
Markov Random Field Based Automatic Image Alignment for Electron Tomography

Farshid Moussavi*

Department of Electrical Engineering
Stanford University
Stanford, CA 94305
farshid1@stanford.edu

Fernando Amat*

Department of Electrical Engineering
Stanford University
Stanford, CA 94305
famat@stanford.edu

Luis R. Comolli

Life Sciences Division
Lawrence Berkeley National Laboratory
Berkeley, CA 94704
lrcomolli@lbl.gov

Gal Elidan

Department of Computer Science
Stanford University
Stanford, CA 94305
galel@cs.stanford.edu

Kenneth H. Downing

Life Sciences Division
Lawrence Berkeley National Laboratory
Berkeley, CA 94704
khdowning@lbl.gov

Mark Horowitz

Department of Electrical Engineering
Stanford University
Stanford, CA 94305
horowitz@stanford.edu

1 Introduction

Cryo electron tomography (cryo-ET) is the primary method for obtaining 3D reconstructions of intact bacteria, viruses, and complex molecular machines ([7],[2]). It first flash freezes a specimen in a thin layer of ice, and then rotates the ice sheet in a transmission electron microscope (TEM) recording images of different projections through the sample. The resulting images are aligned and then back projected to form the desired 3-D model. The typical resolution of biological electron microscope is on the order of 1 nm per pixel which means that small imprecision in the microscope's stage or lenses can cause large alignment errors. To enable a high precision alignment, biologists add a small number of spherical gold beads to the sample before it is frozen. These beads generate high contrast dots in the image that can be tracked across projections. Each gold bead can be seen as a marker with a fixed location in 3D, which provides the reference points to bring all the images to a common frame as in the classical structure from motion problem. A high accuracy alignment is critical to obtain a high resolution tomogram (usually on the order of 5-15nm resolution).

While some methods try to automate the task of tracking markers and aligning the images ([8],[4]), they require user intervention if the SNR of the image becomes too low. Unfortunately, cryogenic electron tomography (or cryo-ET) often has poor SNR, since the samples are relatively thick (for TEM) and the restricted electron dose usually results in projections with SNR under 0 dB. This paper shows that formulating this problem as a most-likely estimation task yields an approach that is able to automatically align with high precision cryo-ET datasets using inference in graphical models. This approach has been packaged into a publicly available software called RAPTOR-Robust Alignment and Projection estimation for Tomographic Reconstruction.¹

*These authors contributed equally to this work.

¹[1] presents an extended version of the results reported in this abstract.

2 Probabilistic Image Correspondence

We seek the most likely projection model given the observed 2D image data. Finding this model involves three steps: detecting marker locations in each image, recovering the marker trajectories across the tilt series, and finally fitting a projection model to the estimated trajectories. The unknowns of the problem are therefore the 3D marker locations, the 2D marker locations in each image, and the projection model. Let \mathbf{R} be a matrix containing the 3D marker locations, \mathbf{S} be a matrix containing the true 2D locations of the markers in each image, and \mathbf{O} be the noisy measured values of \mathbf{S} . The vector of parameters θ is the projection model that relates \mathbf{R} to \mathbf{S} . Finding the best projection model is now the same as finding the arguments which maximize the probability of \mathbf{R} , \mathbf{S} , and θ given the observed data \mathbf{O} . This probability can be expressed as the product of two terms:

$$P(\mathbf{R}, \mathbf{S}, \theta | \mathbf{O}) = P(\mathbf{R}, \theta | \mathbf{S}, \mathbf{O}) P(\mathbf{S} | \mathbf{O}) \quad (1)$$

The right-most term is the solution to the correspondence problem, estimating the probabilities of the ordered marker positions given the image data. The preceding term gives the probability of a projection model, given the correspondence. This decomposition suggests solving two maximum likelihood problems: one for finding markers and trajectories, and the other estimating projection model. The latter step has a convex formulation, and can be solved efficiently. Therefore our main focus here is the marker detection and correspondence.

