Automated construction of generative models from time series cell images: Tools for more complete analysis of perturbagen effects

Robert F Murphy
Ray & Stephanie Lane Professor of Computational Biology and Professor of Biological Sciences, Biomedical Engineering and Machine Learning
Carnegie Mellon University
External Senior Fellow, FRIAS, University of Freiburg
Overview

• Input assumed to be 2D or 3D movie (single or multichannel)
• Want to detect and model perturbations
  – Temporal pattern feature changes
  – Object type composition changes
  – Object type proximity changes
TEMPORAL PATTERN FEATURES
Automated analysis of protein subcellular location in time series images

Yanhua Hu¹,², Elvira Osuna-Highley¹,³, Juchang Hua¹,²,⁴, Theodore Scott Nowicki¹,², Robert Stolz⁵, Camille McKayle⁵ and Robert F. Murphy¹,²,³,⁴,⁶,*

¹Center for Bioimage Informatics, ²Department of Biological Sciences, ³Department of Biomedical Engineering, ⁴Department of Machine Learning, Carnegie Mellon University, Pittsburgh, PA 15213, ⁵Division of Science and Mathematics, University of the Virgin Islands, St Thomas, VI 00803 and ⁶Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
Descriptive approach

• Calculate features that measure temporal patterns
  – Object tracking
  – Temporal texture
  – Normal (optical) flow
  – Fourier transform
  – Autoregression

• Use for classification or clustering
Results

- Evaluate ability to classify movies of 12 different GFP-tagged proteins

<table>
<thead>
<tr>
<th>Temporal feature type</th>
<th>Without static features</th>
<th>With static features</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>-</td>
<td>66</td>
</tr>
<tr>
<td>Object tracking</td>
<td>nd</td>
<td>68</td>
</tr>
<tr>
<td>Temporal texture</td>
<td>66</td>
<td>77</td>
</tr>
<tr>
<td>Normal flow</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Fourier transform</td>
<td>60</td>
<td>69</td>
</tr>
<tr>
<td>Autoregression</td>
<td>nd</td>
<td>59</td>
</tr>
</tbody>
</table>
Speed and utility

- Object tracking (slow)
- Temporal texture (fast, accuracy++)
- Normal flow (slow, accuracy++)
- Fourier transform (fast)
- Autoregression (fast)
• Can distinguish or group movies by their temporal patterns using these features
• As with most feature-based methods, limited ability to interpret differences
OBJECT TYPE COMPOSITION MODELS
Movie Analysis via Object Type Changes

• HeLa cells expressing GFP-tagged growth factor receptor-bound protein 2 (Grb2)
• TGF added at t=0
• 3D movie over 9.2 minutes
• 8 sec / frame
• Alexander Sorkin group, Univ. Pittsburgh School Medicine
HeLa cells expressing growth factor receptor-bound protein 2 (Grb2)

TGF added at t=0

single slice from 3D movie

A. Sorkin U. Pitt
Three “patterns” from visual analysis

Cytosol  Plasma membrane/Coated Pits/Vesicles  Internal Vesicles
Goal

• Analyze temporal dependence of pattern changes with minimal assumptions
• Major assumption: Patterns representable by composition of objects
  • “Bag of visual words”
• Can be calculated “on the fly”
Method

• Segment objects with adaptive thresholding
• Cluster objects by geometric features
• Describe frame as a vector of object type proportions
• Cluster vectors (specify number of clusters)
• Fraction of each pattern contained in model as normalized inverse relative distance to each cluster centroid

\[
ird_i = \left( \sum_{j=1}^{k} d_j \right) / d_i
\]

\[
nird_i = ird_i / \sum_{j=1}^{k} ird_j
\]

\[\text{(0.95, 0.025, 0.025)}\]

\[\text{(0.35, 0.35, 0.30)}\]
Capturing phases of pattern changes

Greg Johnson

Devin Sullivan
• Can find and visualize temporal pattern changes
• Still descriptive
OBJECT TYPE TRANSITION MODELS
Model building and intelligent acquisition with application to protein subcellular location classification

C. Jackson\textsuperscript{1,2}, E. Glory-Afshar\textsuperscript{1,2}, R. F. Murphy\textsuperscript{1,2,3,4,5} and J. Kovačević\textsuperscript{1,2,3,6,*}

\textsuperscript{1}Center for Bioimage Informatics, \textsuperscript{2}Department of Biomedical Engineering, \textsuperscript{3}Lane Center for Computational Biology, Carnegie Mellon University, 5000 Forbes Ave., \textsuperscript{4}Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Ave., \textsuperscript{5}Machine Learning Department, Carnegie Mellon University and \textsuperscript{6}Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA
Object type A =

Object type B =

Object type C =

$t=0$

$t=1$
Object type A =
Object type B =
Object type C =

A ⇔ B

$\text{t}=0$

$\text{t}=1$
Object type A = 
Object type B = 
Object type C = 

A \rightarrow B 

<table>
<thead>
<tr>
<th>From object type</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Object type A =
Object type B =
Object type C =

