

"Molecules in the Mist"

Informatics Challenges in Cryo-Electron Tomography

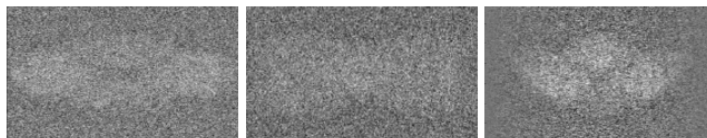
Christoph Best

Dept. of Molecular Structural Biology, Max-Planck-Institut fuer Biochemie,
Martinsried, Germany (WWW: <http://www.rzg.mpg.de/~cbest>)

Cryo-electron microscopy is a unique biological imaging technique that provides molecular-resolution volumetric images of macromolecular complexes and subcellular compartments in a near-native state without the need for invasive labeling.

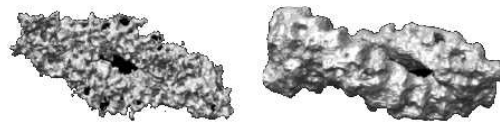
Recent progress in specimen preparation and automation of the acquisition process poses new challenges for image informatics to reliably and efficiently process and analyze these images and to extract meaningful biological information. In contrast to classical computer vision and fluorescent-light microscopy, cryo-electron microscopy is affected with a very low signal-to-noise ratio and thus requires sophisticated methods to reliably infer high-resolution information by combining a large number of images. Currently, up to $O(10^5)$ images are regularly used to obtain resolutions as high as 4 Å.

Informatics techniques, especially from machine learning and bioinformatics, help in dealing with this problem by providing a framework in which noisy and unreliable data can be efficiently integrated and combined with prior knowledge. We present give several examples of such methods in our recent work: Non-parametric robust estimation methods can be used for reliable registration and tilt-geometry estimation even for extremely low-contrast images without strong features. Level set methods that allow the incorporation of prior knowledge about the geometry of objects such as surface features and solidity can replace classical tomographic reconstruction methods like weighted backprojection. Spectral clustering and probabilistic modeling of image space are used for three-dimensional projection image classification and "computational purification" of the sample, while supervised learning with support vector machines is used for image selection.

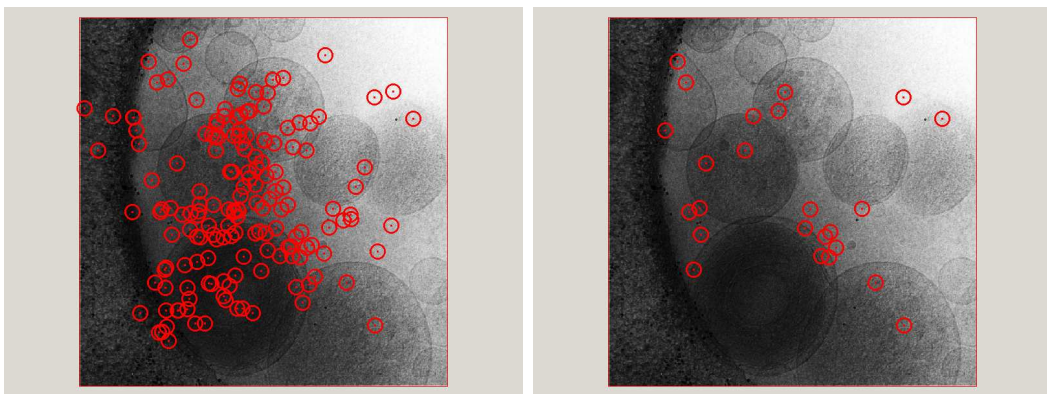


On the technical side, XML-based data formats for image data have turned out to be useful for solving annotation problems, and the Python/C++ environment allows for efficient combination of fast algorithms with interactivity and rapid prototyping.

Finally, these challenges will not remain unique to cryo-electron microscopy. Automation of fluorescent light microscopy especially in the area of developmental biology and in particular of the brain, e.g. in *in vitro* studies of neuronal stem cells, will pose similar problems of the cellular level, while on the molecular level future imaging techniques being planned today such as lens-less imaging with extremely bright X-rays from free-electron lasers will require the integration of ever larger number of images to obtain atomic-resolution molecular structures.



Level set reconstruction of the TPP II macromolecular complex from extremely noisy data.



Robust registration and geometry estimation: selection of tracking regions (Image courtesy of G. Kong, Martinsried).