Graph-Shifts

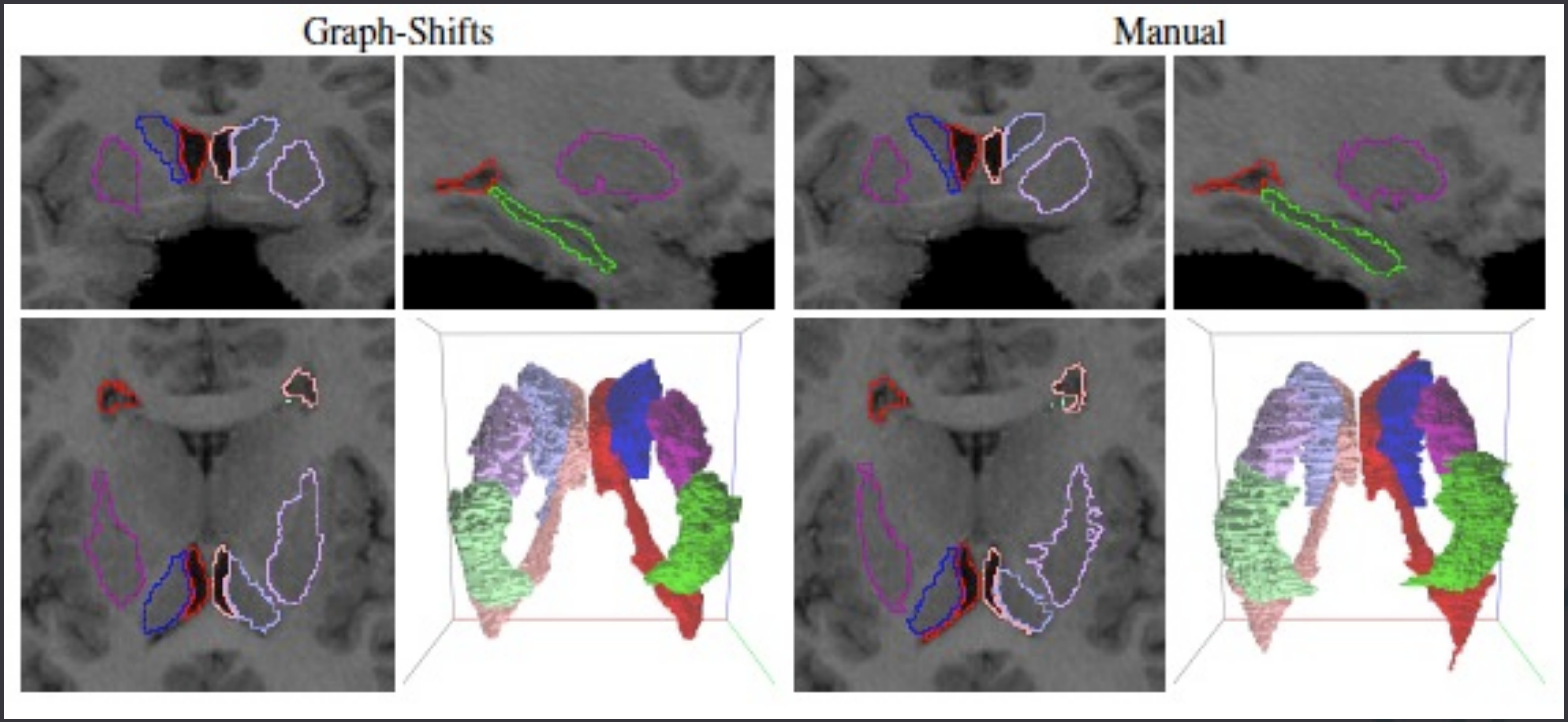
Anatomic 3D Segmentation by Dynamic Hierarchical Minimization

Jason Corso
Postdoctoral Fellow
UCLA LONI/CCB

jcorso@ucla.edu

Motivation

- The work deals with the problem of automatically labeling 3D data into anatomic (and pathologic) structures of interest.
- The resulting segmentation can be used for a variety of analyses.
- Work on sub-cortical brain structures.



Problem Statement

- Class-based segmentation / partitioning.
  - Given a set of models of interest, associate one model label with each voxel.
  - The models correspond to different anatomical regions of interest m_k.
  - A solution is represented by $\{m_\omega : \omega \in D\}$, D is voxel lattice.
  - The class of energies we consider in this formulation is

$$\sum_{\nu \in D} E_1(\phi(\mathbf{I})(\nu), m_\nu) + \frac{1}{2} \sum_{\substack{\nu \in D, \mu \in D: \\ N(\nu,\mu)=1}} E_2(\mathbf{I}(\nu), \mathbf{I}(\mu), m_\nu, m_\mu)$$