<table>
<thead>
<tr>
<th>From object type</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To object type</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A \rightarrow B
C \rightarrow C

t = 0

t = 1
Modeling mitochondrial response to hyperosmotic stress

- 3T3 cells expressing GFP-tagged mitochondrial protein
- Pre-equilibrated with Hoechst 33342 to mark nuclei
- Add 5M NaCl to increase NaCl concentration by \( \approx 74\text{mM} \)

- Model Components
  - \( m_\lambda \) – \( k \)-by-1 vector representing the proportion of objects of type \( \lambda \)
  - \( m_{\lambda,\lambda} \) – \( k \)-by-\( k \) matrix representing the proportion of objects of type \( \lambda \) that have a nearby object of type \( \lambda' \) in the subsequent frame
  - \( m_{\lambda,0} \) – \( k \)-by-1 vector representing proportion of objects of type \( \lambda \) with no nearby objects in the subsequent frame
  - \( m_{0,\lambda} \) – \( k \)-by-1 matrix representing the proportion of objects of type \( \lambda \) that have appeared from no nearby object of type \( \lambda' \) in the previous frame
Adaptive image acquisition protocol

For each time point $t$
  Wait until next $t$
  Do
    Image 5 frames
    Add to model$_t$
  while model$_t$ error > error threshold

$t = 0$
$t = 961$
$t = 1921$
$t = 5761$
Example images

Before

After
Features for each frame pair

- Proportion of object types (7)
- Proportion of object transition types for each frame pair (63)
  - Object to object
  - Object disappear
  - Object appear
- 3D SLF (85)
- Z-scored all features
- Kalman smoothing
Conclusions: Temporal models

• Various approaches available
  – Extend features to capture spatiotemporal information
  – Learn changes in object composition over time
  – Learn generative model of how objects change, appear, disappear
Update on static generative models

• Have previously described methods for building generative models of nuclei, cell shape, organelle pattern
• Recently extended to 3D
• Collected tools under CellOrganizer framework
The **CellOrganizer** project provides tools for:

- learning generative models of cell organization directly from images
- storing and retrieving those models in XML files
- synthesizing cell images (or other representations) from one or more models

Model learning captures variation among cells in a collection of images. Images used for model learning and instances synthesized from models can be two- or three-dimensional static images or movies.

Current components of **CellOrganizer** can learn models of:

- cell shape
- nuclear shape
- chromatin texture
- vesicular organelle size, shape and position
- microtubule distribution.

These models can be *conditional* upon each other. For example, for a given synthesized cell instance, organelle position is dependent upon the cell and nuclear shape of that instance.

Cell types for which generative models for at least some organelles have been built include human HeLa cells, mouse NIH 3T3 cells, and Arabidopsis protoplasts. Planned projects include mouse T lymphocytes and rat PC12 cells.

Support for **CellOrganizer** has been provided by grants GM075205 and GM090033 from the National Institute of General Medical Sciences, by a Forschungspreis from the Alexander von Humboldt Foundation, and by the School of Life Sciences of the Freiburg Institute for Advanced Studies.
Overview

• Choose parametric or non-parametric way of representing a particular component (nucleus, cell shape, lysosome, microtubule) in a single cell (may be conditional upon other components)

• Combine results from many cells to build statistical model of variation -> model

• Randomly sample from model -> instance
Generative model structure

- Nuclear shape
- Cell shape
- Object positions
- Object number
- Object appearance
Example 3D instances
3D Endosome Tilt Series
Endosomes
Endosomes
Endosomes
Lysosomes
Mitochondria
Mitochondria
Chloroplasts Rotation Series
Combining Patterns in One Cell

Plasma membrane
Nuclear membrane
Endosomes
Lysosomes
Mitochondria
Multicomponent conditional models

Plasma membrane
Nuclear membrane
Microtubules
Lysosomes
Conclusions: Representation

- Generative model parameters are generalizable, transportable means for comparing and communicating effects of perturbagens across experiments, laboratories, cell types, and technologies.
Acknowledgments

• Past and Present Students and Postdocs
  – Michael Boland (Hopkins), Mia Markey (UT Austin), Gregory Porreca (Harvard), Meel Velliste (U Pitt), Kai Huang, Xiang Chen (Yale), Yanhua Hu (Bristol-Myers Squibb), Juchang Hua, Ting Zhao (Qiushi Acad), Shann-Ching Chen (St. Jude’s), Elvira Garcia Osuna (CMU), Justin Newberg (Baylor), Estelle Glory, Arvind Rao (M.D. Anderson), Henry Lin (Microsoft), Tao Peng (Microsoft), Luis Coelho (U Lisbon), Aabid Sharif, Rumi Naik, Josh Kangas, Jieyue Li, Baek Hwan Cho, Taraz Buck, Charles Jackson, Gregory Johnson, Devin Sullivan

• Funding
  – NSF, NIH, Commonwealth of Pennsylvania, Alexander von Humboldt Foundation, Freiburg Institute for Advanced Studies, Wellcome Trust

• Collaborators/Consultants