2.1 Marker Detection and Correspondence for One Pair of Images

Markers are detected by cross correlation with a template. The template is generated in two steps. Firstly, three patches containing markers are selected by hand and averaged to obtain a rough template. This rough template is then used to find many more patches in low tilt (higher quality) images for further averaging, thus obtaining a refined final template.

The correspondence problem is divided into a subproblem of corresponding markers in one pair of images, and another subproblem of obtaining global correspondence and building trajectories for markers. Corresponding groups of markers in two images amounts to assigning each of the markers in a group in one image to one (or none) of the markers of the other image and is described below. The global correspondence (described in the next section) combines the results of image pair correspondence.

Once we have the list of probable marker locations in each image from the first step, we need to correspond those locations in two images. Focussing on one marker at a time can be very prone to error, since multiple candidates for that marker may exist which are close together and difficult to discern. For this reason, it is preferable to correspond groups of markers at a time using contextual information. For example, five markers may form a rather unique pattern, or "constellation", which is easier to match in a candidate image.

In order to utilize multiple markers in each decision, we define a distribution over the assignments of the markers of one image to another. Specifically, let $\mathbf{A} = \{A_1, \dots, A_M\}$ be the set of markers in the first image and let $\mathbf{B} = \{B_1, \dots, B_K\}$ be the set of markers in the second image. Also, let I_1 and I_2 be the two images themselves. We now treat each A_i as a random variable that can take a value in the set \mathbf{B} in addition to an "unassigned" value. The joint distribution $P(A_1, \dots, A_M | I_1, I_2)$ assigns a probability value for each assignment of the markers in \mathbf{A} . Our goal is to find the assignment that maximizes this distribution. That is,

$$\mathbf{A}^* = \operatorname{argmax}(P(\mathbf{A}_1, \dots, \mathbf{A}_M | I_1, I_2)) \quad (2)$$

Since the complexity of this inference is in general exponential with the number of markers M , we simplify the problem by assuming conditional independence, allowing us to use factorized distributions. Therefore, we define a Markov Random Field (MRF) over the set \mathbf{A} . For computational efficiency, we limit ourselves to factors over at most two variables. Therefore, our choice of model for the distribution of \mathbf{A} is:

$$P(A_1, \dots, A_M | I_1, I_2) = 1/Z \prod_i \phi_{s_i}(A_i) \prod_{i,j} \phi_{p_{i,j}}(A_i, A_j) \quad (3)$$

This distribution encodes a cluster graph in which there is a clique for each factor. There is a singleton factor (and therefore clique) for each variable. The pairwise factors (and corresponding cliques) indicate direct dependence between a pair of markers, and are chosen to exist only if that pair of markers are within some fixed distance of each other. This distance is a model parameter.

To complete the specification of the MRF, we need to define the singleton and pairwise factors (also known as potentials). Rather than model the underlying physics and geometry behind the projections, we try to emulate the preferences that a human doing manual correspondence would have. These are primarily proximity and similarity. A single marker is more likely to correspond to one candidate if they are physically similar (measurable by cross correlation), and if they are within close proximity of each other. By the same token, a pair of markers is considered more likely to correspond to a pair of candidates when the lines that each pair form are similar (similar distance and angle), and if the two pairs are within close proximity of each other.

To capture these preferences, we first define the following exponential rolloff factor:

$$f_k(\vec{a}, \vec{b}) = \exp(-(\|\vec{a} - \vec{b}\|/k)^2) \quad (4)$$

We can now define the singleton and pairwise potentials formally:

$$\phi_s(i_1, j_1) = NCC(i_1, j_1) * f_{k_1}(\vec{x}_{i_1}, \vec{x}_{j_1}) \quad (5)$$

$$\phi_p(i_1, i_2, j_1, j_2) = f_{k_2}(\vec{v}_i, \vec{v}_j) * f_{k_1}(\vec{x}_{i_1}, \vec{x}_{j_1}) * f_{k_1}(\vec{x}_{i_2}, \vec{x}_{j_2}) \quad (6)$$

where $NCC(i, j)$ is the normalized cross correlation score of the two patches containing markers i, j ; k_1 and k_2 are static constants; and \vec{v}_i is a vector connecting marker i_1 to marker i_2 .