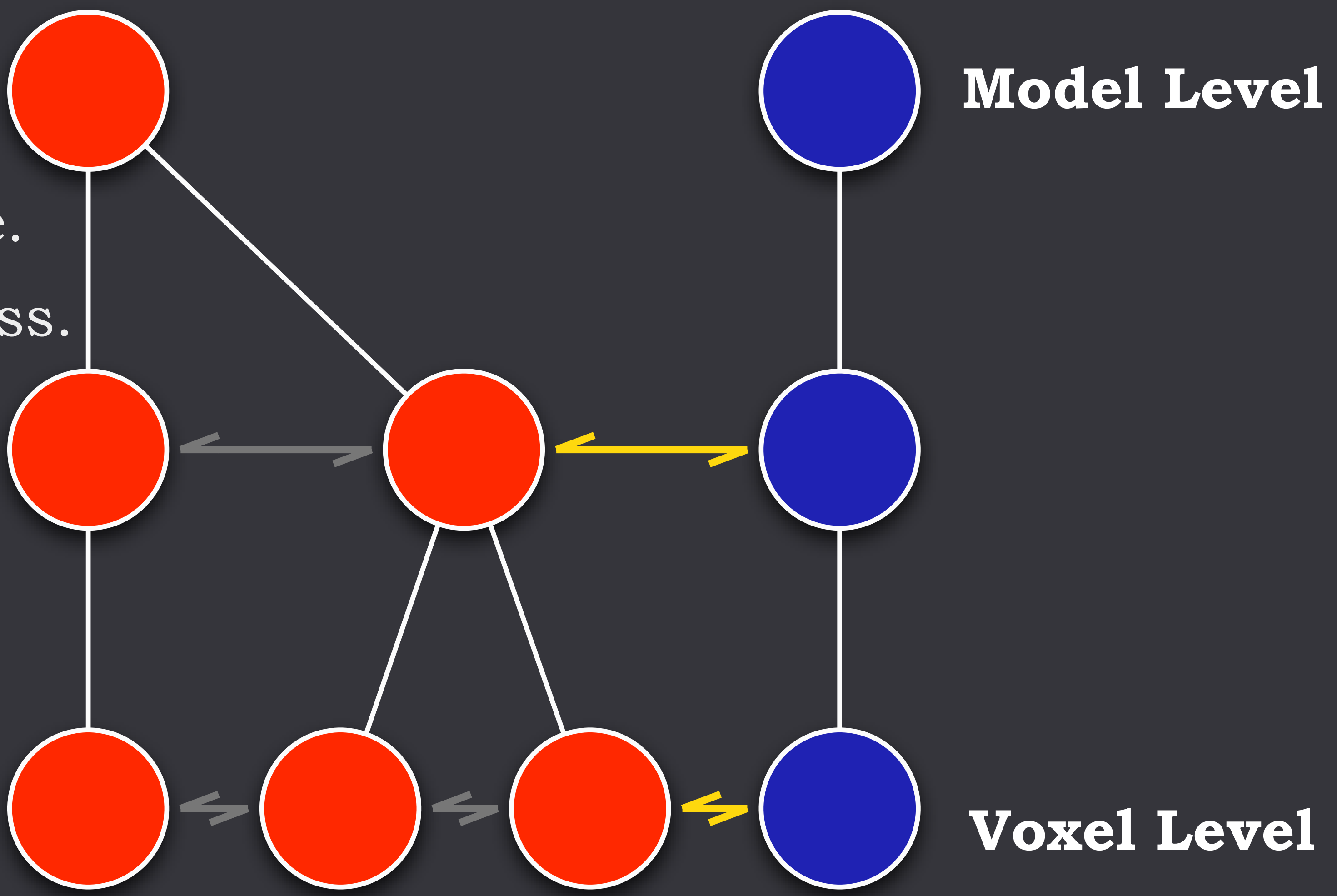
- E_1 is a unary term on voxel likelihood for a given model
  - The $\phi(\mathbf{I})(\nu)$ is a non-linear filter incorporating context and is learned from training data.
- E_2 is a binary term on pair-wise voxels.
  - This can include conventional PDE-type functions such as $\int_{\delta R} ds$ and $\int_{\delta R} |\nabla \mathbf{I}|^2 ds$
  - Or it can include pairwise terms learned from data like conditional random fields.

Prior Art

- Deterministic Methods
  - Level-Set / PDE methods
    - Operate at a single level only causing slow convergence and local minima risk.
  - Graph-Cut Methods
    - Take global cuts, but only guaranteed to converge for a small class of energies.
- Stochastic Methods
  - Markov Chain Monte Carlo and DDMCMC
    - Take samples from a global probability distribution.
    - Very slow convergence.
    - How to design proposal distributions to activate the split, merge and other moves?
- Hierarchical Methods
  - Segmentation by Weighted Aggregation
    - Does not minimize any objective function.
    - Instead, outputs regions satisfying certain homogeneity properties.
    - Soft representation requires huge amounts of memory (especially in 3D).
  - Hierarchical Swendsen-Wang
    - Again, stochastic.
    - Limited dynamics in the hierarchy.

Graph-Shifts Algorithm

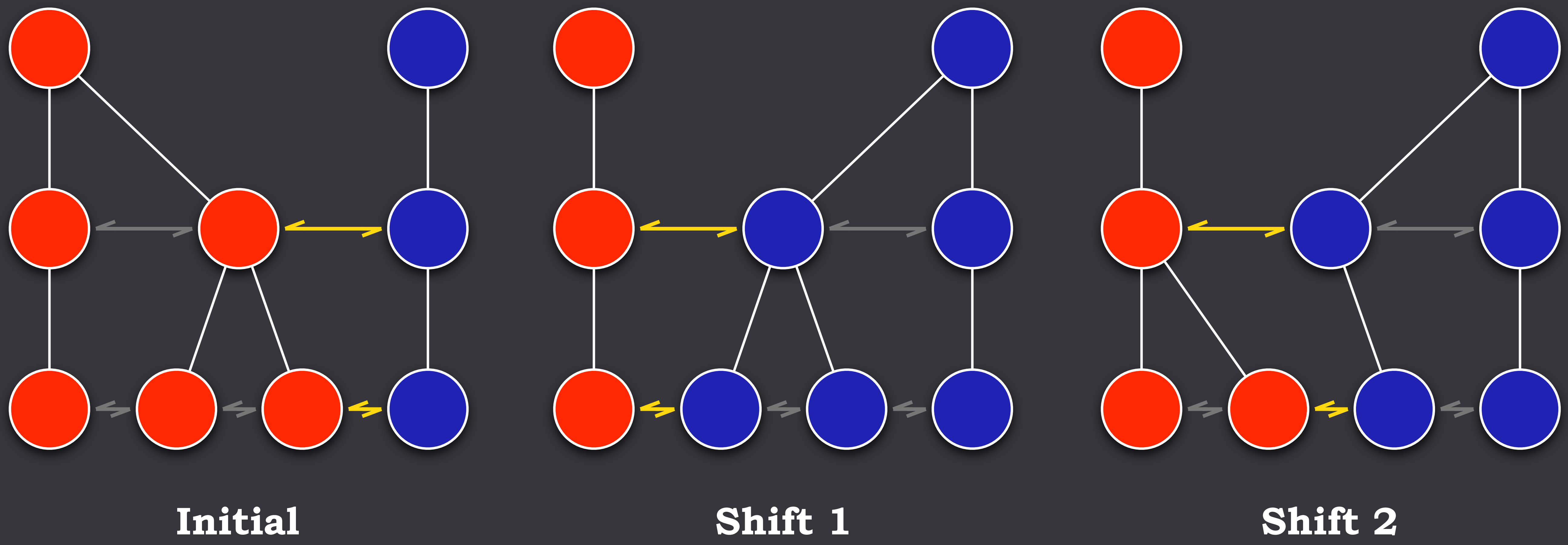
- Manipulates a dynamic hierarchical representation of the image.
- Can take large (split-and-merge) and small (PDE-style) moves.
- A discrete, steepest descent minimizer.
- Novel representation and graph dynamics make it possible to quickly explore the combinatoric space and take the optimal move (in a local sense) at every iteration.
- Very rapid convergence (orders of magnitude faster than others).
- Graph Example
  - Eg. red class and blue class.
  - One model node per structure.
  - Each node inherits parent class.
- Recursive energy definition:
  - Compute energy at any node.



