Recent Developments in Electron Microscope Tomography

Albert Lawrence

Center for Research in Biological Systems
University of California, San Diego

October 13, 2011

Beijing
The National Center for Microscopy and Imaging Research

Mission:

- Develop and deploy technologies to determine and reveal supramolecular details in their cellular and tissue contexts.

- Focus on the ‘Meso-scale’ - \(~0.5\text{nm}\) to > \(100\text{um}\)
Solving the Puzzle

- Basic Biology
  - Genetic Control
  - Metabolic Pathways
  - Cellular Ultrastructure
  - How do these three aspects of the eucariotic cell fit together?
  - Compartmentalization, signaling and siting of cell processes.

- New Tools in Imaging
  - Fluorescence Microscopy
  - Electron Microscopy

- Faithful Representation of Biological Processes down to the Molecular Scale is an Exascale Problem
Why Exascale?

• Personalized medicine based on genetics of the individual
  – Doubling time of genetics database much shorter than Moore's law

• State model of a living cell is tremendously complex
  – Details of dynamics may extend over many orders of magnitude
  – Hardware limitations make it impossible to put chemical bond rearrangement and protein folding into a single dynamical model

• From EM to light microscopy:
  – Scale Change: 100K
  – Volume Change: 1 X 10^15
  – Details of dynamics may extend over amplification factor in biological processes
Cell death (Apoptosis) is an important component in the maintenance of the integrity of the organism. This fails in many cancers. Current non-surgical therapy techniques induce cell death in a highly non-selective way. More selective means of drug delivery are desirable.
Example: Problem of Drug Delivery in Cancer Therapy

- More selective drug delivery requires better knowledge of metabolic pathways and siting of critical reactions within the cell.
- Picture of cell ultrastructure may be incomplete. Current research tends to concentrate on specific organelles.
- Degree of organization of the cell nucleus and nucleolus is a subject of controversy.
- New techniques in microscopy can resolve these issues.
- Work on P53 gene may be a critical element of the research.
As the network becomes larger intuition becomes more fuzzy...
And problems in validation become more complex.
Inference can be computationally nontrivial for two reasons. In order to compute the partition function, the number of terms in the sum is equal to $m$ which can be very large since many applications of graphical models require that the models have large numbers of random variables. One may easily encounter $n = 200$ binary random variables, in which case

$$m = 1606938044258990275541962092341162602522202993782792835301376.$$ 

The success of graphical models has been due to the possibility of efficient inference for many models of interest. The organizing principle is the generalized distributive law which gives a recursive decomposition according to the graph underlying the model.
1961: Osamu Shimomura and Frank Johnson, University of Washington, observed a protein in jellyfish that produced green fluorescence when illuminated with ultraviolet light.

1992: Gene for the “Green Fluorescent Protein” (GFP) was cloned.

1994: Scientists inserted the gene for GFP into worms and bacteria and noticed that they glowed!

1994-present: Scientists have been attaching the gene for GFP to that of specific proteins, making glowing proteins. Can observe these glowing proteins in living organisms!
UCSD Scientist and NCMIR Co-Principal Investigator Shares 2008 Nobel Prize in Chemistry

Oct. 8, 2008 — UCSD Scientist Dr. Roger Tsien was awarded 2008 Nobel Prize in Chemistry along with Osamu Shimomura and Martin Chalfie, for their discovery of the green fluorescent protein GFP and its development as a tool for observing otherwise invisible cellular processes. Dr. Tsien is a leader in the development of new indicator systems and their application to cell biology and has been a driving force behind the NCMIR core efforts to develop improved labeling technologies for correlative light and electron microscopy.
Building Structural and Functional Descriptions Requires Many Tools

**Chemistry**
- Labeling and specimen development
- Molecular-specific structure detection and mapping
- Correlated light and electron microscopic imaging

Mark or introduce specific reporters for cells, target molecules and metabolic pathways

**Instrumentation**
- Microscope development
- Multidimensional imaging
- Multispectral detection
- Advanced optics
- Robotics for automation
- Large-scale and sensitive imaging detectors

