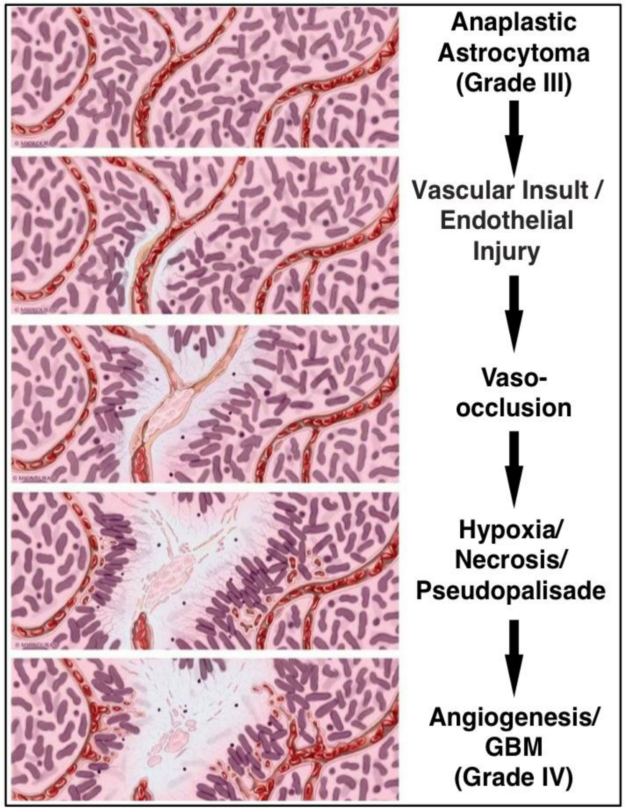




3D Digital Pathology for Tumor Microenvironment

Study of the evolution of tumors will help researchers to develop predictive models of tumor progression and response to treatment that will help oncologists make informed treatment decisions for each patient

- With the development of necrosis, there is a massive restructuring of the TME
- Glioma stem cells (GSCs) and tumor associated macrophages (TAE) promote angiogenesis and glioma progression
- Spatial and temporal analysis could help us understand much better the dynamics of TME



GBM tumor expansion processes

3D Spatial Queries and Analytics

Stony Brook University

Relationship based

- Containment/range: identify only cells of interest contained in or within certain distance from a blood vessel
- Spatial join: compare two spatial datasets to find spatial relationships

Distance based

- Spatial density computation: compute the 3D histogram of cell density in the space
- Nearest neighbor: find the closest vessel for each cell
- Proximity estimation: distributions of different types of objects

Global patterns

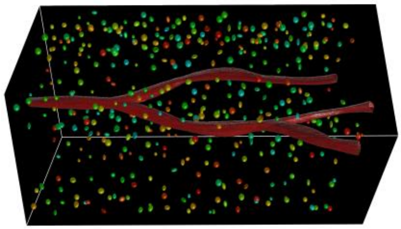
- Spatial clusters
- spatial point patterns

Our Goal: Analytical 3D Digital Pathology

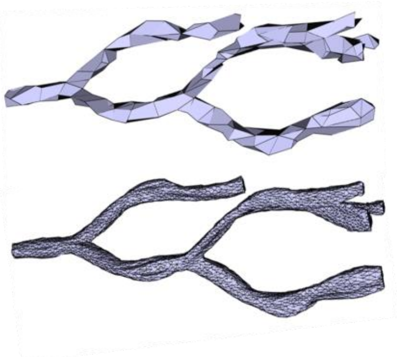
- Quantitative 3D pathology imaging: characterize a set of histopathological objects of common interest to the biomedical research community
  - Including nuclei/cells, blood vessels, lumens, ducts, among others
  - Registration, pathology object segmentation, and 3D reconstruction
- Explore spatial relationships among derived 3D biological objects, and discover correlations across spatial patterns of these biological objects, disease progression, and corresponding genomic signatures
  - 3D data compression, storage, indexing and querying methods

Challenges

- Explosion of 3D data
  - 10 billion pixels per image, tens of or hundreds of slides per volume, tens of millions of 3D objects per volume
  - High I/O and communication cost for data processing



- Complex 3D structures and representations
  - Bifurcations in blood vessel
  - Multiple levels of detail (LOD)



- High computational complexity
  - Heavy duty geometric computation

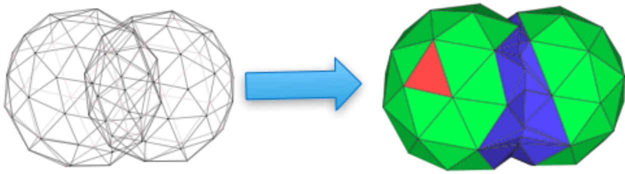


Image Registration (with Jun Kong@Emory)

- Use only small sub-volumes of tissues (disease sites or histology hallmarks) to overcome memory limit
  1. Macro-level global structure registration
  2. Transformation mapping and propagation from low to high resolution
  3. Micro-level local subtle deformation

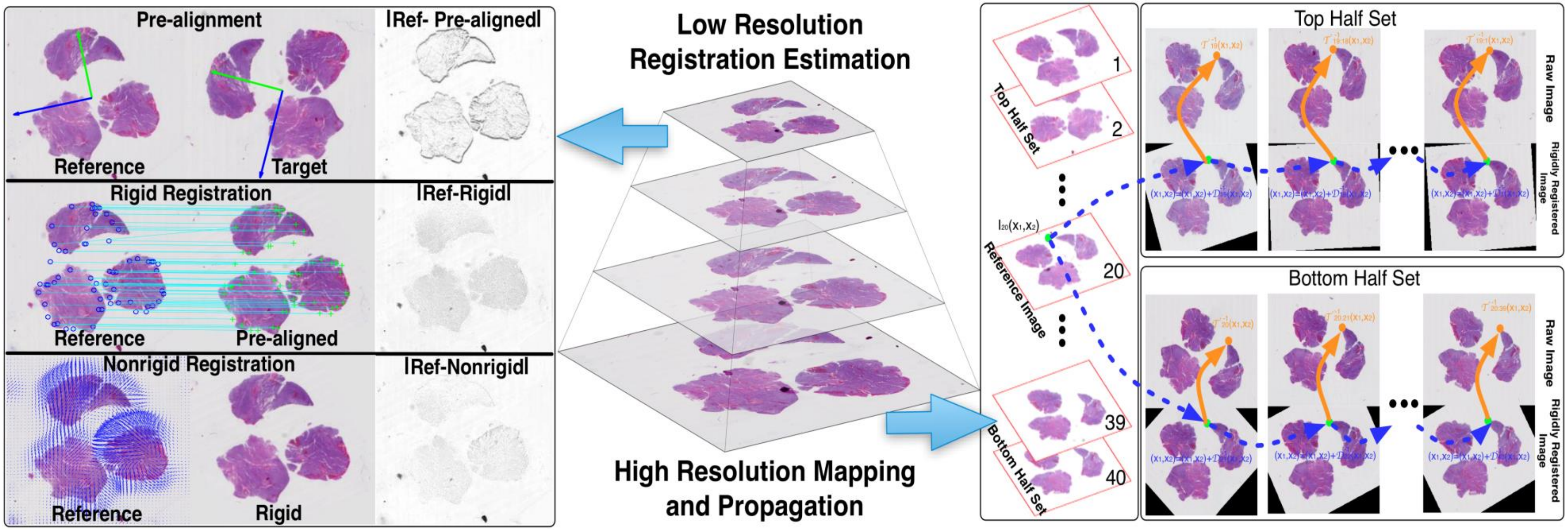
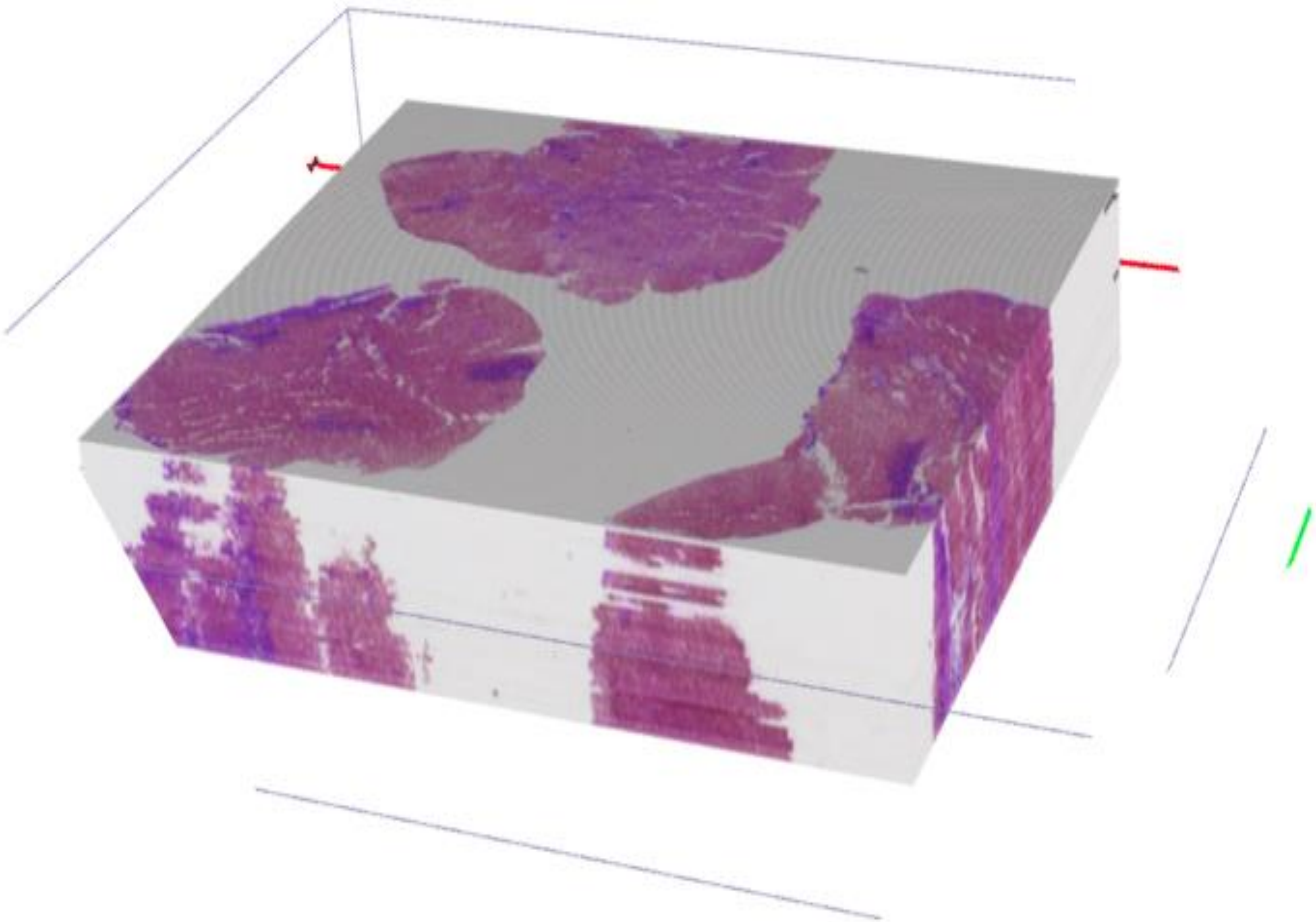
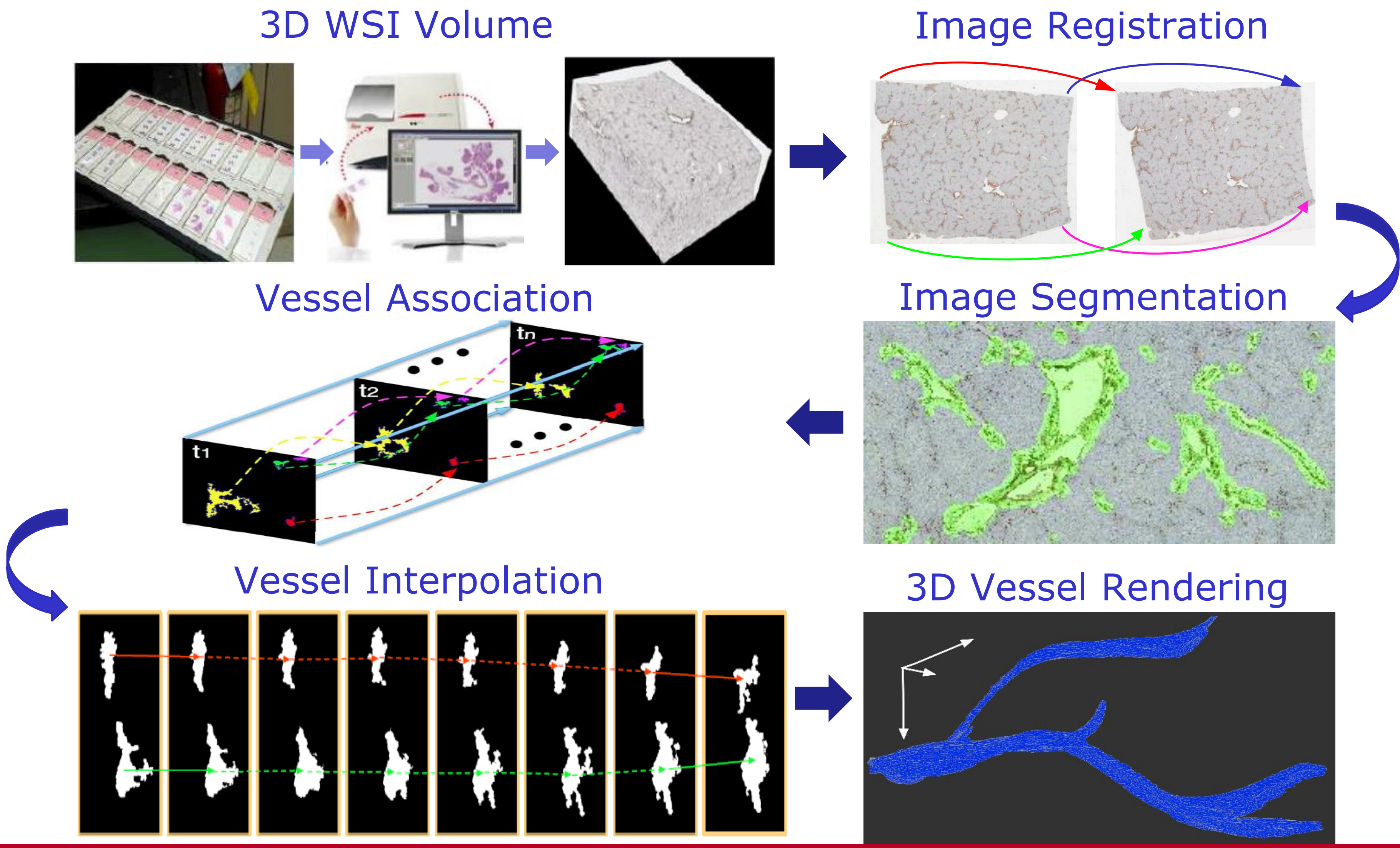