Graph-Shifts Algorithm

...or, the Godfather Algorithm

- A graph shift is when one node takes the parent of a neighbor.
- The complete subgraph of the shifted node takes the new label.
- Potential shifts are shown in yellow.



Computing and Selecting Shifts

- Each shift stores the exact resulting change in the energy function.
  - Called the *shift-gradient*:

$$\Delta E(\mu \to \nu) = E_1(\mu, m_\nu) - E_1(\mu, m_\mu) + \sum_{\eta : N(\mu,\eta)=1} \left[E_2(\mu, \eta, m_\nu, m_\eta) - E_2(\mu, \eta, m_\mu, m_\eta) \right]$$

- Actual number of potential shifts is very small.
  - Empirically shown to be about 1% of all edges in graph.
- The complete set of potential shifts is stored at all times.
- Upon taking a shift, the potential shifts along the shift boundary are updated.
  - Number of affected shifts is logarithmic in input size.

Rough Sketch of Algorithm

GRAPH-SHIFTS

Input: Volume **I** on lattice D.

Output: Label volume **L** on lattice D.

0 Initialize graph hierarchy.

1 Compute exhaustive set of potential shifts S.

2 while S is not empty

3 s gets the shift in S that best reduces the energy.

4 Apply shift s to the graph.

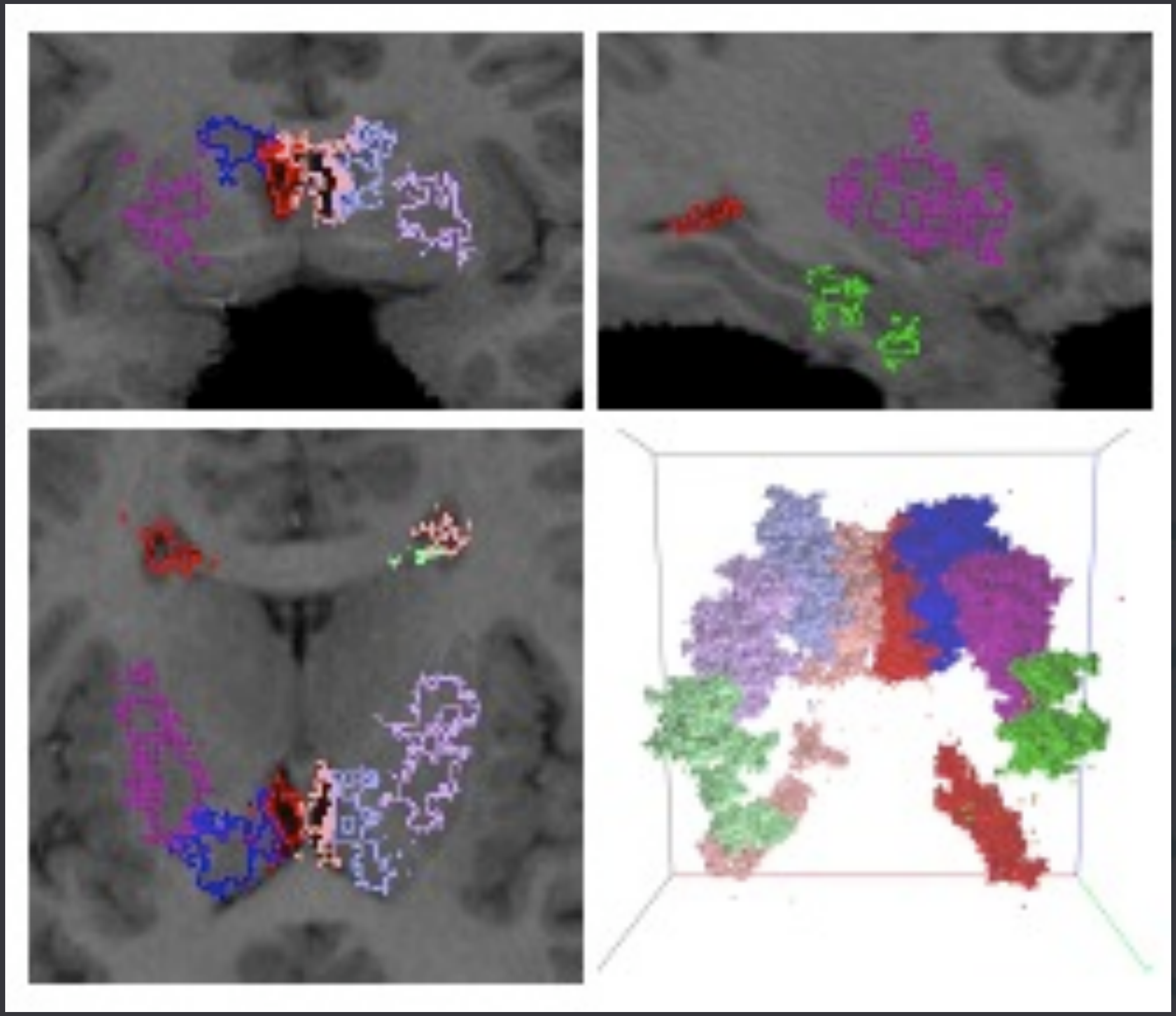
5 Update affected region and edge properties.