With the model fully specified, we now need to run inference. We use a variant of loopy belief propagation (LBP) ([3], [9]) called residual belief propagation (RBP) ([6]), which has good convergence properties and is computationally efficient.

2.2 Global Correspondence and Subsequent Alignment

While the pair-wise correspondence now makes use of all the markers in the two images, it still only considers two images at a time. To generate the marker trajectories across the tilt series, \mathbf{S} , we compare each image with several of its following images. We apply a multi-level approach to use the redundant information both to allow trajectories to skip over a missing marker in some images, and to reduce the probability that a bad correspondence will affect the global trajectories.

The first level contains correspondences between the j^{th} image and the $j + 1^{th}$ image. The second level contains the correspondences between the j^{th} and the $j + 2^{th}$ image. We use up to 3 levels in images spaced up to 2° apart because after that it is difficult to find correspondences due to the tilt difference between projections. We initialize the first trajectory as the first point in image 1 and level 1. We call this point p11 (in general, we define pij as the point in trajectory i and image j). Local correspondence gives us the location of this trajectory in image 2, i.e. p12. Next, we try to find p12 in the local correspondence between image 2 and image 3. If we find it, that returns p13 and we repeat the process for this new location to find p14.

If a local correspondence for a track is missing, we look in the next level of the global correspondence to complete the track. For example, if we cannot find p13 from p12 in the first level, we try to find p14 from p12 in the second level, which contains local correspondence between image 2 and image 4. If we find p14, we continue the process for this new location to try to extend the trajectory. If we do not find it, we try to find p15 from p12 in the third level, which contains local correspondence between image 2 and image 5. We only use the latest point found in the trajectory to try to find the next one. We iterate this process until the first trajectory can not be extended longer,

either because we reach the last image (full trajectory) or because one location can not be found in any level (partial contour).

To create new trajectories, we use points in each level that have not been used in previous trajectories, since they are potential seeds of new trajectories. Having multiple match levels allows trajectories to recover from occlusion.

The second way of using a multi-level approach is to add confidence in the trajectories obtained in the first step. We do this using the redundant information contained in the local correspondences that have not been used in the first step. For example, suppose we found p12, p13, and p14 using the first step. We can check this path against the local correspondence for p12 in image 4, which is contained in the second level. We count the number of times such comparisons fail in each trajectory. If this occurs more than one fifth of the length of this particular trajectory, we drop the trajectory. Otherwise, we retain the consistent part of the trajectory.

Finally, only trajectories of certain length are kept to estimate the projection model. The length is selected automatically to assure that we have enough points in each image to estimate the projection model.

3 Results

Table 3 shows an exhaustive comparison between RAPTOR and a manual alignment by an expert user using IMOD [8]. A cross-validation technique to assess resolution (NLOO3D) presented in [5] was used to compare performance in variety of scenarios². In most of them, RAPTOR performance is comparable or even better than manual alignment. In addition to cryo-ET tomograms, which are the main focus of this project, we successfully tested RAPTOR in other tomography scenarios relevant in biology like X-Ray and plastic embedded sections, where the challenges and imaging conditions vary from cryo-ET.

Analysing the results makes clear that the method has a limitation in the number of markers that can be tracked. Datasets with over 100 markers present a challenge. In these cases, pairwise correspondence using inference still works. However the recovered trajectories are too short because with the current multilevel approach it is difficult to keep track of all the markers across several images without losing them. An immediate future direction is to improve this step.

RAPTOR also usually performs better if the degree separation between projections is less than $1 - 2^\circ$. This corresponds to the assumption in the model that relative position and orientation between markers in adjacent projections do not change much. In biological tomographic applications this is not a limitation, since it is very unusual to find datasets with a larger degree of separation than 2° .