Image Registration Visualization

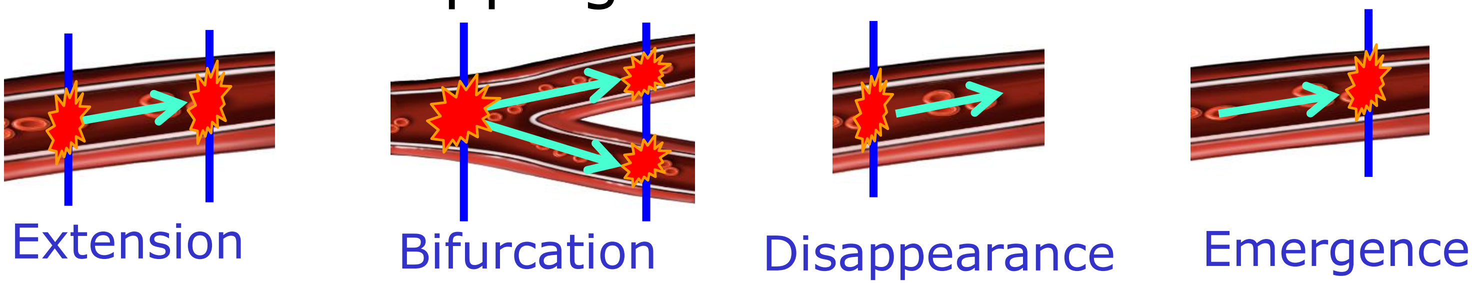


3D Primary Vessel Reconstruction

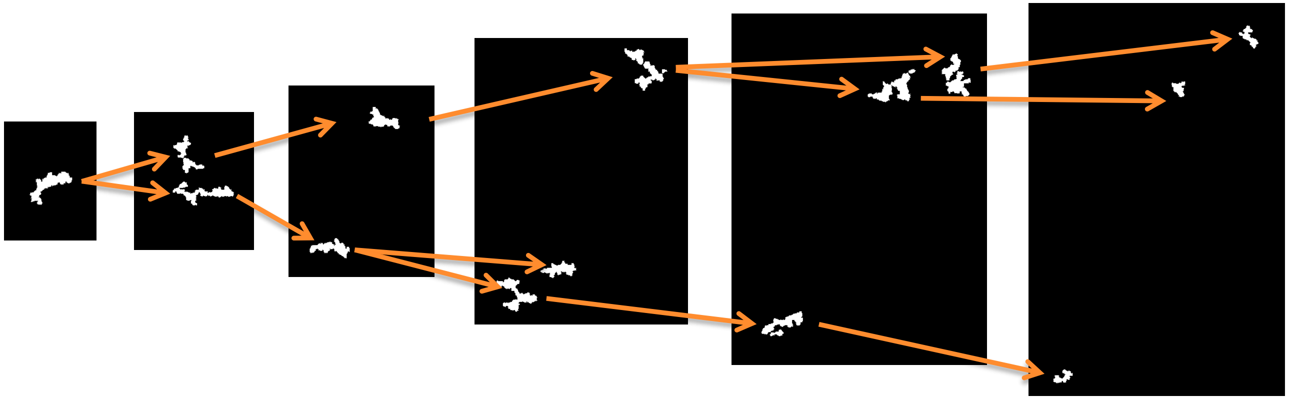


Two-stage Vessel Association

- Local bi-slide mapping: four vessel association cases



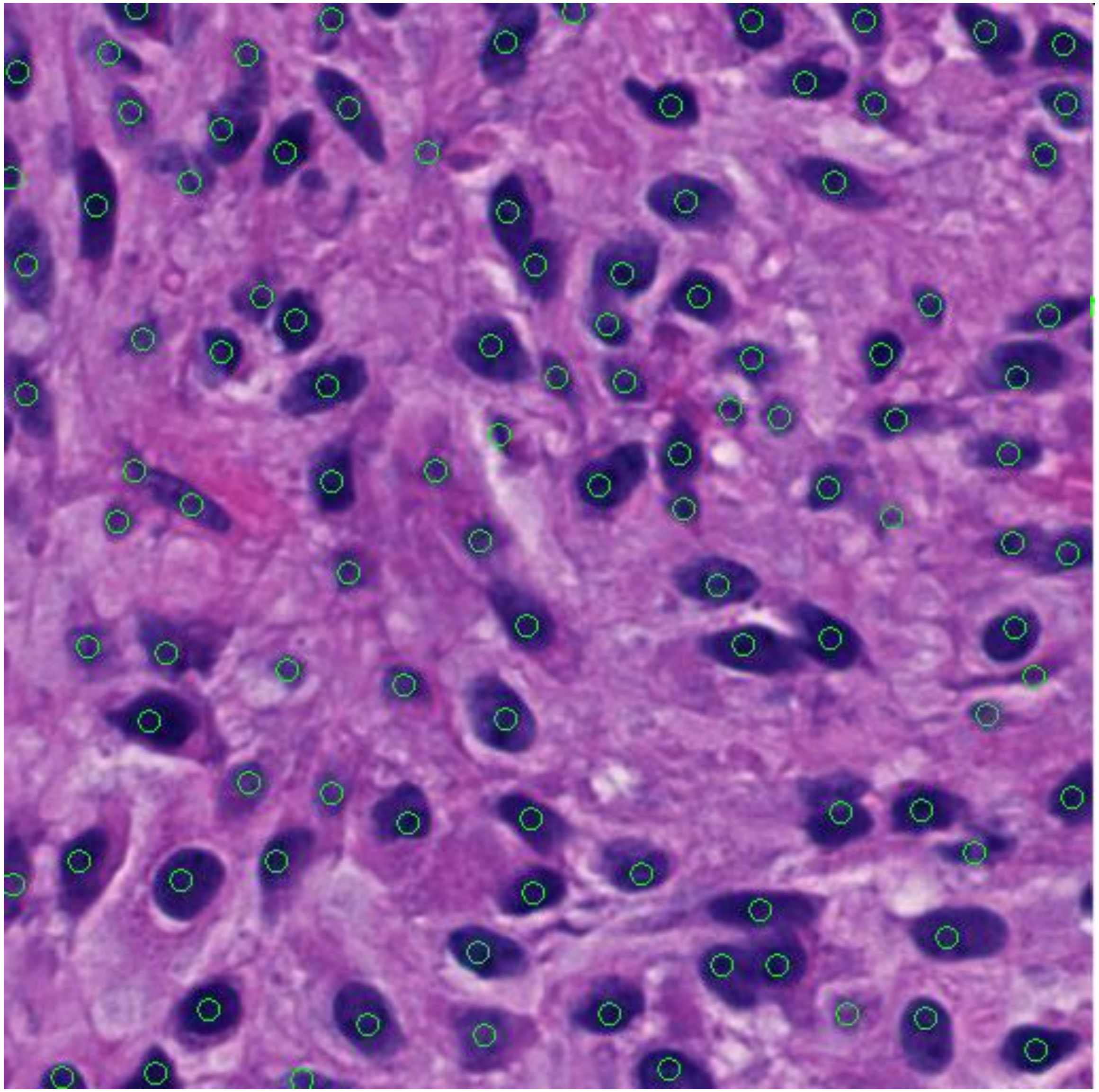
  - Similarity functions
    - Shape descriptor, spatial relationship, trajectory smoothness
- Global vessel structure association
    - Bayesian Maximum A Posteriori (MAP) framework



