Pattern unmixing

• Some proteins may be found in more than one organelle
• Clustering sees each combination of organelles as a new pattern
• Can we “unmix” such mixed patterns?

Unmixing approach

• Assume that each fundamental subcellular pattern can be represented by some combination of distinct object types (10% small round objects and 90% long skinny objects)
• Assume that a mixed pattern is formed by adding together the objects from two or more fundamental patterns and that no new object types are created
Learning object types

- Find all objects in all images of fundamental types
- Describe each object by features such as size, ellipticity, distance from nucleus
- Cluster objects to find types
- Represent each fundamental pattern as probabilities of observing each object type

Test samples

- How do we test a subcellular pattern unmixing algorithm?
- Need images of known mixtures of pure patterns – difficult to obtain “naturally”
- Created test set by mixing different proportions of two probes that localize to different cell parts (lysosomes and mitochondria)
Pattern unmixing results

Communicating patterns

- How do we communicate results learned about subcellular patterns?
- Proposal: Use generative models learned from images to capture pattern and variation in pattern
Nuclear shape models

- Modified medial axis model
- Diffeomorphic model


Nuclear Shape - Medial Axis Model

Threshold
Rotate
Extract medial axis

Fit splines to two curves:
11 parameters
Shape generation

- 11 parameters for each object
  - 5 parameters for each curve
  - the length of the medial axis
- Learn the distribution of parameters over many nuclei
  - Assume multivariate normal
- Randomly sample parameters from distribution
- Construct nuclear shape using the sampled parameters

Synthesized nuclear shapes

Diffeomorphic analysis of nuclear shape

- Can use distance between shapes to characterize shape space instead of parameters of model – Gustavo Rohde
Concept: measure distances between all examples as means of characterizing shape space

LDDMM references

Finding deformation field
- Goal: Find a function $g(x,t)$ which smoothly transforms an image $I_n$ into an image $I_m$ as $t$ goes from 0 to T
- Choose $g(x,t)$ to minimize sum of
  - Total deformation in $g$ from 0 to T
  - Distance between $I_m$ and $I_n(g(x,T))$
Finding deformation field

Solve differential describing evolution from \( x \) to \( g(x) \):

\[
\begin{aligned}
\frac{dg(x,t)}{dt} &= v(g(x), t) \\
g(x, 0) &= x
\end{aligned}
\]

\[ v = \arg \min_{v(x,t)} \left( \int_0^T \| v(x,t) \|_t dt + \| I_m(x) - I_m(g(x,T)) \|_t^2 \right) \]

Finding deformation field

Geodesic distance between two images:

\[ d(L_m, L_m) = \int_0^T \| v(x,t) \|_t dt \]

Where

\[ \| v(x,t) \|_t = \| v(x,t) \|_{t_c} \]

for some operator \( t_c \).

Mapping two shapes to each other
Characterizing shape space

- Find deformation fields from each image to every other image
- Calculate distance between each pair of images as total deformation required between them
- Use multidimensional scaling (MDS) to find variables (principal components) that compactly represent variation
Finding mean shape

- For a population of images/shapes, find mean shape as that from which all shapes can be generated with minimum total deformation

Cell shape models

- Conditional radial distance ratio model
- Diffeomorphic model (in progress)

Examples of natural variation in cell shape
Cell Shape
Description: Distance Ratio

\[ r = \frac{d_1 + d_2}{d_2} \]

Represent single shape as vector of ratios for \( n \) angles and represent variation using PCA.

Diffeomorphic analysis of cell shape

Models for protein-containing objects

- Object library
- Gaussian objects
  - Mixture of Gaussians with number of objects determined from number of local minima
  - Learn distributions for number of objects and object size
  - Learn probability density function for object position relative to nucleus and cell shape
Modeling Vesicular Organelles

Position Model

r: normalized distance, a: angle to major axis

Synthesized Images

- Lysosomes
- Endosomes
- SLML toolbox - Ivan Cao-Berg, Tao Peng, Ting Zhao
- Have portable tool for generating images from model
Framework for conditional subcellular location models

- SLML: slots for different parts of cell model
  - Nucleus
  - Plasma membrane
  - Specific protein
- Each slot can hold one of multiple types of models, each of which is probabilistic
- Each slot’s model can be conditional (dependent) on another

Combining Models for Cell Simulations

Integrating with Virtual Cell (University of Connecticut) and M-Cell (Pittsburgh Supercomputing Center)