Markov Random Field Modeling of the Spatial Distribution of Proteins on Cell Membranes

Jun Zhang[1], Stanly L. Steinberg[2], Bridget S. Wilson[3], Janet M. Oliver[3], and Lance R. Williams[1]

Short Abstract — **Biologists use gold nanoparticles visualized by transmission electron microscopy (TEM) to label proteins on the cell membrane. However, due to limitations of applicable gold particle size and shape, typically only two proteins can be labeled at a time. A challenge is to integrate experimental data across multiple experiments where there are from 10 to 100 different proteins of interest and only the positions of two proteins can be observed simultaneously. Markov random field (MRF) is proposed to characterize the distribution of proteins on the cell membrane. We develop a multiscale MRF model and use mathematical programming techniques to infer the conditional distribution of a MRF for proteins of three types from observations showing the spatial distribution of only two types. Application to synthesized data shows that the spatial distribution of three proteins can be reliably estimated. Results on experimental data are also presented.**

I. Introduction

To observe events associated with cell signaling, several groups have generated high resolution topographical maps of colloidal gold nanoprobes marking receptors and signaling proteins in native membranes [1-3]. However, due to limitations of applicable gold particle size and shape, only two protein species can be labeled with confidence in the same experiment. Consequently, there is a need to integrate experimental data across multiple experiments where there are between 10 and 100 different proteins of interest and only the positions of two proteins can be observed simultaneously. We propose to use Markov random field (MRF) modeling to solve this problem.

MRF provides a powerful and robust framework for modeling correlations between states associated with neighboring sites in spatial lattices [4]. Cell membranes can be modeled as a 2D lattice, and the presence or absence of a protein of a specific type at a point on the cell membrane is a state. Since only two protein types can be observed, the problem is one of deducing the conditional distribution of a MRF with unobservable (hidden) states.

II. Modeling Protein Spatial Distributions

Let us consider a simple, idealized case, where there are three proteins of interest called *{R,G,B},* and where only two proteins can be observed in any single sample. For this scenario, we use a hidden process D with four hidden states *{R,G,B,X}* to characterize the distribution of proteins on the cell membrane. The additional state X corresponds to background. In addition, there are three observable processes, O_r, O_g and O_b, to model observations where only two kinds of proteins can be observed at a time. These processes have four observable states *{R',G',B',X'}*. There are also three observation matrices, $\boldsymbol{Q}_r$, $\boldsymbol{Q}_g$ and $\boldsymbol{Q}_b$, mapping the hidden states to the observable states. The problem is to infer the conditional distribution of D given O_r, O_g, and O_b.

A multiscale MRF model with three layers [5] is used for the hidden process, and multi-resolution representations of observed data at three scales are realizations of the observable processes. We use maximum likelihood estimation to infer conditional distribution of the hidden process, which is solved by using mathematical programming techniques.

Application of our approach to synthesized data has demonstrated that the multiscale MRF model is good at characterizing both short and long range statistical properties and that the spatial distribution of three proteins can be reliably estimated. Application to experimental data from two biological experiments has also demonstrated that the spatial distributions of three proteins in reconstructed samples are consistent with those in observed TEM images.

III. Conclusion

We have described a novel approach for reconstructing the spatial distribution of three proteins on cell membranes from samples showing only two proteins.

Acknowledgements: This work was funded in part by NIH grants R01 GM49814, R01 CA119232, R01 AI051575 and P20 GM065794.

[1]Department of Computer Science, (E-mail: jz@cs.unm.edu) [2]Department of Mathematics and Statistics, [3]Department of Pathology, University of New Mexico, Albuquerque, NM 87131.

References

[1] Wilson B, Pfeiffer J, Oliver J (2000) Observing FcεRI signaling from the inside of the mast cell membrane. *Journal of Cell Biology* **149**(5), 1131-1142.

[2] Prior I, et al. (2003) Direct visualization of ras proteins in spatially distinct cell surface microdomains. *Journal of Cell Biology* **160**(2), 165-170.

[3] Volna P, et al (2004) Negative regulation of mast cell signaling and function by the adaptor lab/ntal. *Journal of Experimental Medicine* **200**(8), 1001-1013.

[4] Li S (1995) Markov random field modeling in computer vision. Springer-Verlag, London, UK.

[5] Bouman C, Shapiro M (1994) A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing* **3**(2), 162-177.