Locate molecules in cells and observe their structural relationships and responses

**Algorithms**
- Increase experimental throughput, information content and quality of observations
- Analysis tools and AI
- Correlated data structures

Refine observations and connect the new knowledge to other findings and frameworks

\[ R_i^{\text{quadratic}}(x_j,y_j,z_j,\theta_j) = \left[ a_{\theta_1}x_j^2 + a_{\theta_2}x_j + a_{\theta_3}x_j y_j + a_{\theta_4}y_j^2 + a_{\theta_5}y_j z_j + a_{\theta_6}z_j^2 \right] \\
+ \left[ b_{\theta_1}x_j + b_{\theta_2}y_j + b_{\theta_3}z_j + c_{\theta_1} \right] \cdot \frac{1}{d_{\theta_1}x_j + d_{\theta_2}y_j + d_{\theta_3}z_j + 1} \quad (i=1,2) \]
Purpose of Electron Microscope (EM) Tomography Development at CRBS

- Increase the range of spatial scales of 3D reconstructions from EM images
  - Connect with light microscopy at longer scales ($10^{15}$ change in volume).
  - Improve resolution to level of molecular assemblies at shorter scales
- Integrate EM and fluorescence microscope data
  - Superresolution techniques localize emitting molecule to $\sim$1nm
  - Fluorescence microscope techniques give functional and dynamical data, EM gives structural context
- Integration of fluorescence microscopy data and EM data will become an increasingly important (if not essential) technique for the validation of genetic and molecular biology data.
- Speed requirements are a driver in computer technology
  - Increase in data set size over past 10 years: $10^5$-$10^6$
  - Improvement in processing times: $10^3$-$10^4$ (algorithms + GPU)
  - Actual decrease in turnaround times due to increase of parallelism
  - Long term objective: Real time tomography. Computational part invisible to user.
• Used for constructing 3D views of sectioned biological samples
• Sample is rotated around an axis and images are acquired for each ‘tilt’ angle
• Electron tomography enables high resolution views of cellular and neuronal structures.
• 3D reconstruction is a complex problem due to low signal-to-noise ratio, curvilinear electron path, sample deformation, scattering, magnetic lens aberrations...

High quality 3D reconstruction under these conditions is a difficult mathematical and computational problem...
The Results Should Be Worth the Trouble

Tomography Workflow

Tilt Acquisition

Track Fiducials

Align Fiducials

Remap

Filter

Backproject

GPU Code
• Tomography, in practice, requires many steps
  – Sample preparation
  – Data collection
  – Feature isolation and tracking
  – Image alignment
  – Image filtering
  – Volume reconstruction
  – Object segmentation
• Each step carries its own set of problems
• Choice of methods on one step affects subsequent steps
• Lab Code: Transform Based Tracking, Bundle Adjustment and Reconstruction
• In addition TxBR provides user choice of model, automated reconstruction, and statistical error reporting
EM Tomography: Technical Problems

- Noise. Particle flux is limited.
- Low contrast. Imaging is through electrons scattering out from beam.
- Missing data. No flux thru sample at high angles.
- Discretization. Number of angles exposure-limited.
- Positioning accuracy. Sub micron information required.
- Curvilinear electron trajectories. Electrons travel in helical paths in focusing fields.
- Sample mass loss. High energy electrons remove material. Sample warping.
- Imperfect lenses.
Focus is changed in steps so focal plane moves through object. Note effects due to Helical trajectories and differential magnification.
Projection in Electron Microscope

γ = electron path
Si = point of entrance
Sf = point of exit
θ = tilt angle
x = (x, y)\text{image}
X = (x, y, z)
u = object density
ν = image intensity

\[ \nu = \ln(\hat{v}(\theta; x)) = -\int_{Si}^{Sf} u(\gamma(\theta;x,s))ds \]

Linear “single scattering model”

\[ \hat{v}(\theta; x) = I_0 \cdot e^{-\int_{Si}^{Sf} u(\gamma(\theta;x,s))ds} \]
Mathematical Model

\[ \Gamma = \{ \gamma_{(\theta; x, y)} \} \]

