Protein Dynamics Reconstruction from Unordered Images

Christian Choe[1,2], Min Cheol Kim[1]
Departments of Electrical Engineering[1] and Chemical Engineering[2], Stanford University

Introduction/Background

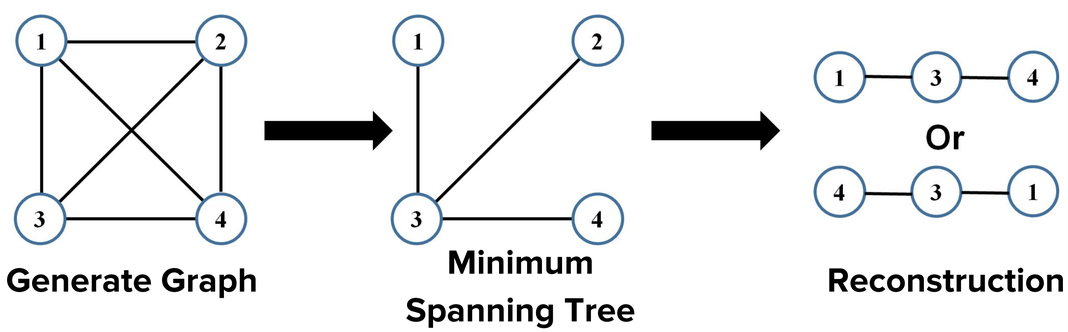
This problem is inspired by the data generated by the LCLS facility at the Stanford Linear Accelerator Center. This facility can take an "image" of a biological molecule, acquiring data such as the (x, y, z) coordinates of all the atoms. However, because the procedure destroys every molecule after each "snapshot," we end up with many images of the molecule in different states in its dynamics trajectory. The task is to order these images so that we can learn something about the dynamcics about the protein.

Methods

- **Nonlinear Dimensionality Reduction**

$$\begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} \xrightarrow{\text{Isomap}} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad f_i \in \boldsymbol{R}^{n \times m};\ x_i \in \boldsymbol{R}^{n \times r},\ r \ll m$$

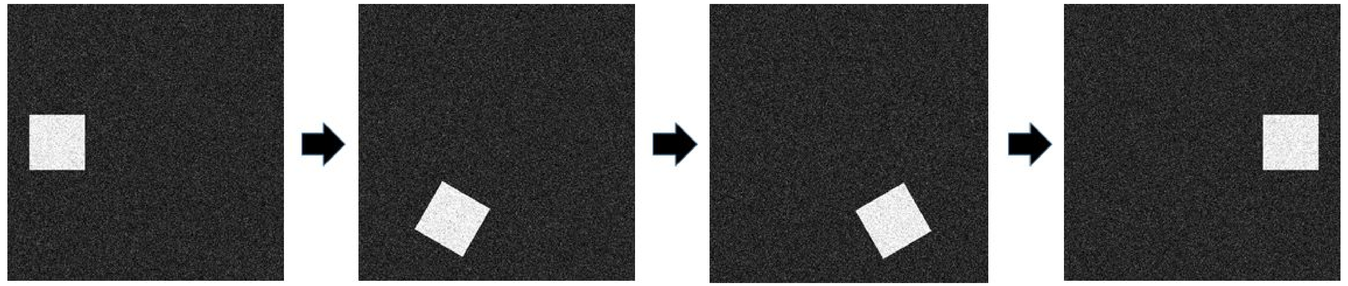
- **Frame order reconstruction**



Results

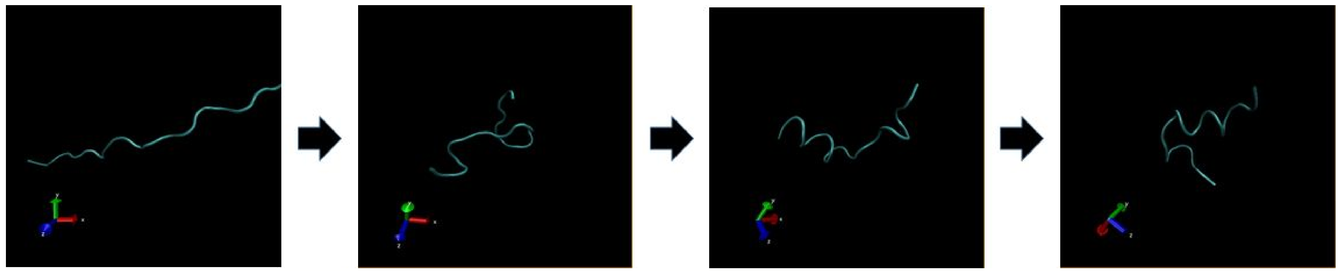
Reconstructing Path of Cube in 2D/3D

- 2D image of moving cube ✓
- 2D image of moving cube with noise ✓
- 3D image of spiraling cube ✓



Reconstruction of AK peptide folding simulation

- Raw XYZ simulation data ✓
- Phi/psi angles ✗
- XYZ data with alignment ✓



Conclusion/Future Work

Nonlinear dimensionality reduction via Isomap then reconstruction via a graph algorithm seems like a promising approach to reconstructing the protein dynamics trajectory. More improvements can be made both on the dimensionality reduction and on the trajectory reconstruction, by considering different algorithms (e.g., rearranging the minimum spanning tree, t-NSE instead of Isomap). This approach would have to be tested on the true SLAC data, which comes in a similar form as the simulations.