Image Alignment of Heterogeneous Macromolecules from Electron Microscopy

Maxim Shatsky[1,2], **Richard Hall**[2], **Robert Glaeser**[2], **Steven E. Brenner**[1,2]

Keywords: Particle alignment, heterogeneous data, 2D alignment, EM reconstruction

1 Introduction.

Electron microscopy (EM) is one of the key experimental techniques in modern structural biology[4]. Three-dimensional structures of large macro-molecular complexes can be solved by EM which otherwise would be impossible using X-ray crystallography or NMR. A common reconstruction technique, from 2D EM images to 3D structure, assigns 2D images to 2D projections of some initial 3D template[5, 1]. Due to high levels of noise and unknown orientation of imaged particles, the quality of the reconstruction procedure relies on two factors: an accurate alignment between a raw image and a template, and a scoring function that should be able to assign a raw particle to a correct template projection.

Most of the methods for particle alignment apply a very efficient Fast Fourier Transform to compute a superposition between two images. This approach implies that only a normalized cross-correlation (NCC) function, or its variants, can be used as a similarity measure between two images[5, 1]. One of the major problems in EM reconstruction is heterogeneous data. In such cases a single template model does not accurately account for the heterogeneous data. Consequently, an application of the cross-correlation function produces inaccurate alignments. Here we present an alignment method that is able to accommodate various similarity scoring functions while efficiently sampling the 2D transformational space. In our preliminary results we apply a scoring function based on Mutual Information of two images. For a heterogeneous sample containing incomplete molecules it allows accurate alignment using a model of the complete structure. In our preliminary results, we successfully tested our approach on a model data set containing a mixed population of 70S and 50S E. coli Ribosomes.

2 Method.

The method is composed of five major stages:

1. Feature detection. This stage is done separately for each image. First, we detect local extreme (maximum and minimum) values: $E = \{e_i\}$. Then, we define local significant values around each point $e_i \in E$: $N(e_i) = \{p : |value(p) - value(e_i)| > 2\sigma(e_i)\}$, where $\sigma(e_i)$ is the standard deviation of gray values in the neighborhood of e_i.

2. Construction of transformations. To define a 2D transformation that superimposes image A onto B we align two vectors, one from A and one from B. For each image the vector set is defined by $F = \{(p, e_i) : e_i \in E, p \in N(e_i)\}$. Therefore the set of 2D transformations is defined as the product of all possible vector pairs: $F(A) \times F(B)$.

3. Pose-clustering. We utilize an assumption that high scoring alignments have a large number of features with similar transformations. Therefore, the number of transformations (generated at Stage 2) can be significantly reduced by a pose-clustering technique[3]. After

[1]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA. E-mail: maxshats@compbio.berkeley.edu

[2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, CA 94720, USA.

the clustering only one thousand transformations, from the largest clusters, are evaluated with the scoring function.

4. Scoring Function. Let $p_A(a)$ define a distribution of pixels with gray value a in image A (the marginal probability) . Given two aligned images A and B let $p_{AB}(a, b)$ define a distribution of aligned pixels with gray values a and b (the joint probability). Then, the mutual information [2] of two aligned images is given by: $MI(A, B) = \sum_{a,b} p_{A,B} \log \frac{p_{AB}(a,b)}{p_A(a)p_B(b)}$.

5. Re-scoring. Due to different gray level distributions, some images tend to receive high scores even when aligned to an incorrect template. We apply an iterative rescoring that selects alignments with the highest, but non-random scores.

3 Results.

Due to the problem of reconstructing heterogeneous data there are no available experimental examples to validate our approach. Therefore, we selected the well studied structure of the 70S ribosome and its 50S sub-complex. 70S ribosome was used as a template for the alignment of both 70S and 50S projections. As expected, 70S projections were correctly assigned to the template by our method as well as by the standard NCC based methods, Imagic[5] and Spider[1]. However, NCC based methods give 100% errors in assigning 50S projections to the 70S template even in the case of noise-free projections. Our method correctly assigns more than 50% of the projections with the signal to noise ratio (SNR) greater than one. The results are summarized in the Table 1. Our current work in progress is to analyze whether the accuracy of our method is sufficient for the final 3D reconstruction.

	Images are Pre-aligned		Complete Alignment	
	num of errors	after rescoring	num of errors	after rescoring
no noise	0	0	1	0
SNR=5	23	6	182	65
SNR=2	9	9	175	104
SNR=1	58	67	166	110
SNR=0.5	170	100	178	145

Table 1: Alignment of 50S ribosome projections against the template of 70S ribosome. There are 212 projections of each macromolecule. Noise is added to 50S projections to simulate real data. The first column shows the number of 50S projections assigned to a wrong template. The second column shows the same type of errors after applying the iterative re-scoring. In the first, easy, test no transformation is applied (images are pre-aligned), this is to verify the scoring function. In the second test the complete alignment is performed.

References

[1] J. Frank. Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu Rev Biophys Biomol Struct*, 31:303–19, 2002.

[2] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multi-modality image registration maximization of mutual information. In *MMBIA '96: Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*, page 14, Washington, DC, USA, 1996. IEEE Computer Society.

[3] C. F. Olson. Time and space efficient pose clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 251–258, Seattle, Washington, 1994.

[4] A. Šali, R. Glaeser, T. Earnest, and W. Baumeister. From words to literature in structural proteomics. *Nature*, 422:216–225, 2003.

[5] M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, and A. Patwardhan. Single-particle electron cryo-microscopy: towards atomic resolution. *Q Rev Biophys*, 33:307–69, 2000.