6 Recompute affected shifts on boundary and update S.

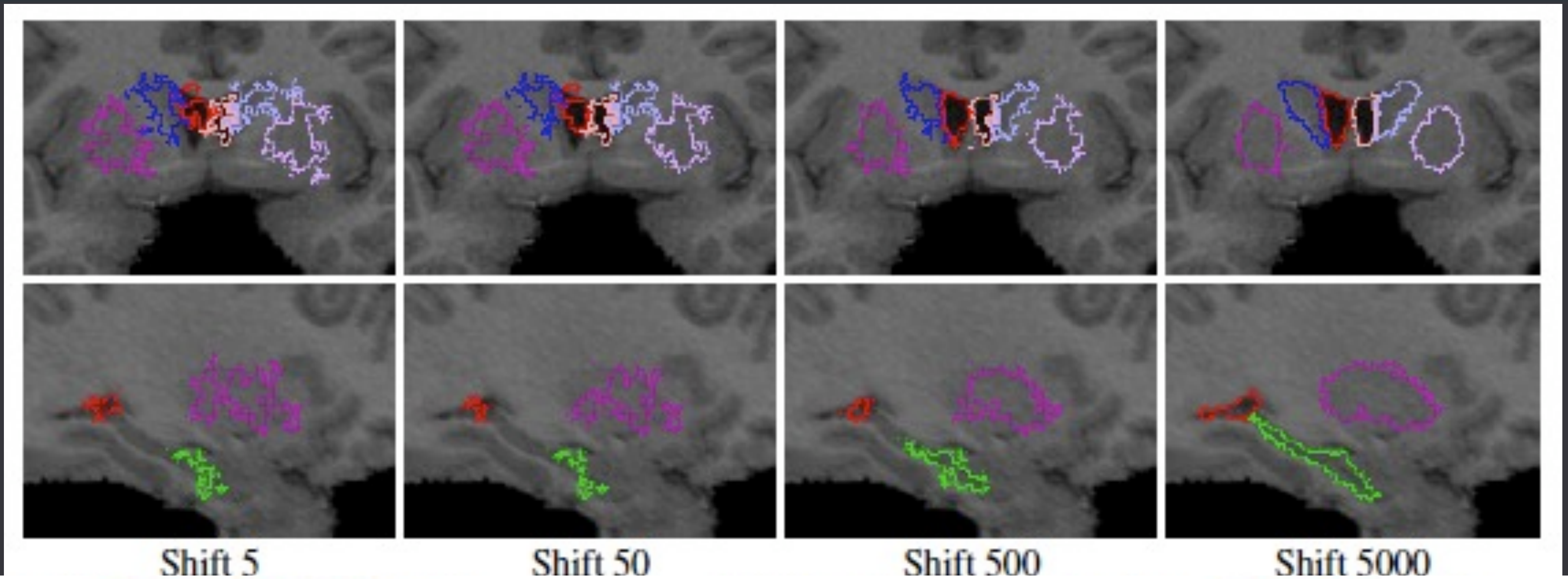
7 Compute label volume L from final hierarchy.