Nuclear Segmentation

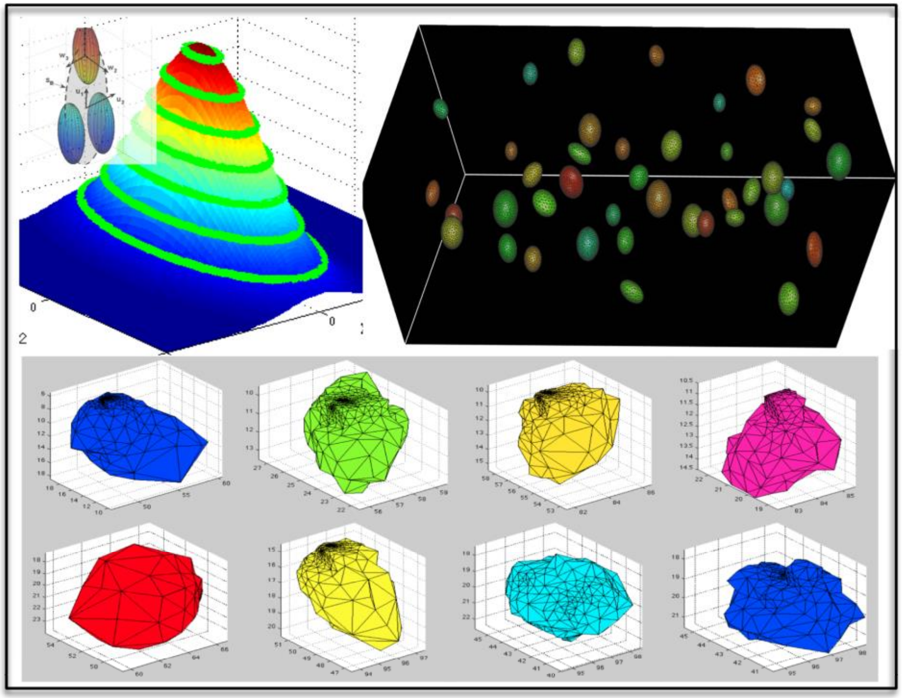
- Segmentation initialization with seed detection

- Iterative contour convergence



Reconstruction Method for Cells

- Reconstruct 3D cells by a cross section association method that solves a geometrical model fitting problem using the statistical multi-variate Gaussian distribution



In-Memory Based 3D Spatial Data Management and Queries

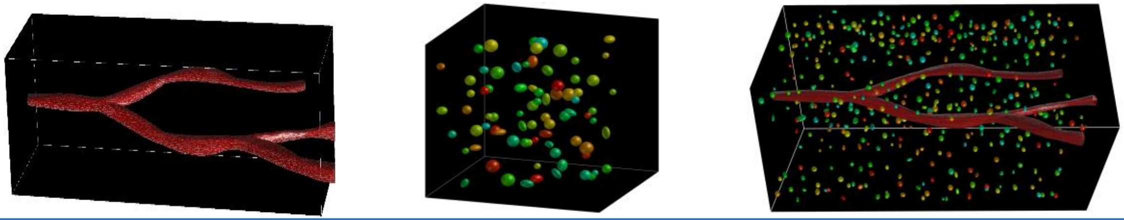
Stony Brook University

- iSPEED (**i**n-memory **sp**atial query system for thr**ee** **d**imensional spatial data)
  - Take advantage of memory to store data and indexes, minimize I/O, communication, and shuffling cost
  - Support effective and scalable spatial queries and analytics on Hadoop and Spark
- Effective progressive compression for individual 3D objects
- In-memory based data storage and indexing
- Multi-level spatial indexing
- On-demand 3D spatial query pipelines scalable for Hadoop/Spark

iSPEED Architecture

Stony Brook University

3D Spatial Data



Distributed File System

Data Pre-processing

- 3D Data Compression
- 3D Spatial Partitioning
- 3D Global Indexing

In-memory Storage

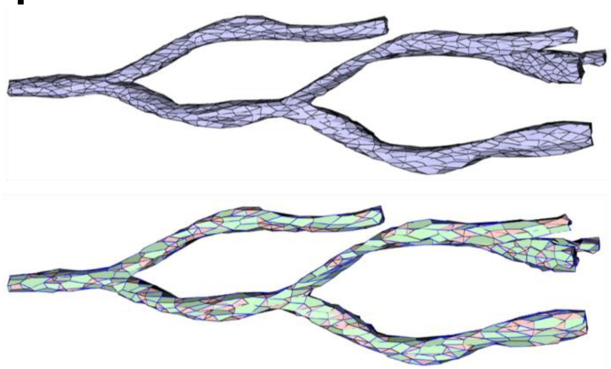
3D Spatial Query Engine

- On-demand Local Indexing
- 3D Spatial Query Processing
- Results Normalization

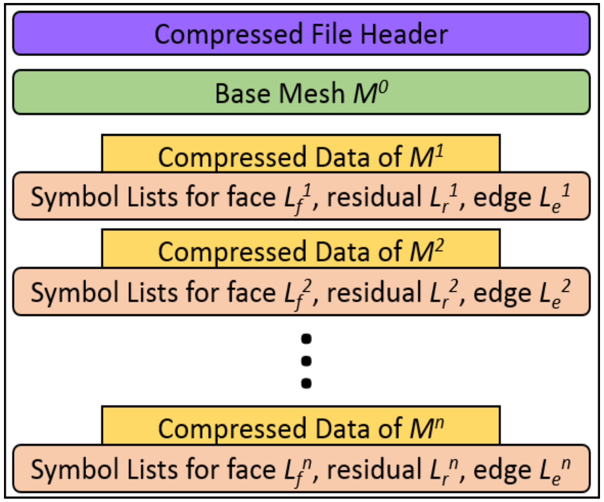
Distributed Computing Platform

3D Data Compression

- Each 3D object is compressed individually with successive levels of details
- A progressive approach consists of three steps
  - Decimation: remove vertices and add edges
  - Patch and edge encoding: generate symbol lists
  - Entropy Coding: further compress the symbol lists
- Compressed file size: base mesh: 1%; all LODs: 3%
- Compressed file stored in memory and replicated on all nodes



Mesh decimation

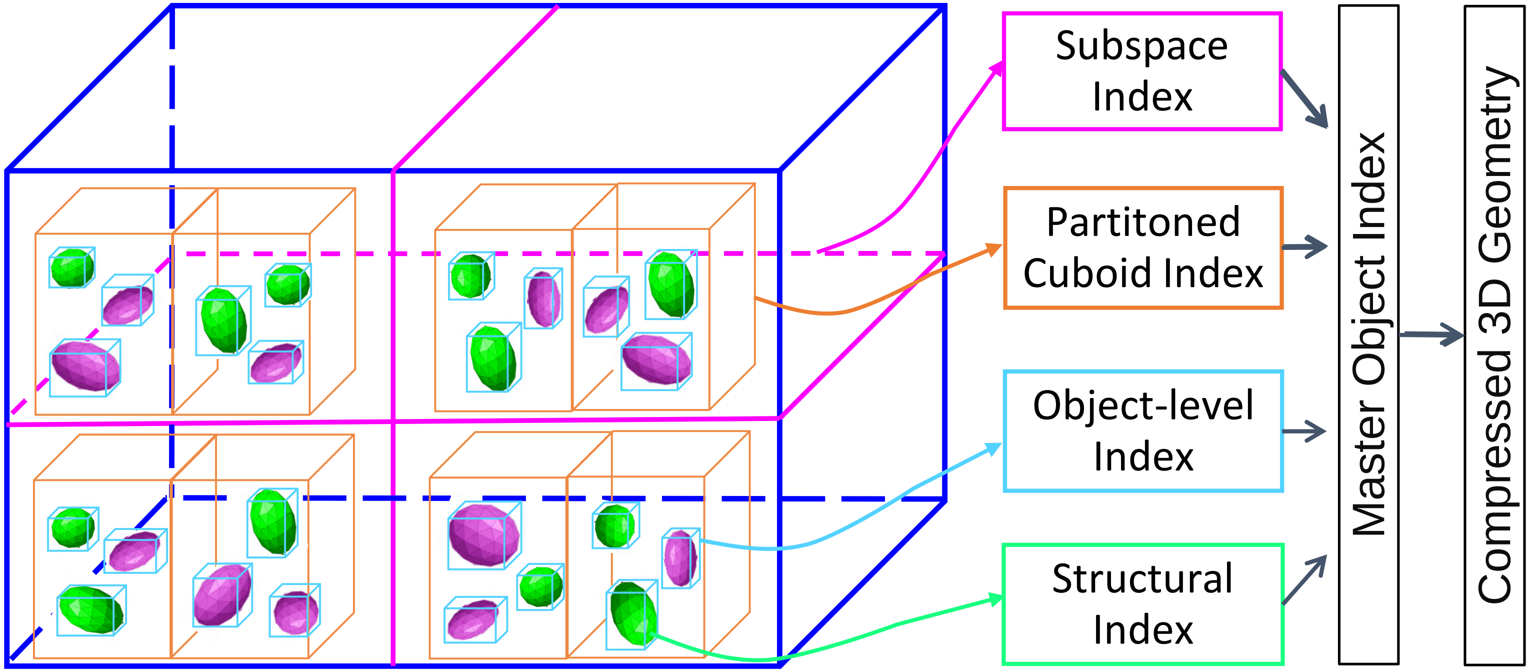


Compressed mesh file structure

Multi-level Spatial Indexing: Global Indexing

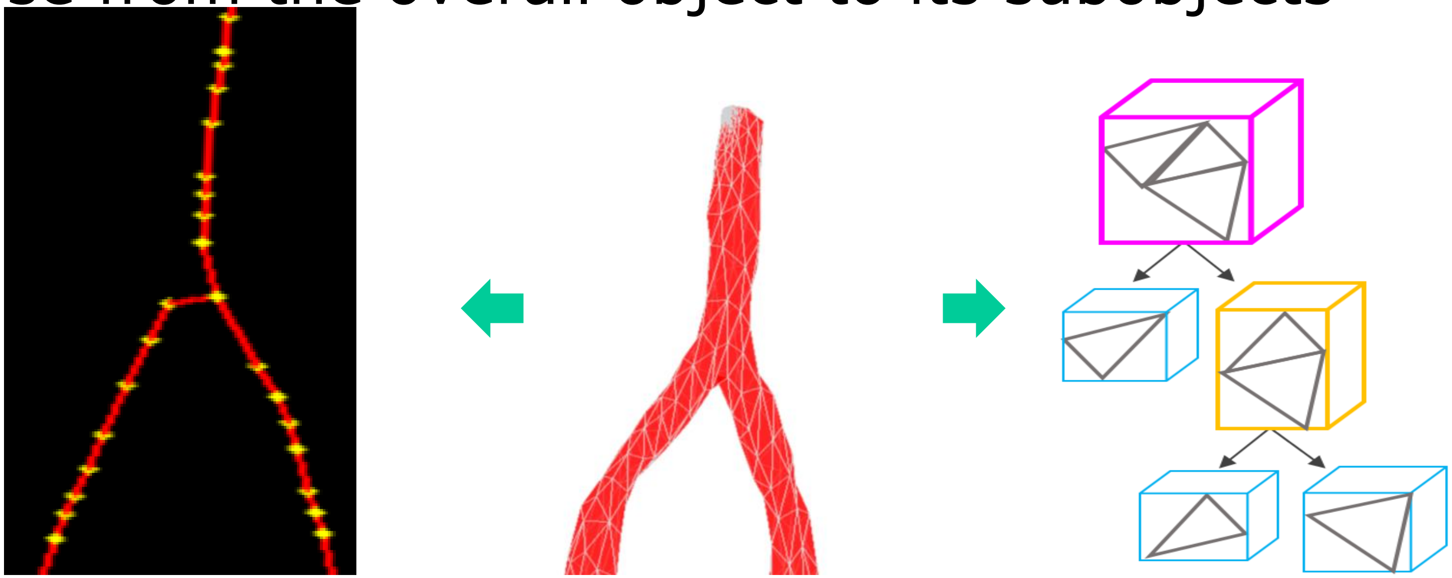
Stony Brook University

- Global indexing for space level data filtering and task grouping
  - Cuboid indexing for managing cuboids: all objects contained in a cuboid – grouping objects for tasks
  - Subspace based spatial indexing for grouping multiple neighboring cuboids: filtering data