Family of trajectories
Indexed by image point and sample angle.

\[ R_\Gamma(u)(\theta; x, y) = \int_{s=s_i}^{s=s_f} u(\gamma_{(\theta; x, y)}(s))ds \]

Transform defined by integration of density along trajectories.

\[ R_\Gamma^*(v)(x, y, z) = \int \gamma_{(\theta; x, y)}(s)=(x, y, z) R_\Gamma(u)(\theta; x, y)d\theta \]

Adjoint transform defined by integration over sample orientations.
Backprojection Along Curvilinear Trajectories
Tracks gold particles deposited on surface of sample through tilt series images.

- Accepts general set of sample orientations.
- Constructs series of geometrically nonlinear projections simultaneously with 3D model of gold particle positions via generalized bundle adjustment.
- Corrects projection maps to 9th order polynomials.
- Remaps (warp) tilt series images to align tracks to run orthogonally to projected tilt axis.
- Applies one of the common r-weighted filters to tilt series images.
- Backprojects via adjoint of curvilinear projection calculated in the bundle adjustment.
- Utilizes fast recursion, MPI code for backprojection.
• Aligns on surface contours, fibers and point features; reconstructs surfaces as during alignment process
• Dewarping of objects distorted by mass loss
• Backprojection code runs on GPU boards
• Cross validation for elimination of limited angle artifact, discretization artifact and sampling bias.
Alignment Requirements

- Alignment must be accurate over ~10K pixel images.
- For high quality reconstructions require average reprojection errors to within a pixel.
- Feature model must avoid distortions.
- “Gauge ambiguity” pushes nonlinearities to projection maps.
- Iterated bundle adjustment gives orthogonal feature model and polynomial projection.
Geometrical Characterization of the stage motion

Verification of tracking and alignment
Calculation of tilt axis allows correction of filter
Automated estimation of object dimensions
Quality Comparison: EM Tomography

Cardiac tissue reconstruction sections
• Surface $S_\rho$ of a 3D object $\rho$ is parameterized with $(t,u)$.

• We restrict $S_\rho$ to be small patches. Use of polynomial expressions for $S_\rho(t,u)$.

• Curvilinear rays tangent to surface. Use of a polynomial expression for the projection map $P_\omega$. Index $\omega$ represents a sample orientation.

• Contour in surface where $u_{\omega\rho}(t)$ projects to contour $C_{\omega\rho}$ in image $I_\omega$. 
Contour Alignment Model

- **Contour Tracks:**
  \[
  C_{\omega p} = (x_{\omega q_1}(t), x_{\omega q_2}(t))
  \]
  \[
  x_{\omega q_i}(t) = \sum_{k=0,\ldots,N_1} c_{\omega q_i k} t^k
  \]

- **Projection Map:**
  \[
  P_{\omega} (X_1, X_2, X_3) = (P_{\omega 1} (X_1, X_2, X_3), P_{\omega 2} (X_1, X_2, X_3))
  \]
  \[
  P_{\omega i} (X_1, X_2, X_3) = \sum_{N_2 \geq j, k, l \geq 0} b_{\omega i j k l} X_1^j X_2^k X_3^l
  \]

- **Structure Patches:**
  \[
  S_{\omega i} (t, u) = (S_{\omega 1} (t, u), S_{\omega 2} (t, u), S_{\omega 3} (t, u))
  \]
  \[
  S_{\omega i} (t, u) = \sum_{N_3 \geq k l \geq 0} a_{\omega i k l} t^k u^l
  \]

- **Surface Contours:**
  \[
  u_{\omega q} (t) \approx \sum_{k=0,\ldots,N_4} d_{\omega q k} t^k
  \]
A Generalized Bundle Adjustment

Two Error Terms to minimize:

• A Projection Error:

\[ E_{\omega q}^{P} = \int_{t_0}^{t_1} \left\| P_{\omega q} (t, u_{\omega q} (t)) - C_{\omega q} (t) \right\|^2 dt \]

\[ P_{\omega q} (t, u_{\omega q} (t)) = C_{\omega q}^{(r)} (t) \]