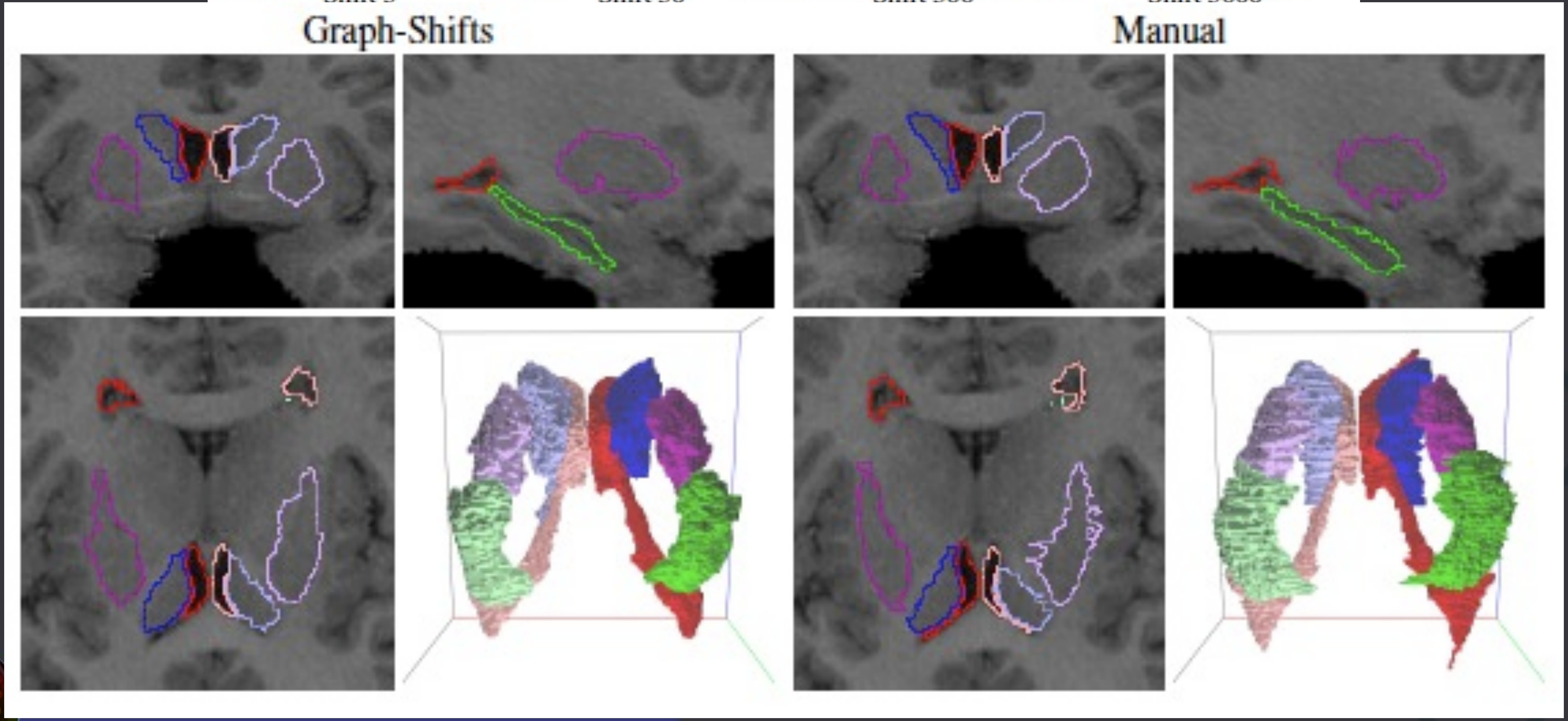
Initialization

- Rapid, bottom-up hierarchy initialization.
- Take insight from the SWA algorithm and Statistical Affinities.