Structural Indexing for Complex Objects

- Complex objects(e.g., vessels) can not be approximated as points or MBBs for distance based queries
- Topological skeleton based: effective shape abstraction to capture the essential topology of complex structures
- Hierarchical tree based: binary tree (e.g., AABB tree) based hierarchical representation of the MBBs which traverse from the overall object to its subobjects



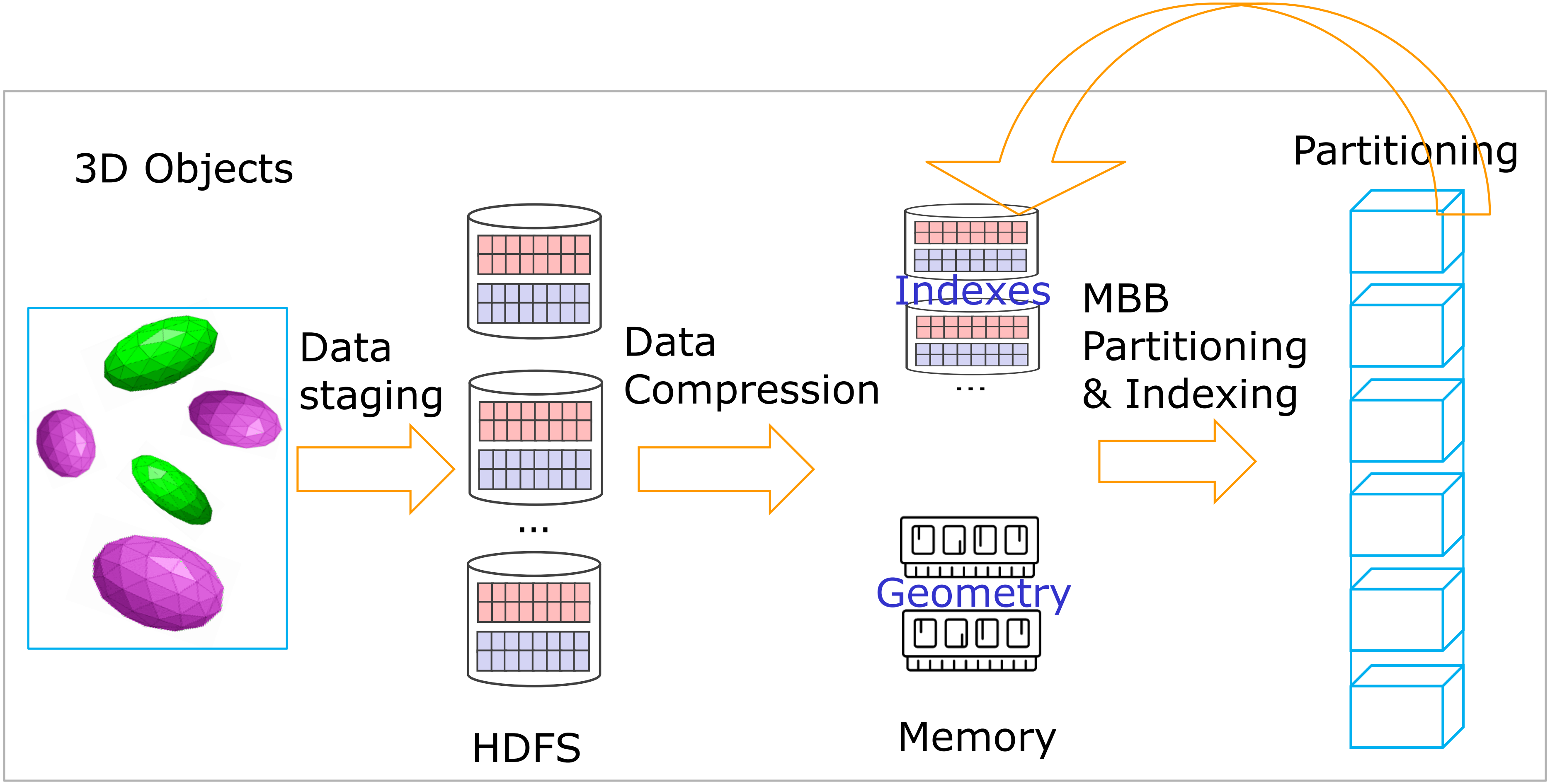
Skeleton based indexing　　original vessel　　hierarchical indexing

A General Framework of 3D Spatial Data Processing in MapReduce

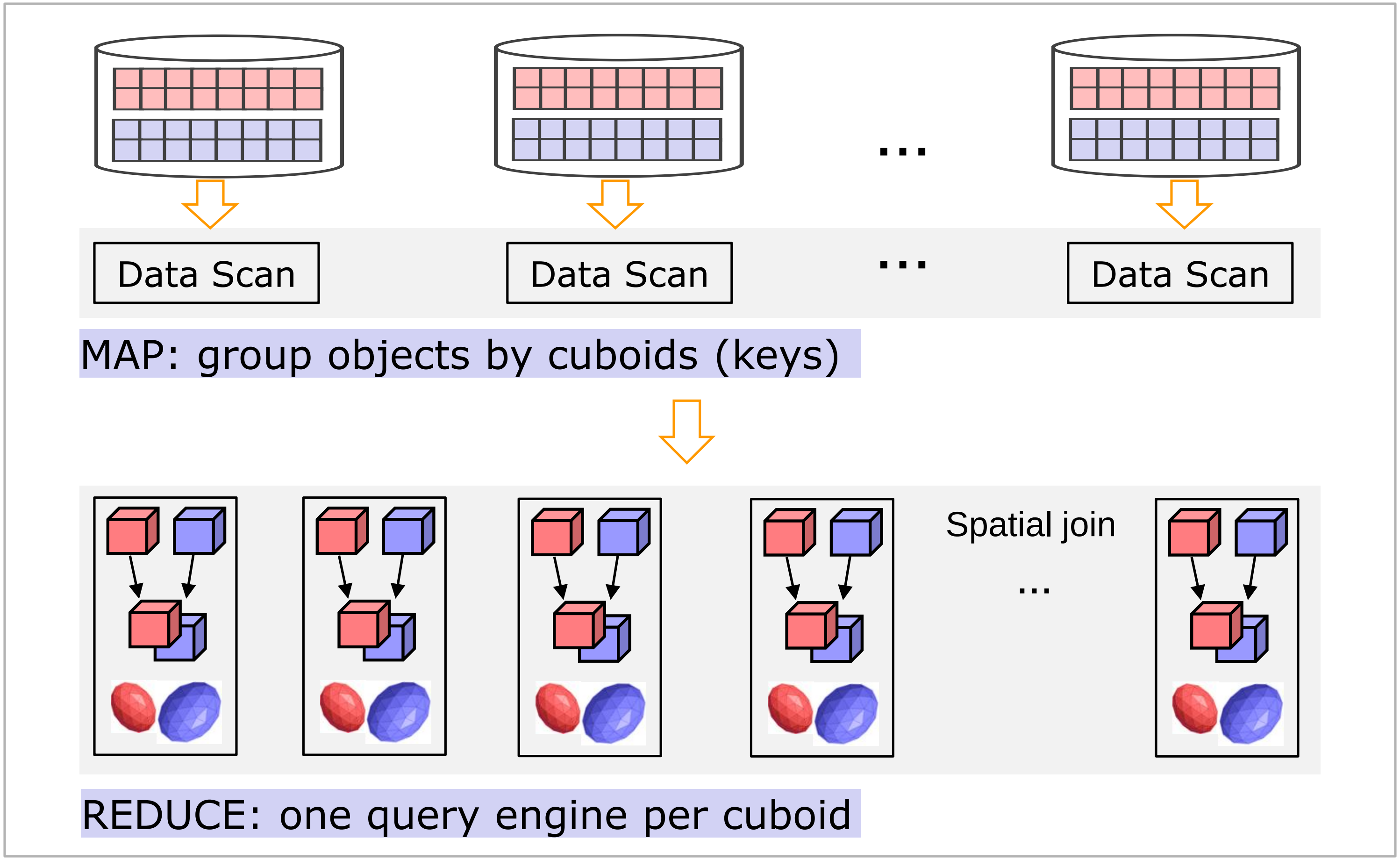
Stony Brook University

A. Raw data staging on HDFS
B. Data compression
C. Spatial partitioning and indexing
D. Block based filtering with global indexes
E. For each *cuboid* in *input_collection* do (in parallel)
    a. Cuboid based filtering based on cuboid indexes
    b. On-demand indexing for objects in the cuboid
    c. Cuboid based spatial query processing
F. Boundary-crossing object handling
G. Result storage on HDFS

Example: 3D Spatial Join in MapReduce: Data Compression and Partitioning



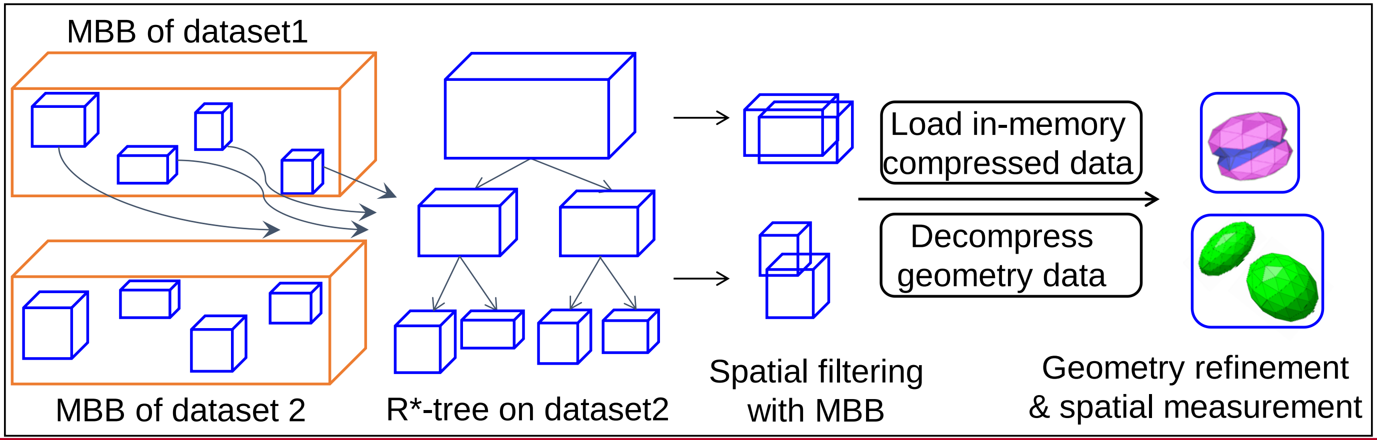
Example: 3D Spatial Join in MapReduce: Query Processing



On-demand 3D Query Engine (INTENSE)