• A Tangency Error:

\[ P_{\omega i} \left( \gamma_{x,\omega}^1 (t), \gamma_{x,\omega}^2 (t), \gamma_{x,\omega}^3 (t) \right) = x_i \quad \nabla P_{\omega i} \cdot \dot{\gamma}_{x,\omega} = 0 \]

\[ E_{\omega q}^{T} = \int_{t_0}^{t_1} \left\| \nabla P_{\omega 1} \times \nabla P_{\omega 2} \cdot \left( \frac{\partial S_{\theta}}{\partial t} \times \frac{\partial S_{\theta}}{\partial u} \right) \right\|^2 dt \]
Structure Segmentation

Caulobacter Crescentus

Gia
Parameterization of the contours

What choice? For a tilt series:

- $t$ parameterizes the projection of a surface point onto the camera plane.
- $u$ parameterizes the tilt index

Simultaneous parameterization of tracks allows to assess optimal patch order to describe an object.

$\Rightarrow$ Minimization is then implemented with independent parameterization for each contours.
Initialization of the problem

- Coefficients $a_{gikl}, b_{wijkl}, c_{wqik}, d_{wqk}$ need to be adequately evaluated at initialization to avoid sticking to a bad local minima during optimization.

- Projection Map $b_{wijkl}$ is evaluated from experimental evaluation of sample orientation.

- $c_{wqik}$ is evaluated from tracking contours on images

- Dual space method to evaluate $a_{gikl}$. Dimension of dual space: 3d or 4d.

- Initialization of $d_{wqk}$ tied to choice for Parameterization.
Re-projection Error at Minimum

Bundle adjustment applied on Caulobacter Crescentus dataset (500ptsx500pts) with parabolic patches
Volume and Patches Reconstruction
Working Toward the Ultra-widefield

From montages to supermontages

3 Dimensional Montage of Tomograms

S-distortion mitigation and stitching
Towards Large High Resolution 3D reconstruction with TxBR

- Large high resolution 3D reconstructions can be achieved in Electron Tomography (ET):
  - with montages (X-Y directions)
  - with serial sections (Z direction)

- TxBR treats each tile series as an individual set, thus allowing for the best reconstruction quality; at the same time we are able to stitch automatically the mosaic components while the reconstruction (backprojection) is processing.

- TxBR offers the capability to flatten automatically a whole section by making use of physical markers (gold particles). This allows stacking of specimen sections that became warped during the data acquisition.
XY Montage Alignment with TxBR

- Extension of multiple tilt series methods (single area of interest with different orientations) to montages (different areas with overlap)
- Simultaneous alignment of all the tilt series
- No 2D stitching
- More variables (projection maps) to accommodate the 3D stitching of the tiles. Calculated during the bundle adjustment.

Input: $X_{\mu,\omega}^T$

Minimize:

$$E = \sum_{\tau} \sum_{\mu,\omega} \left\| P_{\mu,\omega}(X^T) - x_{\mu,\omega}^T \right\|^2$$

Output:

$$P_{\mu,\omega}$$

Index of exposure: $\omega$
Index of track: $\tau$
Index of tile: $\mu$

Traces
Projection Map
Marker Positions
XY Montage Reconstruction Example (I)

- XY Montage of a whole *drosophila* cell infected by *flock house* viruses
  - 37 tiles of 4096x4096 pixels within a 6x7 grid
  - Alignment using 4733 gold markers
  - Total volume of $20824 \times 19064 \times 1240$ pixels
  - Smooth transition in the intensity map between the tiles
  - Complex out-of-plane warping because of the long exposure time

(Data obtained on FEI Titan microscope)
XY Montage Reconstruction Example (II)

- XY Montage of part of a *drosophila* cell infected by *flock house* viruses
- Dual tilt series
- 18 tiles of 4096x4096 pixels within a 3x3 grid
- Alignment using 4733 gold markers
- Total volume of 11568x12140x497 pixels

(Data obtained on FEI Titan microscope)
Flattening Reconstructions within TxBR (I)