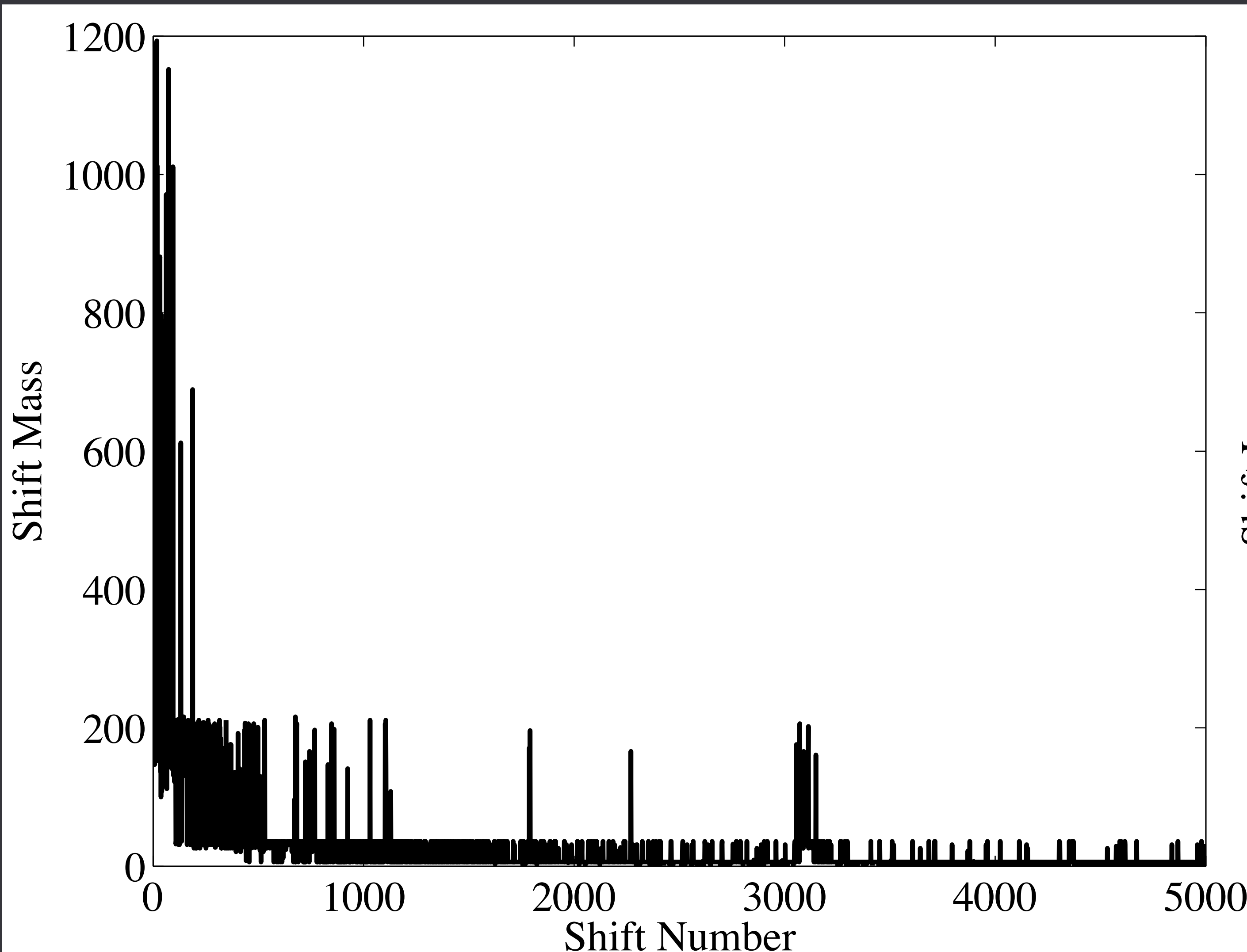
Graph-Shifts Process and Final



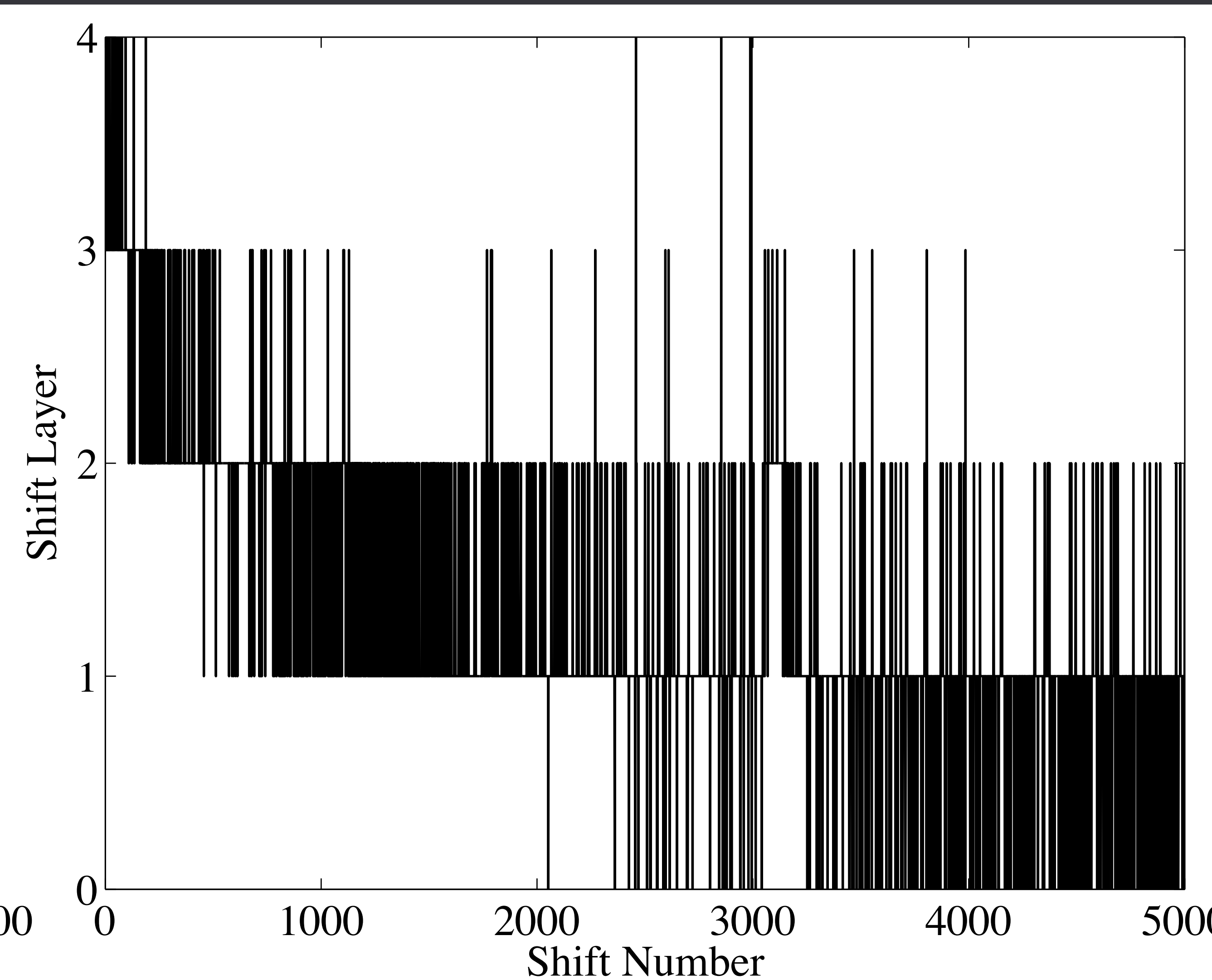


Graph-Shifts Measurements

- Early shifts occur at high-levels in the hierarchy corresponding to large changes in the energy.
- Shift mass is the number of voxels that had their labels changed.