Stony Brook University

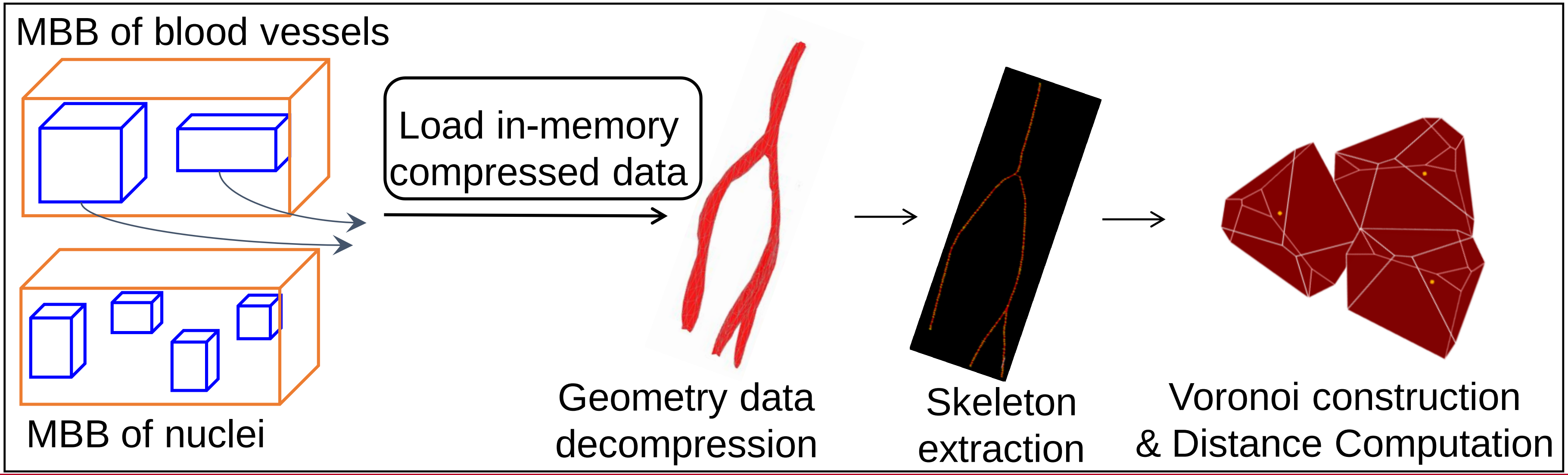
- Effective querying methods that can run in parallel in distributed computing environments
  - Spatial join, containment, nearest neighbor, proximity estimation, and can be extended
  - Computational Geometry Algorithms Library (CGAL)
- On-demand indexing based query processing
  - Create indexes on demand in memory: indexing for multiple objects or structural indexing for individual objects
  - Indexing building overhead is a very small fraction



Spatial join

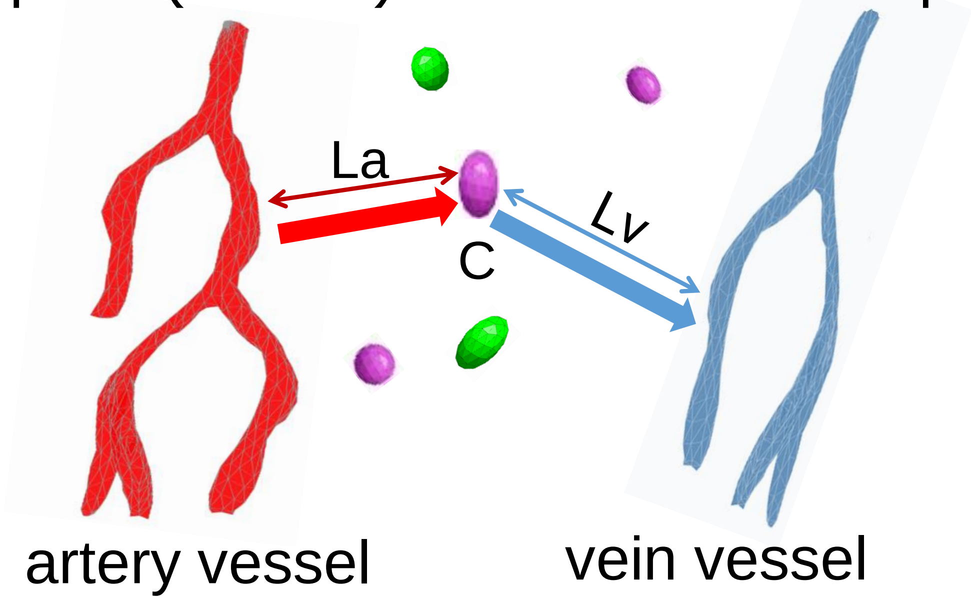
Nearest Neighbor Search in INTENSE

- Nearest neighbor search using skeleton indexing
  - For example: for each cell find the closest blood vessel and return distance to that blood vessel
  - For each vessel in a partition, build skeleton index and use vertices of skeleton as input sites for Voronoi diagram construction
  - For each cell, perform NN search with Voronoi diagram using skeleton indexing for distance computation



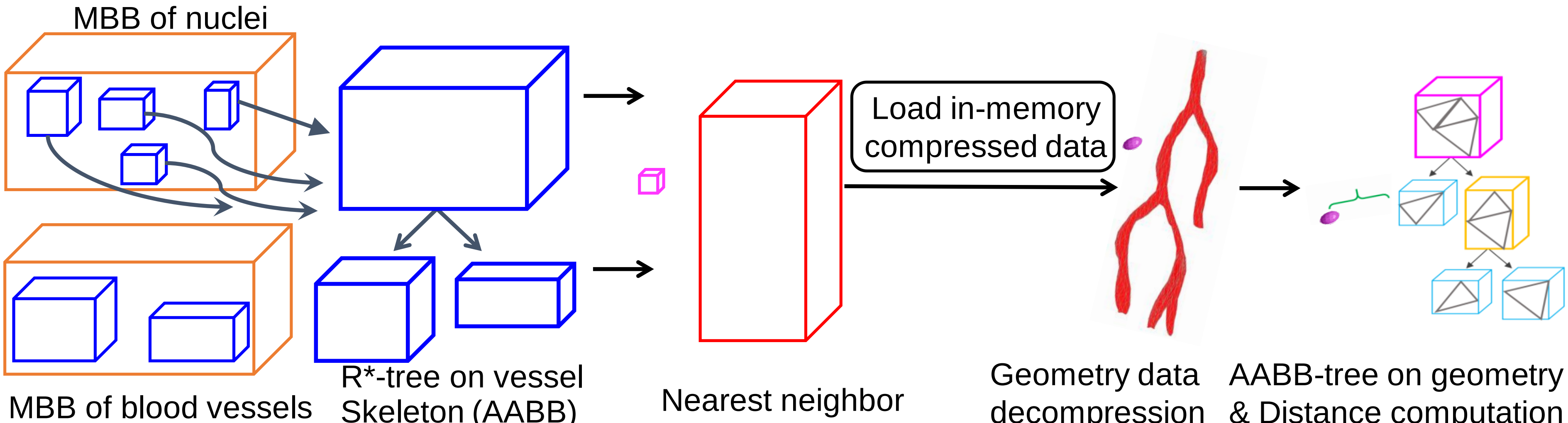
Proximity Estimation in INTENSE

- Explore distributions of different types of target objects (e.g., arteries and veins) with respect to a set of basic objects (e.g., cells)
  - For each cell and for each target vessel type (artery or vein), identify the nearest target object, compute the distances (La and Lv for cell C), and return the sum of the shortest path (La+Lv) for mean and dispersion calculation

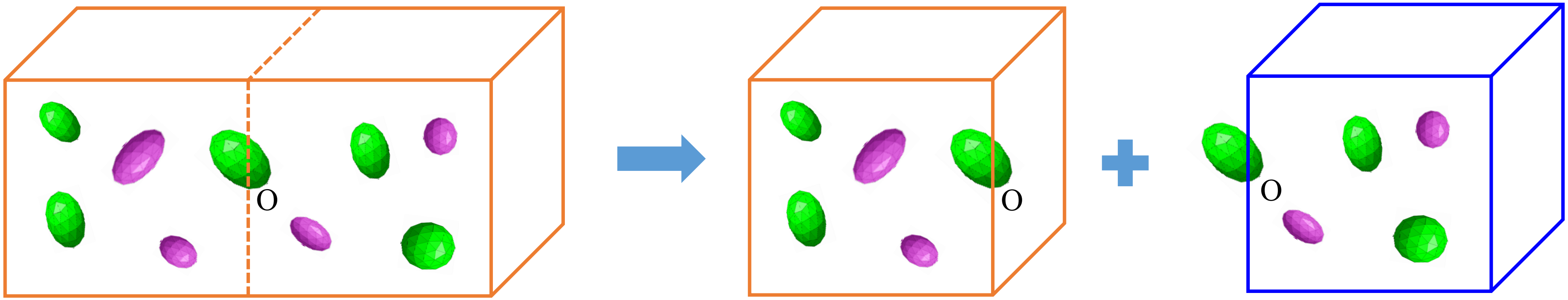


Proximity Estimation in INTENSE (cont'd)

- Three-step proximity estimation query pipeline using structural indexing and R*-tree indexing
  - Build structural index (AABB) of vessels in low LOD
  - Find nearest neighbor using R*-Tree to index MBBs from above structural index
  - Perform accurate calculation of distances using structural index of vessels at high LOD (very few)
- Only small number of vessels with high LOD will be retrieved