**General Approach**  
A flattening gauge is chosen for the bundle adjustment.

**Gauge Ambiguity:** Multiple solutions for \( P_\omega \) and \( X^T \) are possible as long as:

\[
P_{\mu,\omega}(X) = \tilde{P}_{\mu,\omega}(\tilde{X})
\]

with
\[
\tilde{P}_{\mu,\omega} = P_{\mu,\omega} \circ R
\]

\[
\tilde{X} = R^{-1}(X)
\]

**Flattening Transformation**
On the fly flattening reconstruction example for a single tile with a polynomial function of order 3

For large mosaic (more complex warping), a global flattening transformation might bring too much computational effort ... ⇒ Need to redefine the flattening transformation by pieces (for each tile)
P53 Gene Experiments

- Adenovirus infection turns off P53 (cell death) gene
- High-resolution, large-scale 3D reconstructions via TxBR montages
- Extensive structural reorganization of cell nucleus shown in EM tomographic reconstructions
- Indications of cell nucleus structure correlated with cytoplasmic structure
- Further indications of spatial organization of metabolic pathways
Segmentation of Cell Nucleus
Montage
Nuclear Reorganization after Adenovirus Infection
Averaging and More Sophisticated Techniques

- Additional in-plane rotations give better sampling and reduction of the “missing wedge”
- Additional data allows us to average out noise and artifact
- Can we go further?
  - Discretization artifact
  - Missing direction artifact
  - Oversampling of surfaces tangent to ray directions
Multiaxis Tilt Series Data Produces High Resolution in All Directions and Reduces Backprojection Artifacts
Six Fold Wide-Field Tilt Series

- Improved reconstruction quality
- No “bad direction” for samples through the reconstruction
Single Tilt Series Compared to Six-Fold Series

Single Tilt Series.
Tilt axis is horizontal.
Discretization and limited angle artifact layered in planes perpendicular to tilt axis

Difference Between Single and Six-Fold Series.
Note predominance of vertical features correlated to structures in single series
Artifact in EM Tomography

• Sources of Artifact
  • Discretization
  • Missing Data
  • Oversampling

• Properties of Artifact
  • Nonlocal
  • Image correlated
  • Wideband

• Artifact Suppression
  • Averaging
  • Nonlinear methods
  • Cross validation
Double Tilt Reconstruction
Reconstruction from Maxima
Reconstruction from Minima
Nonlinear Methods

- **Maslov Dequantization**

\[ \rho(r) = h \log \left( \sum_\theta \exp \left[ \frac{u_\theta(x)}{h} \right] \right) \]

- **Weighting by Mean Gradient**

\[ \rho(r) = x \rho_A(r) + (1-x) \rho_B(r), \quad \text{with} \quad x = \frac{\langle |\nabla \rho_B| \rangle}{\langle |\nabla \rho_A| \rangle + \langle |\nabla \rho_B| \rangle} \]

- **Weighting by Variance**
Maslov Dequantization

Top Row: $h = -100, -10, -1$; Bottom Row: $h = 1, 10, 100$
Mean Gradient
Weighting of Double Tilt by Mean Gradient
Variance
Weighting by Variance
Moreover, the use of templates provides a remarkable and unexpected unification of the families of linear convolutional filters and the nonlinear operations of mathematical morphology. These have a very similar appearance in image algebra, the sum and product operators that appear in convolutions being replaced by max (sup) and sum, respectively. The structuring elements of mathematical morphology are represented by templates, just like the filter functions of the convolutional procedures.

This algebra was introduced, as we have seen, to simplify and harmonize the many image processing algorithms. But why should we stop there? It would be very satisfying if we could express the whole chain – image formation plus image processing – in terms of this algebra.

• Pachter and Sturmfels - Algebraic Statistics for Computational Biology
Thank You for Your Attention

For further information:

http://ncmir.ucsd.edu/

http://nbcr.sdsc.edu/

And thanks to Mark Ellisman, Director NCMIR; Peter Arzberger, Director NBCR; Michael Holst; Sebastien Phan; James Bouwer; Rajvikram Singh; Masako Terada and many others for the materials of this presentation.