Figure 1 shows a typical projection in cryo-ET and results obtained with RAPTOR. Because of the presence of the whole bacteria, partial trajectories are common due to occlusion. Panels(C)-(D) show how jerky the tilt series can be due to the fact that any small mechanical imprecision in the microscope stage causes big jumps in the image at this resolution. Therefore smooth motion constraints are not applicable in this scenario.

Finally, the system was implemented in a combination of C and matlab stand-alone application for Linux machines. Currently several labs are testing the software and it is available at <http://www-vlsi.stanford.edu/TEM/software.htm>. Execution is very data dependant and so far no optimization or parallelization efforts have been made, although it is an immediate future direction. Typical figures in terms of execution time are 45 min to align a set of 65 2Kx2K images with 20 markers, and roughly 4 h to align a set of 80 images, (2Kx2K in size), targeting 70 markers per image in a 2.4 GHz Pentium 4 PC with 2 GB of RAM. Compared with hours of expert user manual intervention labs can benefit from this approach.

4 Conclusions

In this paper we have presented a new approach to solve the correspondence problem between points in two images under very noisy conditions (SNR lower than 0dB) using MRF without assuming any

²[8, 5] are well-established tools in the cryo-ET community.

Id	Tilt range (deg)	Aprox. number beads	Bead diam. (pixels)	SNR (hi tilt)	number patches	NLOO3D (nm)			
						manual (0.3)	RAPTOR (0.3)	manual (0.5)	RAPTOR (0.5)
CrC1	-58:2:54	15	10	0.038	200	14.26	13.81	30.16	24.48
CrC2	-34:1:36	10	10	0.026	196	11.02	8.63	15.11	14.25
CrC3	-62:2:58	15	10	0.007	199	18.01	16.35	24.82	26.78
CrC4	-61:2:54	10	10	0.009	184	25.21	32.54	36.59	41.14
CrC5	-60:1:60	70	10	0.294	200	10.35	10.34	13.32	13.18
CrC6	-68:2:68	20	10	0.005	190	17.27	18.68	24.09	27.96
CrC7	-54:1.5:51	60	10	0.024	195	15.44	18.34	27.01	27.89
CrC8	-45:1:45	20	10	0.066	195	13.25	13.63	20.01	20.49
CrC9	-60:1:62	10	8	0.032	199	7.51	7.66	14.32	16.95
CrD1	-62:2:57	30	10	0.044	200	19.28	19.64	29.13	30.15
CrD2	-60:1:57	20	7	0.008	190	14.75	10.56	20.55	15.18
CrD3	-25:1:26	40	10	0.044	199	21.04	19.38	30.93	27.65
CrD4	-62:2:62	15	10	0.001	196	-1	-1	-1	-1
PI1	-58:2:60	>200	4,8	0.471	93	23.92	26.69	42.15	45.42
PI2	-60:2:60	>200	8	1.653	36	19.55	25.45	28.07	32.51
XRC1	-90:2:90	30	10	0.095	200	*	*	53.21	52.35
XRC2	-90:2:90	25	10	0.050	200	57.14	56.82	71.43	73.65

Table 1: Columns two to six contain information to present the variety of scenarios where RAPTOR was tested: spacing between projections, approximate number of gold beads to track in the dataset, bead diameter in pixels, SNR of the gold beads in high tilt angles, and number of patches obtained to calculate the SNR. The last four columns show a quantitative comparison of the resolution obtained using human-directed alignment and the automatic full-precision alignment presented in this abstract. The method used for resolution assessment is noise-compensated leave-one-out (NLOO) by Cardone et al., using threshold values 0.3 and 0.5. RAPTOR achieves better or equal resolution (lower numbers in the NLOO3D columns) in many of the datasets. (*) Curve did not reach that point; (-1) method fails to align.