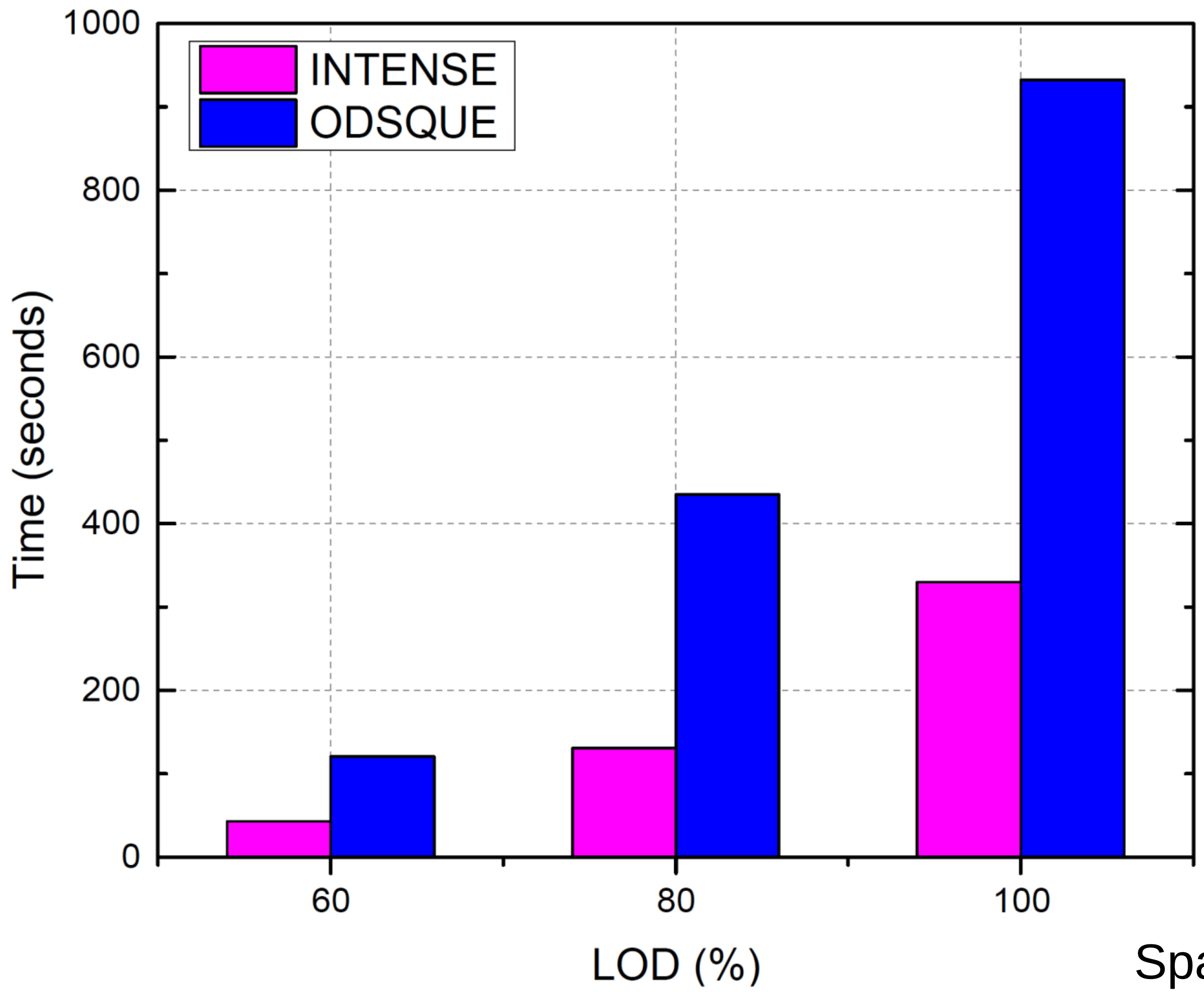


Boundary-Crossing Object Handling



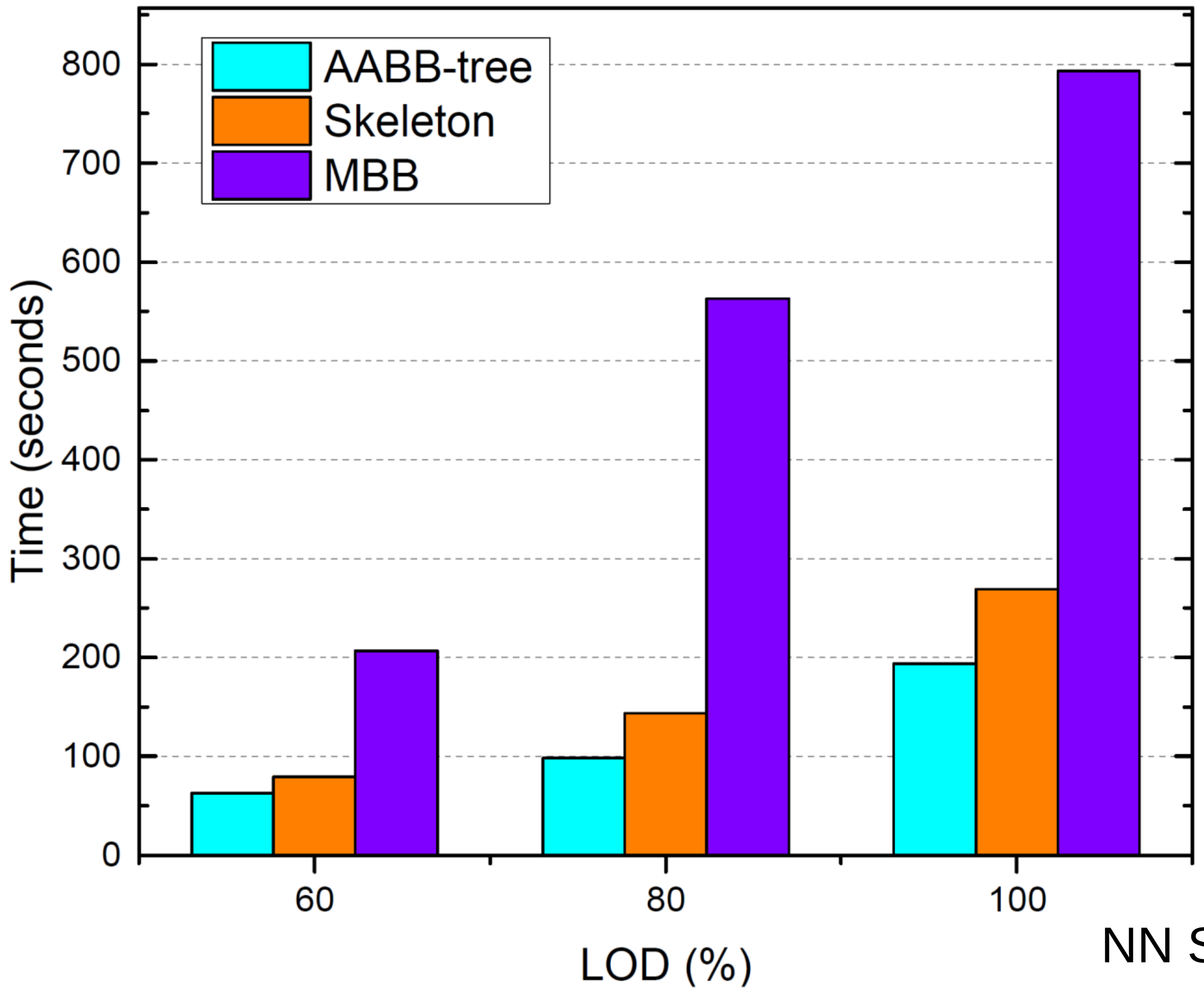
- "Multi-assignment, single-join": replicate objects on boundaries to multiple tiles at partitioning
- Normalization methods are provided for each query type to correct answers
  - Spatial join: removal of duplicates
- Nearest neighbor
  - It always return correct nearest neighbors

Performance: INTENSE vs ODSQUE (disk based storage)