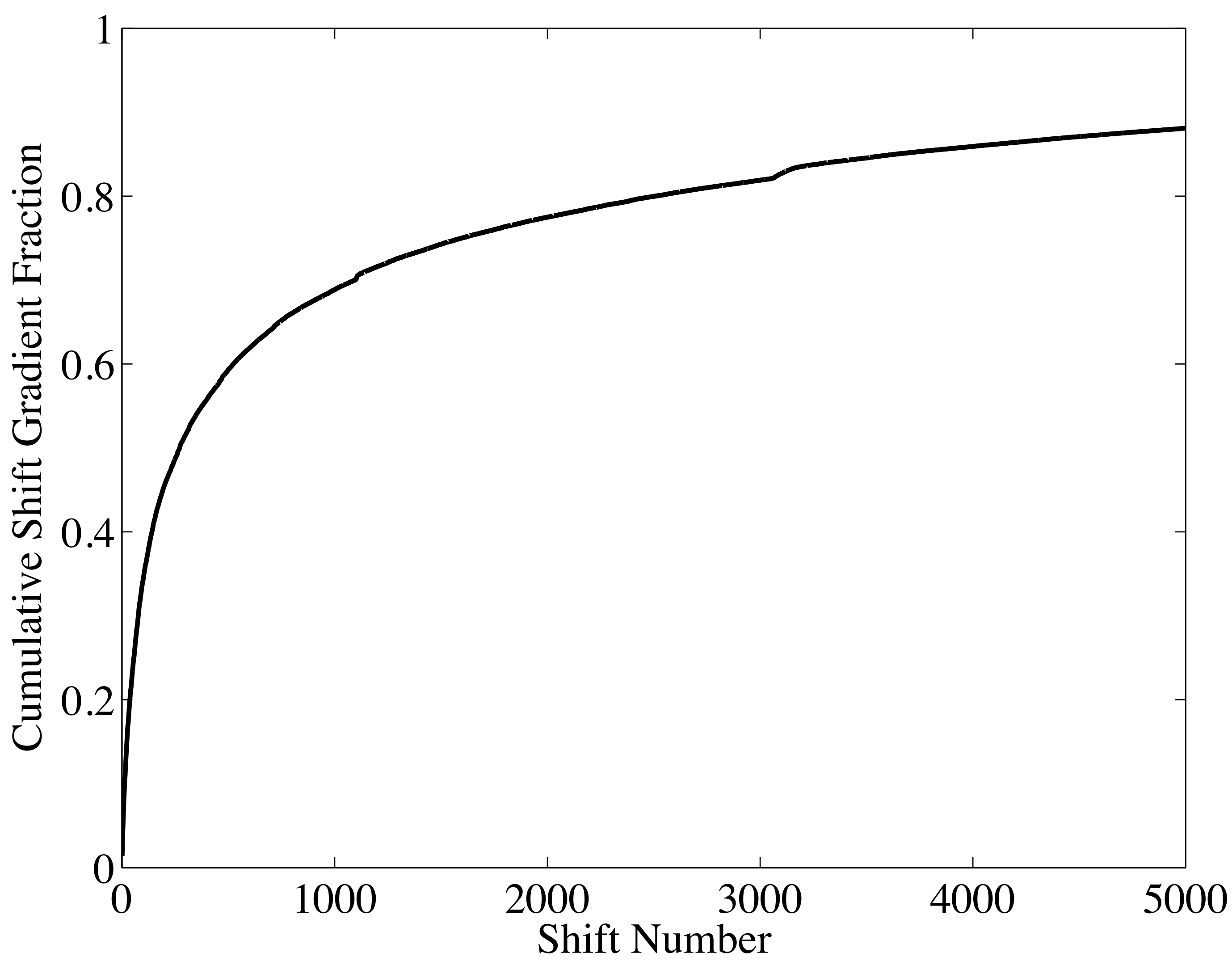
Mass Plot



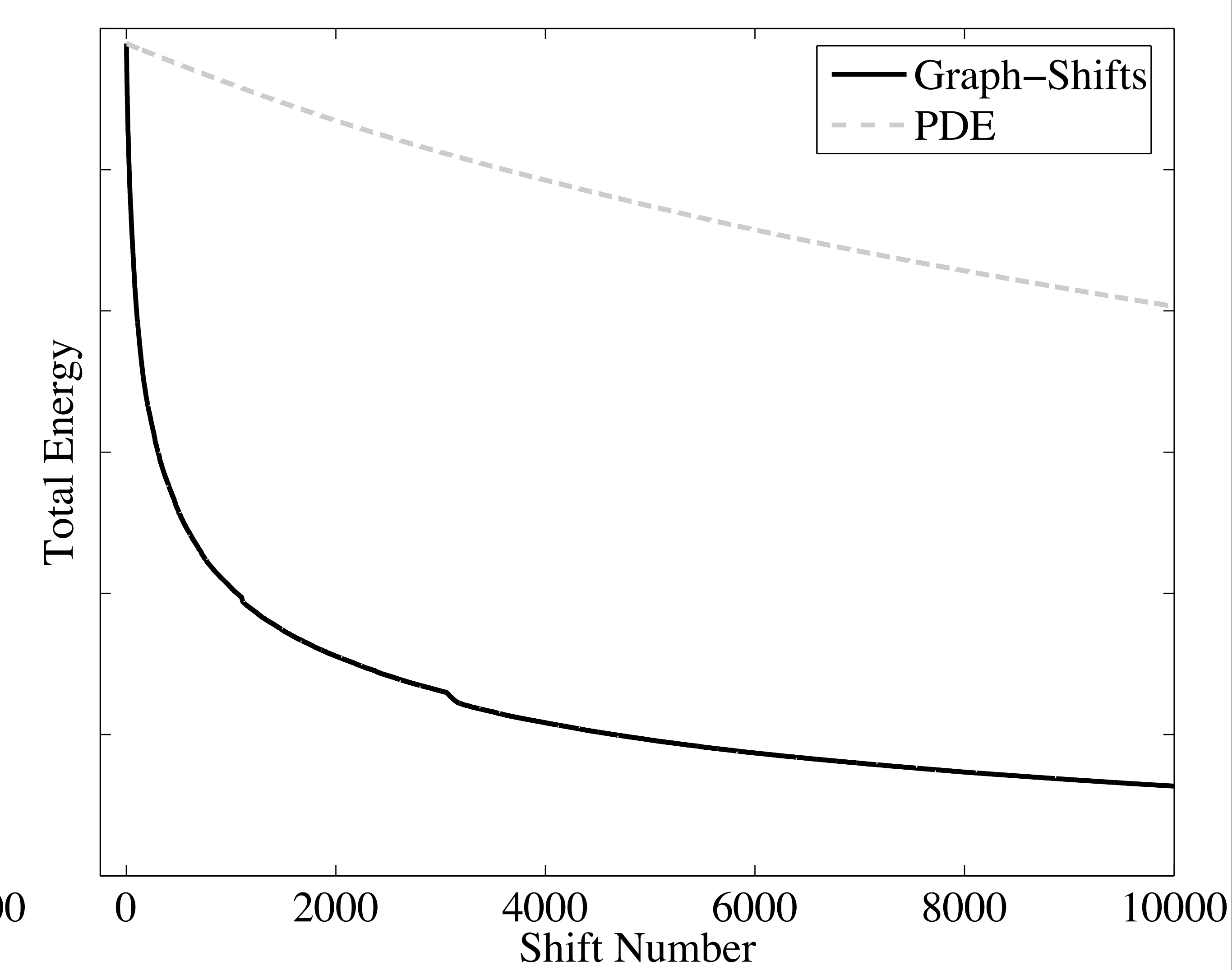
Layer Plot

Graph-Shifts Energy Reduction

- 75% of energy reduction occurs within first 1000 shifts.
- Minimum reached after 30,000 shifts (average) for high-resolution T1-weight MR image (180^3 voxels) in about 50 seconds.
- Orders of magnitude faster than state-of-the-art (5, 30, & 120 min).