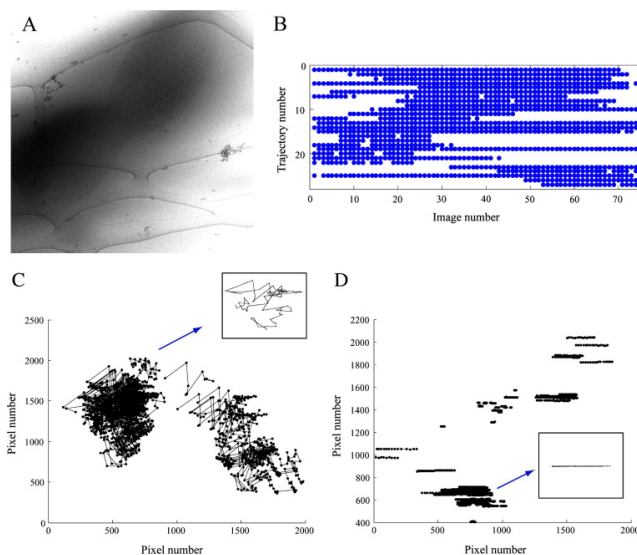


Figure 1: RAPTOR performance tracking markers in a *Caulobacter crescentus* dataset, and aligning the set. (A) The $-58deg$ 2D original projection from the tilt series. The size of the image is 2048 pixels on edge. The cell forms a meniscus of considerable thickness. Both the cell and the grid bars may occlude markers at high tilts. The wider area projected into the images at these angles contains many markers with appropriate SNR, far from the cell, which disappear at lower angles. (B) Trajectories recovered from the tilt series. (C) Overlay of several marker locations in image space (xy coordinates in pixel number) throughout the raw datasets illustrates the trajectories in the raw dataset, (D) trajectories of same markers after RAPTOR automatic alignment. The insets in (C and D) show the trajectory of a single marker [1]

specific transformation between two images. In combination with other techniques graphical models helped to solve the alignment problem for cryo-ET datasets completely automatically, saving expert user time that can be better spent.

The method presented very good results in comparison with a manual alignment done by an expert user in a variety of scenarios. However, when the number of markers to correspond is too large, the global correspondence step tends to lose trajectories. Future work will address this problem. A second limitation is the model assumption that relative distances and orientations between markers does not change much between adjacent images. This assumption is valid in electron tomography but might not be valid in other applications.

Acknowledgments

The authors would like to acknowledge Professor Daphne Koller of Stanford University Department of Computer Science, and Professors Lucy Shapiro and Harley McAdams of Stanford University School of Medicine for their unwavering support of this work.

References

- [1] F. Amat, F. Moussavi, L.R. Comolli, G. Elidan, K.H. Downing, and M. Horowitz. Markov random field based automatic image alignment for electron tomography. *Journal of Structural Biology*, In press, 2007.
- [2] W. Baumeister. Electron tomography: towards visualizing the molecular organization of the cytoplasm. *Current Opinion in Structural Biology*, 12:679–684, 2002.
- [3] Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [4] S.S. Brandt and U. Ziese. Automatic tem image alignment by trifocal geometry. *Journal of Microscopy*, 222:1–14, 2006.
- [5] G. Cardone, K. Grunewald, and A.C. Steven. A resolution criterion for electron tomography based on cross-validation. *Journal of Structural Biology*, 151:117–129, 2005.
- [6] Gal Elidan, I McGraw, and Daphne Koller. Residual belief propagation: informed scheduling for asynchronous message passing. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, 2006.
- [7] AJ Koster, R Grimm, D Typke, R Hegerl, and A Stoschek. Perspectives of molecular and cellular electron tomography. *Journal of Structural Biology*, 120:276–308, 1997.
- [8] J.R. Kremer, D.N. Mastronarde, and J.R. McIntosh. Computer visualization of three-dimensional image data using imod. *Journal of Structural Biology*, 116:71:76, 1996.
- [9] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Constructing free energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51:2282–2313, 2005.