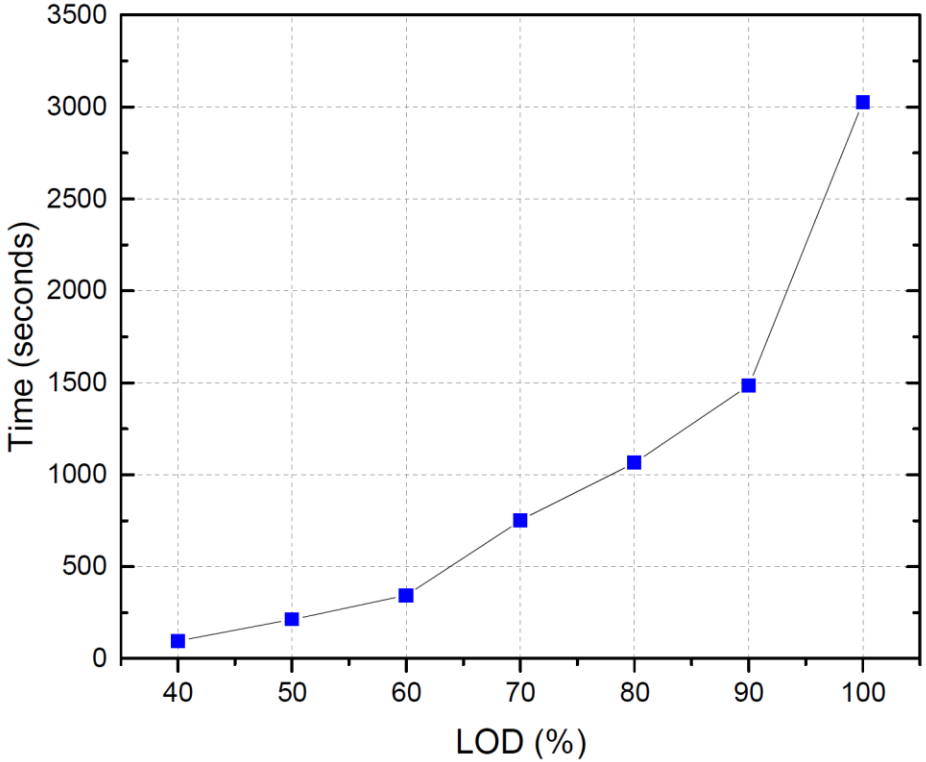
Spatial join

Performance: Structural Indexing vs MBB

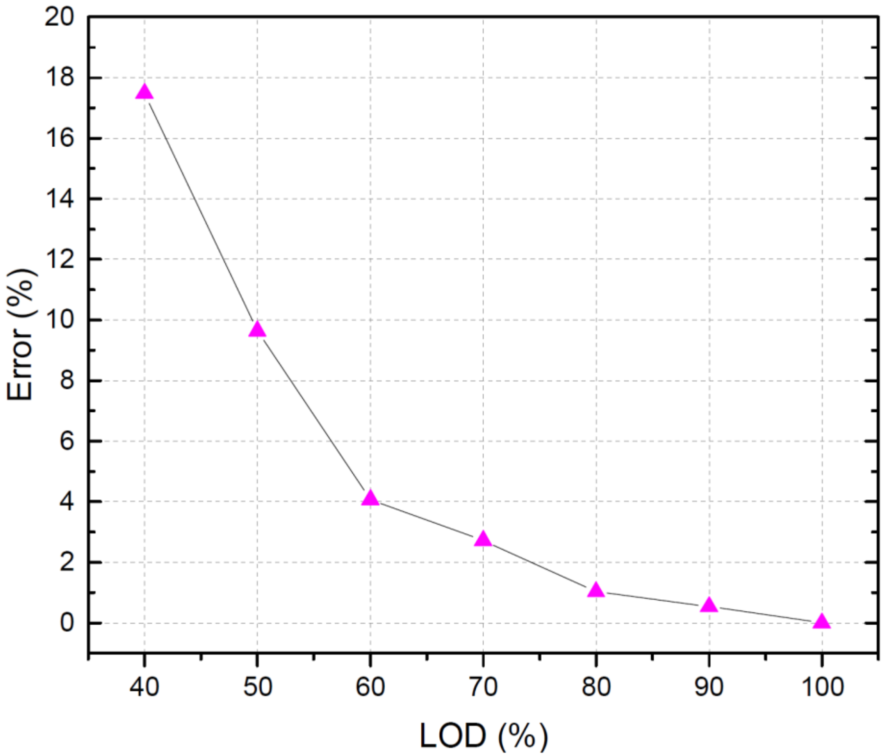


NN Search

Query Performance vs LODs

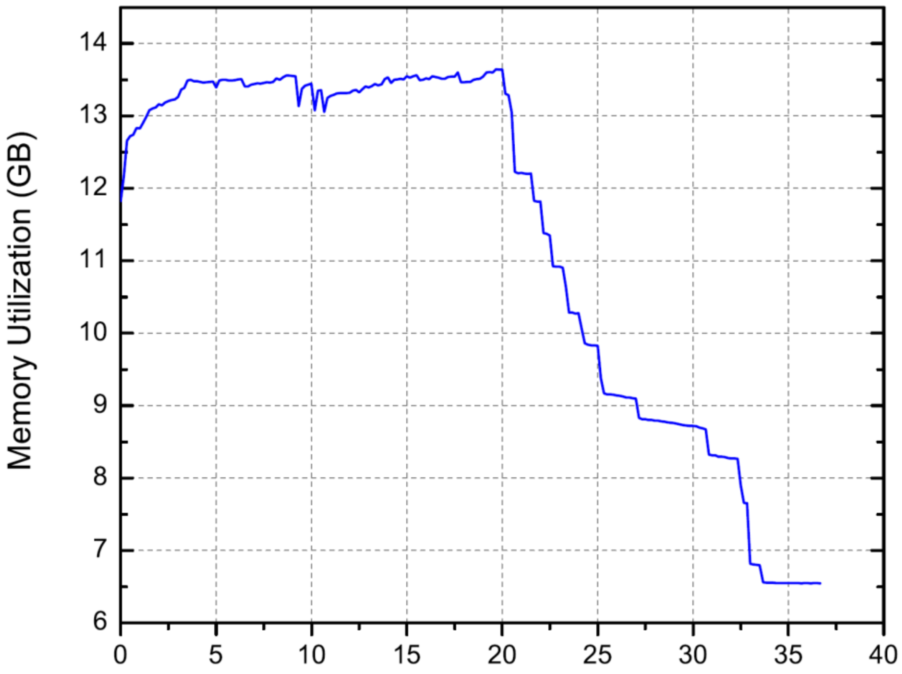


Execution time vs LOD

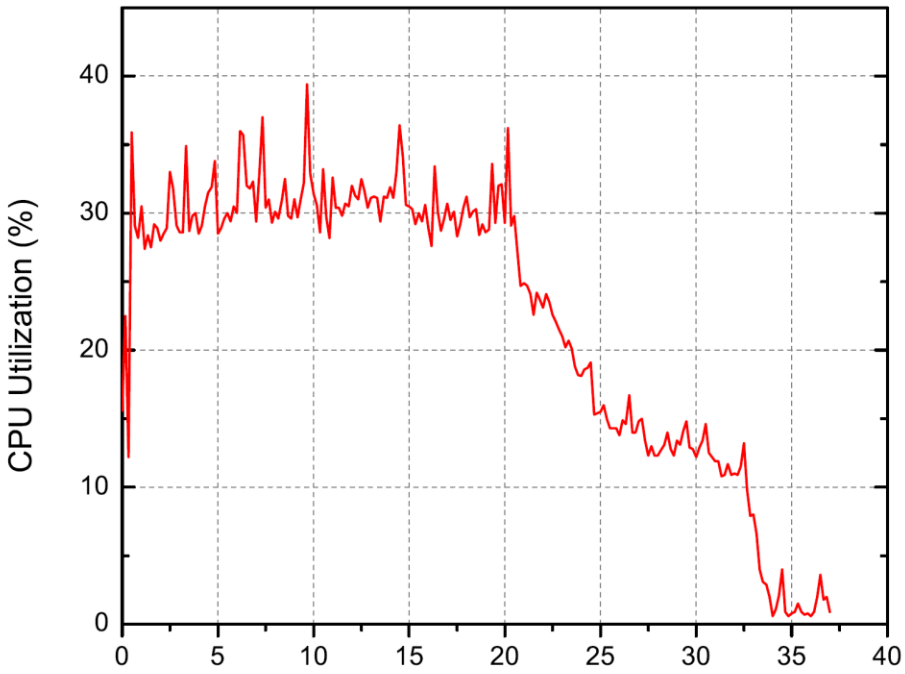


Error vs LOD

CPU and Memory Usage

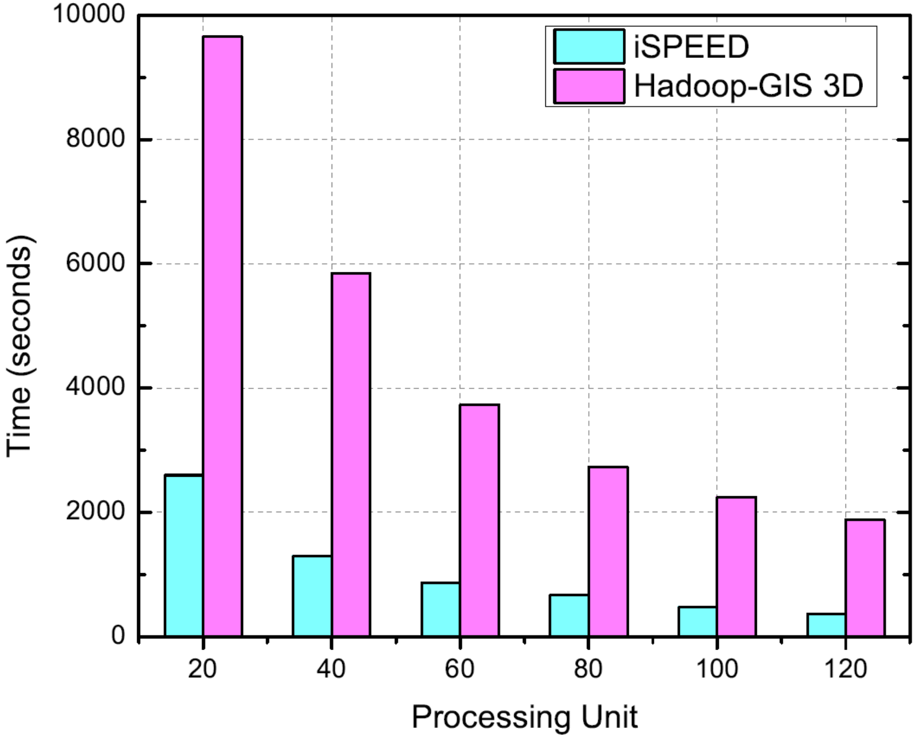


Memory

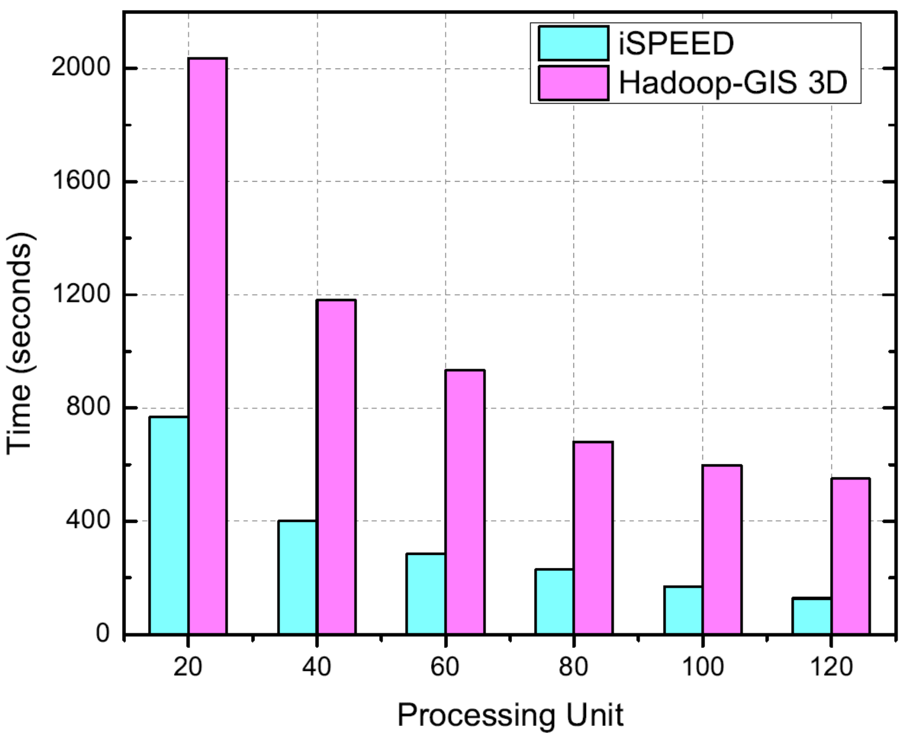


CPU

Performance Comparison between iSPEED and Hadoop-GIS 3D

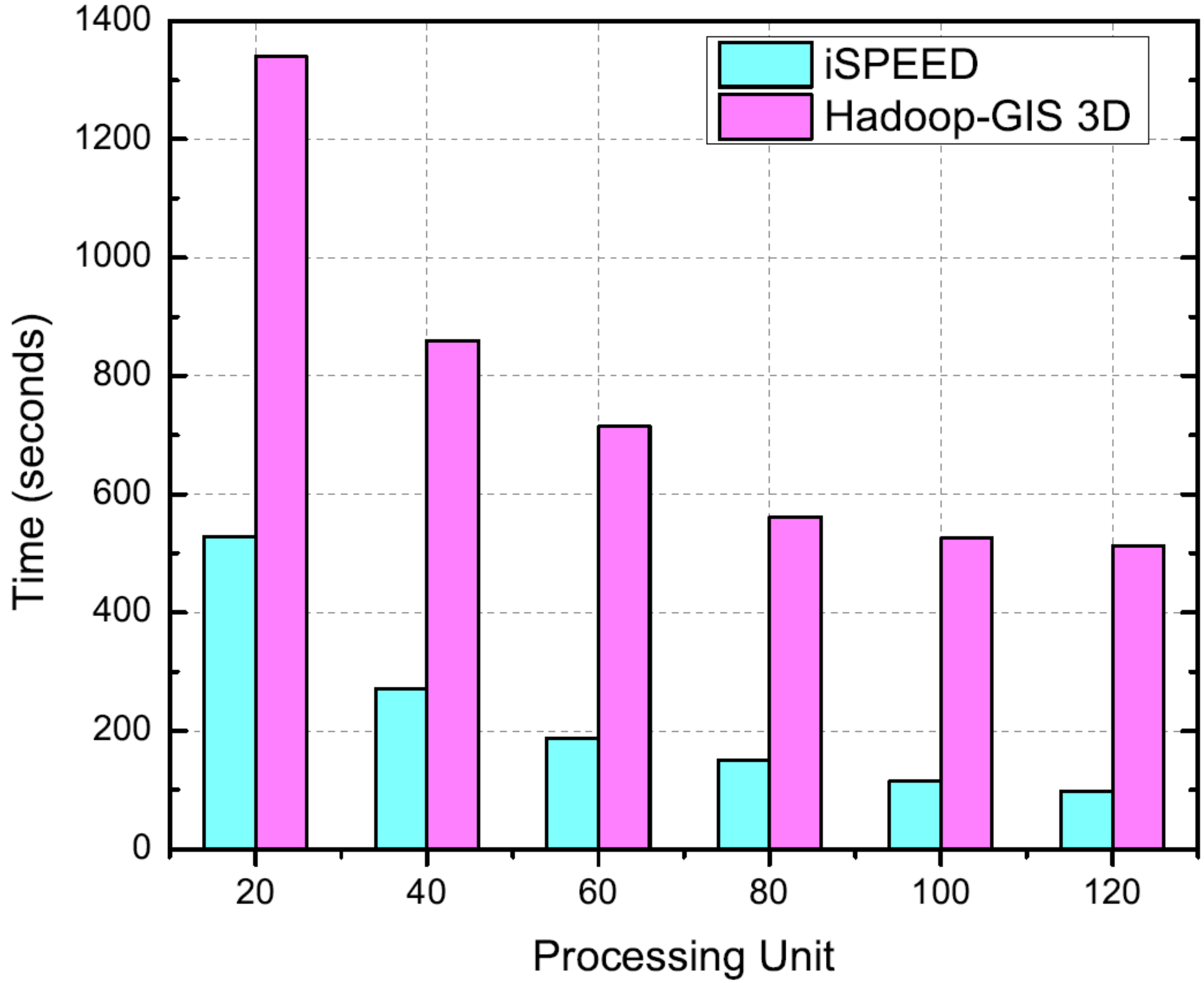


Spatial join



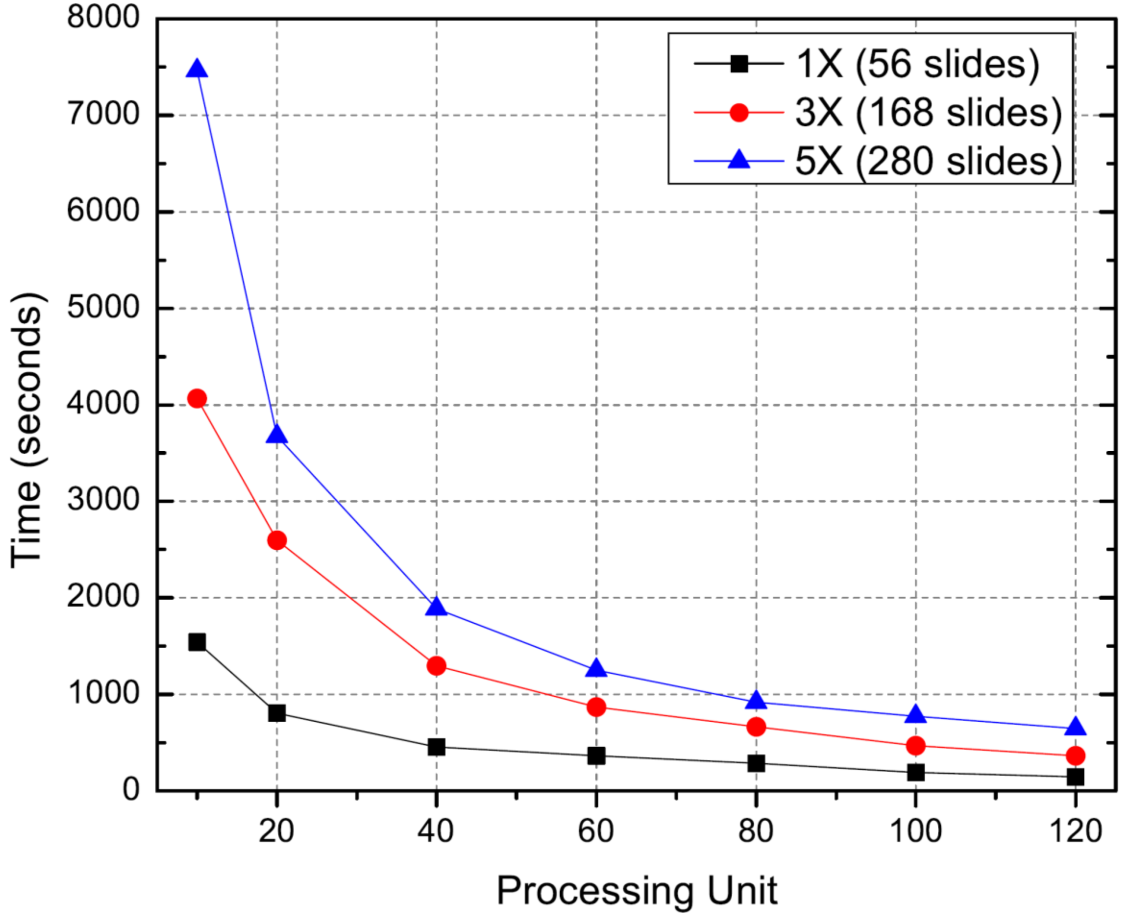
Nearest Neighbor

Performance Comparison between iSPEED and Hadoop-GIS 3D



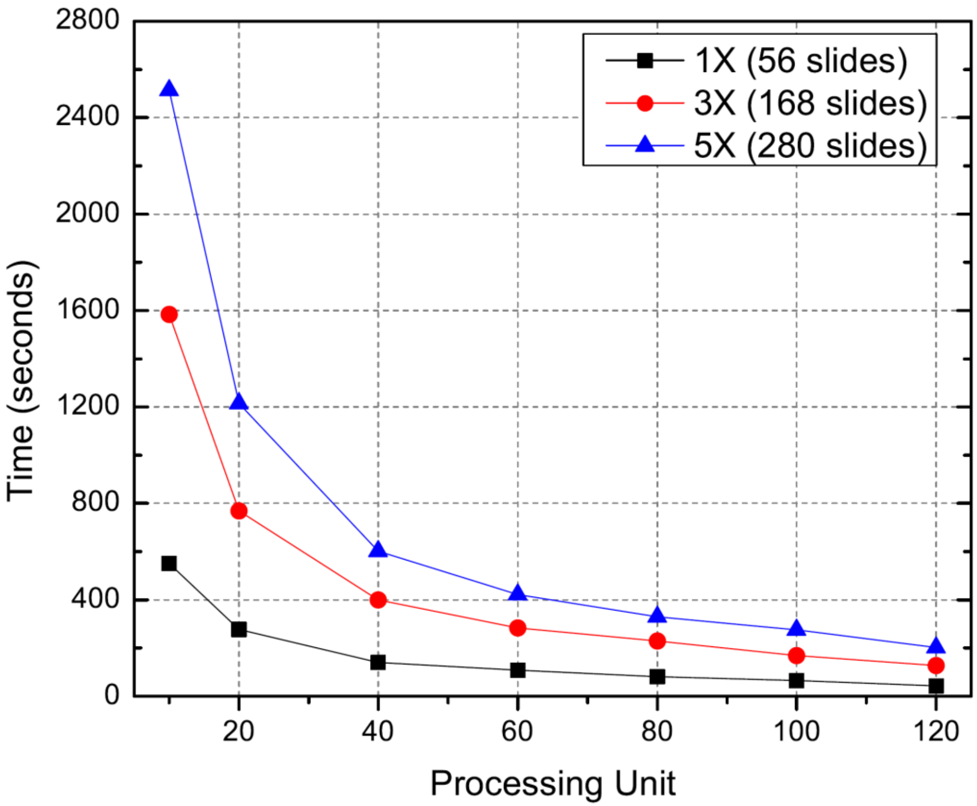
Spatial approximation

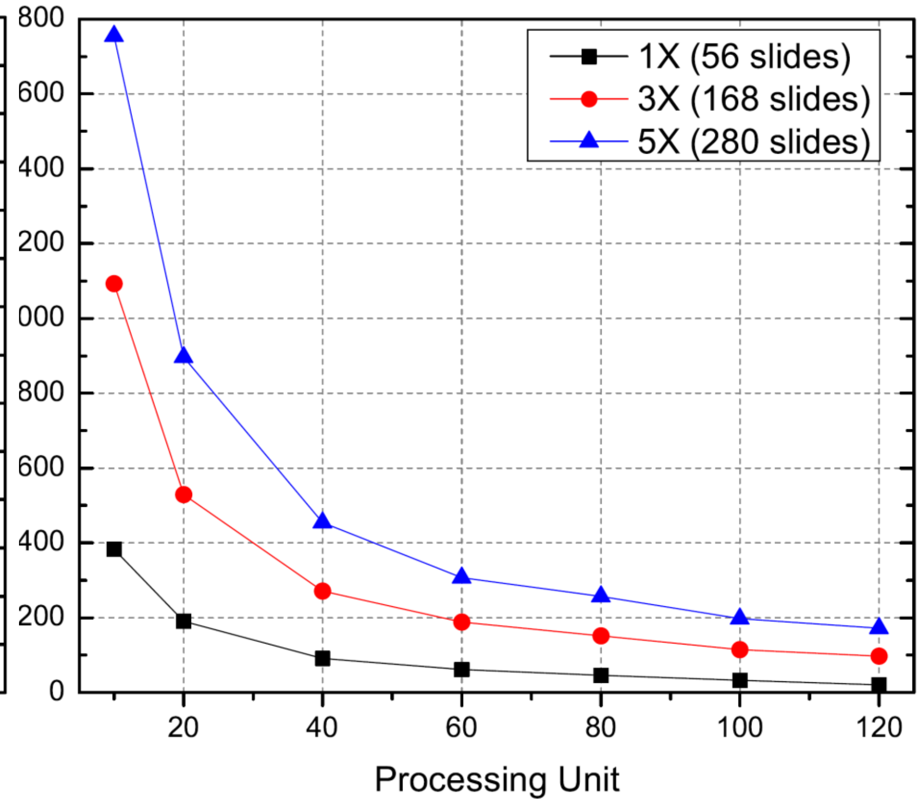
Scalability of iSPEED: Spatial Join



Scalability of iSPEED

Stony Brook University



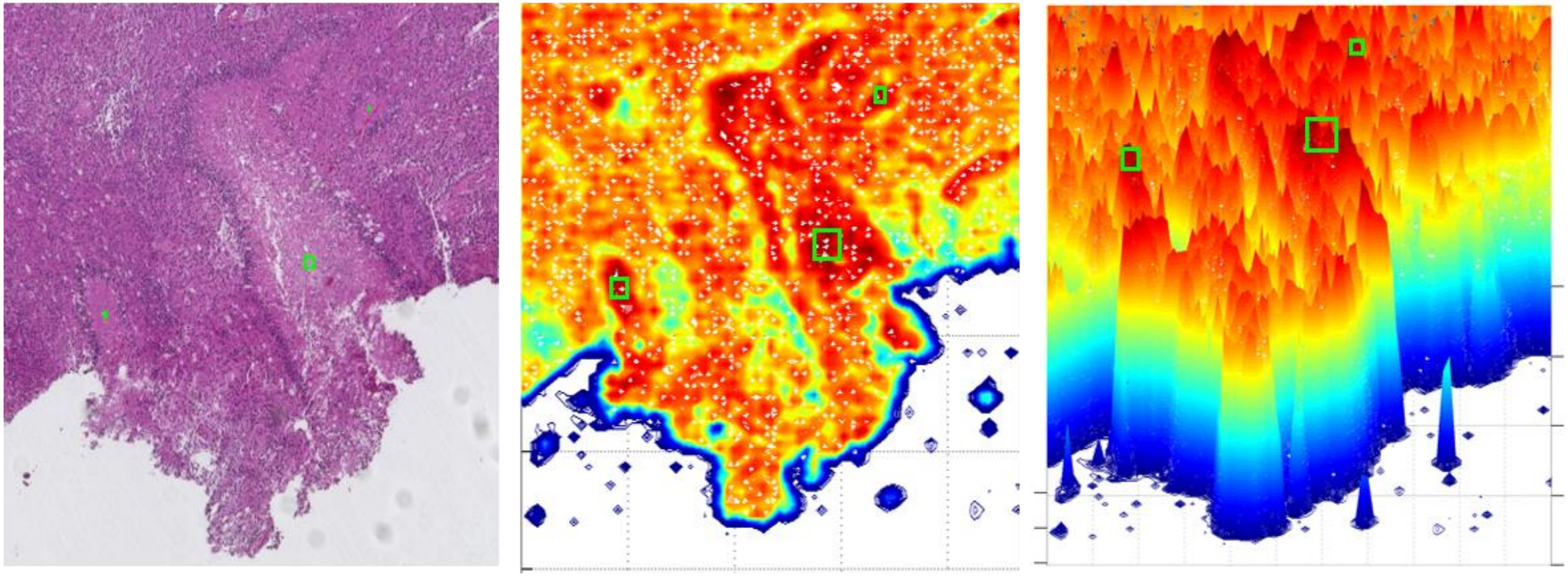


Nearest Neighbor

Spatial Approximation

Integrate Histogogical H&E Images with Biomarker IHC Images

- H&E stained image for pathology structures and features
- IHC stained adjacent image for biomarkers
- Spatially map biomarker data to histology structures



The correlation between necrotic centers and spatial density of hypoxia biomarker

Conclusion

Stony Brook University

- 3D data compression makes it possible to significantly reduce data size for in-memory storage and indexing, leading to much reduced I/O and communication cost
- Model 3D objects with multiple LOD for spatial queries, options for users to balance performance and accuracy
- Multi-level in-memory spatial indexing to reduce search space and accelerate queries, unique structural indexing for boosting distance based queries
- On-demand based 3D spatial query engine using multi-level indexing and data decompression to support high scalability

Future Work

- Study complex spatial analytics due to the heterogeneity of 3D pathology object structures and relationships and complex spatial patterns
- Explore parallelization of complex spatial analytics methods such as 3D spatial clustering and comparison of spatial point patterns
- GPU accelerated spatial querying methods
- Use RDMA to reduce communication cost