Cumulative Gradient Plot



Total Energy Plot

Quantified Results

- Segment sub-cortical structures:
  - Hippocampus, Putaman, Caudate, Ventrices

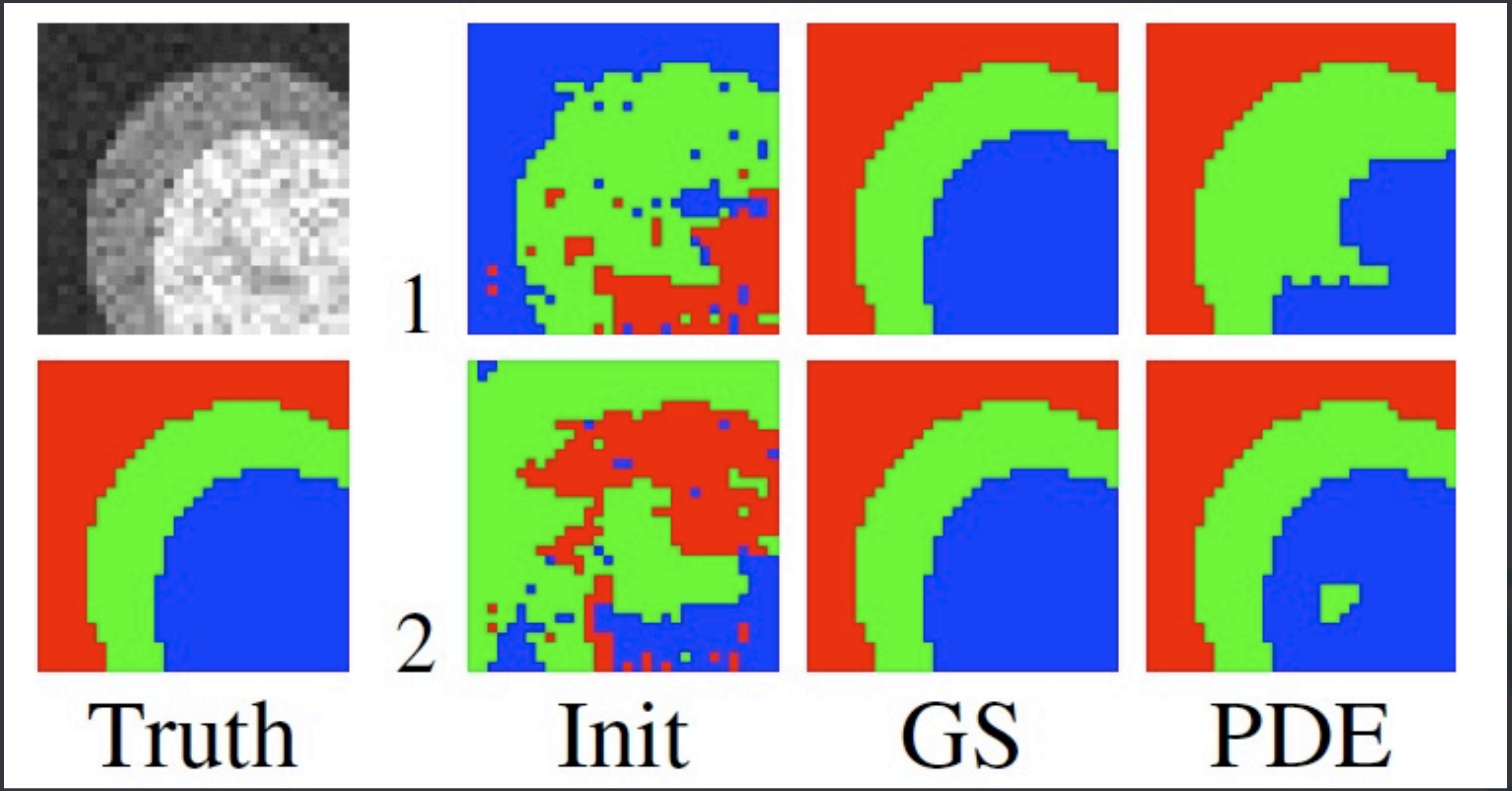
	Training Set								Testing Set							
	LH	RH	LC	RC	LP	RP	LV	RV	LH	RH	LC	RC	LP	RP	LV	RV
Prec.	82%	70%	86%	86%	77%	81%	86%	86%	80%	58%	82%	84%	74%	74%	85%	85%
Rec.	60%	58%	82%	78%	72%	72%	88%	87%	61%	49%	81%	76%	67%	68%	87%	86%
Haus.	11.4	21.6	10.1	11.7	14.7	11.6	26.9	19.0	17.1	26.8	10.4	10.1	15.7	13.7	20.8	21.5
Mean	1.6	4.0	1.1	1.1	2.3	1.8	1.0	0.8	1.8	7.6	1.2	1.2	2.7	2.5	0.9	0.9
Med.	1.1	3.1	1.0	1.0	1.4	1.2	0.4	0.3	1.1	6.9	1.0	1.0	1.6	1.6	0.4	0.5

- Comparison to state-of-the-art FreeSurfer Method:

	LH	RH	LC	RC	LP	RP	LV	RV
Prec.	48%	51%	77%	78%	70%	76%	81%	69%
Rec.	67%	75%	78%	76%	83%	83%	76%	71%
Haus.	25.3	11.5	23.0	26.1	13.1	10.8	31.9	51.8
Mean	3.9	2.1	1.9	2.0	1.8	1.4	1.8	9.6
Med.	2.1	1.5	1.0	1.0	1.3	1.0	0.9	3.9

Avoidance of Local Minima

- High-layer, large-mass shifts give potential to avoid local minima.
- Also increases robustness to initialization
  - Variance of precision and recall when doubling initial energy is 0.0001.
- Below is an example where graph-shifts avoids the minima, but PDE methods fail.



Video of Graph-Shifts in Action

Conclusions

- Graph-shifts is a novel energy minimization algorithm that manipulates a dynamic hierarchical representation of the image.
- It can include terms learned from training data.
- Exhaustively represents energy space and takes optimal move at each iteration.
- Comparable or superior accuracy to state-of-the-art on the difficult problem of sub-cortical structures.
- Run-time is orders of magnitude faster than state-of-the-art.
- Very robust to initialization and shown to avoid some local minima.
- Future
  - Extend class of energy to include model parameter estimation and common quadratic, and TV terms.
  - More sub-cortical structures and thorough experiment.
    - Algorithm scales logarithmically with number of structures to segment.
  - Apply to pathologic segmentation when number of structures is unknown at the outset.

- This work was funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 RR021813 entitled Center for Computational Biology (CCB). Information on the National Centers for Biomedical Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics.