Structure and View Estimation for Tomographic Reconstruction: A Bayesian Approach

Satya P. Mallick spmallick@vision.ucsd.edu
Sameer Agarwal sagarwal@cs.ucsd.edu
David J. Kriegman kriegman@cs.ucsd.edu
Serge J. Belongie sjb@cs.ucsd.edu
Computer Science and Engineering , University of California, San Diego.

Bridget Carragher bcarr@scripps.edu
Clinton S. Potter cpotter@scripps.edu
National Resource for Automated Molecular Microscopy, and
Department of Cell Biology, The Scripps Research Institute, La Jolla.

Abstract

This paper addresses the problem of reconstructing the density of a scene from multiple projection images produced by modalities such as x-ray, electron microscopy, etc. where an image value is related to the integral of the scene density along a 3D line segment between a radiation source and a point on the image plane. While computed tomography (CT) addresses this problem when the absolute orientation of the image plane and radiation source directions are known, this paper addresses the problem when the orientations are unknown – it is akin to the structure-from-motion (SFM) problem when the extrinsic camera parameters are unknown. We study the problem within the context of reconstructing the density of protein macro-molecules in Cryogenic Electron Microscopy (cryo-EM), where images are very noisy and existing techniques use several thousands of images. In a non-degenerate configuration, the viewing planes corresponding to two projections, intersect in a line in 3D. Using the geometry of the imaging setup, it is possible to determine the projections of this 3D line on the two image planes. In turn, the problem can be formulated as a type of orthographic structure from motion from line correspondences where the line correspondences between two views are unreliable due to image noise. We formulate the task as the problem of denoising a correspondence matrix and present a Bayesian solution to it. Subsequently, the absolute orientation of each projection is determined followed by density reconstruction. We show results on cryo-EM images of proteins and compare our results to that of Electron Micrograph Analysis (EMAN)– a widely used reconstruction tool in cryo-EM.

1. Introduction

While the intensity in a photograph is related to the light (radiance) reflected from surfaces in a scene, the intensity at a point in an image produced by modalities such as x-ray, electron microscopy, etc. are related to the integral of the scene density along a 3D line segment between a radiation source and a point on the detector (image plane). Computed Tomography (CT) is a technique for reconstructing the 3D density from a collection of 2D images (aka projections) taken with a known relation between the radiation source/image plane and the scene. This is akin to 3D reconstruction from multiple photographs when the camera geometry is known (multi-view stereo).

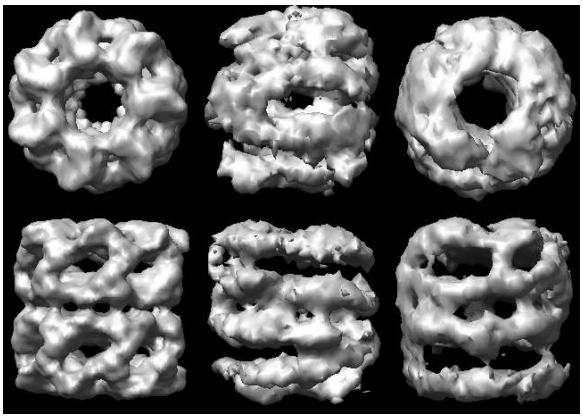

Figure 1. The first column shows the top and side views of a macro-molecule called GroEL produced from a 11.5 Å reconstruction [17] in a publicly available Molecular Structure Database. The middle column shows the initial model estimated using EMAN [16] – A widely used tool in cryo-EM. The right column shows the initial model estimated using our method. The same dataset was used to generate the two initial models.

In this paper, we consider the problem of 3D density reconstruction when the relations between the views are *unknown*. This is analogous to the problem of structure and motion estimation from photographs with unknown viewpoints. However the image formation process is different, and in turn this leads to different types of features

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

and constraints than traditionally encountered in SFM problems. Furthermore, we consider this problem within the context of cryo-EM reconstruction of macro-molecules, and at this resolution, images are very noisy compared to photographs typically used for SFM. Consider the images in Fig. 2.(a,b) of a protein macro-molecule. Unlike SFM from photographs, it is clearly not possible to identify points across these images that correspond to the projection of a common point in 3D, nor is it possible to extract out of images more complex features (e.g., lines, conics or other curves) and establish correspondence between them. Though not obvious a priori, it is possible to determine between every pair of 2D images a single line in each image which is the projection of a 3D line [3]. Hence, the essential challenge is both to identify these pairs of lines in images and to use these lines to estimate the absolute 3D orientations of the image planes. Once estimated, computed tomography is used for 3D density reconstruction. In addition, to achieve a desired resolution (< 10Å) in spite of the noise, researchers use between 1,000 and 100,000 projections, about two orders of magnitude more images than typically used in conventional SFM problems.

Cryo-Electron Microscopy (cryo-EM) is an emerging technique in structural biology for 3D structure (density) estimation of a specimen preserved in vitreous ice. Unlike tomography where a large number of images of a specimen can be acquired, the number of images of a specimen in cryo-EM is limited because of radiation damage. In cryo-EM, the specimen consists of identical copies of the same protein macro-molecule, preserved at random and unknown 3D orientations in ice. Due to larger number of unknowns in cryo-EM as compared to tomography, the problem is more challenging and calls for a different set of techniques.

One of the advantages of cyro-EM over the more widely used technique of X-ray crystallography is that it determines the 3D structure without the need for crystallization. It is often very difficult to crystallize large molecules (Biologists may spend many years trying to do this). Even in the cases when crystallization is possible, the structure constrained in crystalline form can be different from the structure of the macro-molecule in its native environment. Cryo-EM therefore presents an attractive alternative for structure estimation from a biological point of view.

Within the cryo-EM community, a set of techniques for solving the reconstruction problem have emerged [8], and implementations are available [9, 16, 21]. The process is essentially the following: First, a rough, usually low resolution and possibly distorted initial density (initial model) is constructed by some means (e.g., low resolution, higher dose electron micrographs, x-ray crystallography, single axis or random conical tomography, known structure of related molecules, assumed structure from other means, etc.). This model is used to initiate an iterative process where the image plane orientations relative to the current 3D model are determined (pose estimation), and then the 3D density (a new model) is reconstructed using CT techniques. The process repeats with this new model. It should be noted that each iteration may take 12 hours to run, and a full reconstruction may take a few weeks. In the end, the ability of the iterative process to converge to the correct solution depends critically on the accuracy of the initial model, and when it does converge, the number of required iterations also depends upon the accuracy of the initial model.

In this paper we address the problem of generating an initial model of the 3D structure using randomly oriented projections. In the following sections we will show that this is an instance of orthographic structure from motion using line correspondences. More specifically, the problem can be stated as:

Problem Statement: *Consider a set of N planes in $\mathbb{R}^3$ passing through the origin and with unknown orientations $\mathrm{R}_i, i = 1, 2, \cdots N$. A common line $\mathbf{c}_{ij} (= \mathbf{c}_{ji})$ is defined as the line of intersection between two such planes i and j. Since the planes pass through the origin, the orientation of the common line $\mathbf{c}_{ij}$ in the plane i is parameterized by the angle ϕ_{ij} it makes with the x-axis in the local coordinate frame. Given a matrix $\Phi = [\phi_{ij}], i, j = 1, 2, \cdots N$ with noisy entries , the objective is to recover the common lines $\mathbf{c}_{ij}$ and the rotation matrices R_i.*

Our paper makes the following three contributions.

1. It introduces a new, large scale structure from motion problem in which direct correspondence of image features is not possible.

2. It provides the solution to a problem that has plagued and possibly limited the applicability of cryo-EM for reconstructing macro-molecule density for structural biology.

3. We introduce to the computer vision community an application domain (cryo-EM) that can benefit from existing vision techniques, but which also challenge us with a new but relevant set of vision problems that have broader applicability beyond this specific domain.

2. Background and Related Work

In a typical cryo-EM imaging setup, several randomly oriented macro-molecules (aka particles) of the same kind frozen in ice and suspended over holes in a carbon film are placed under an electron microscope and their projections recorded onto a CCD. The projection is orthographic and the intensity at a pixel in the micrograph is directly proportional to the density in the path of the electron(s) that contribute to the intensity at a particular pixel. A typical cryo-

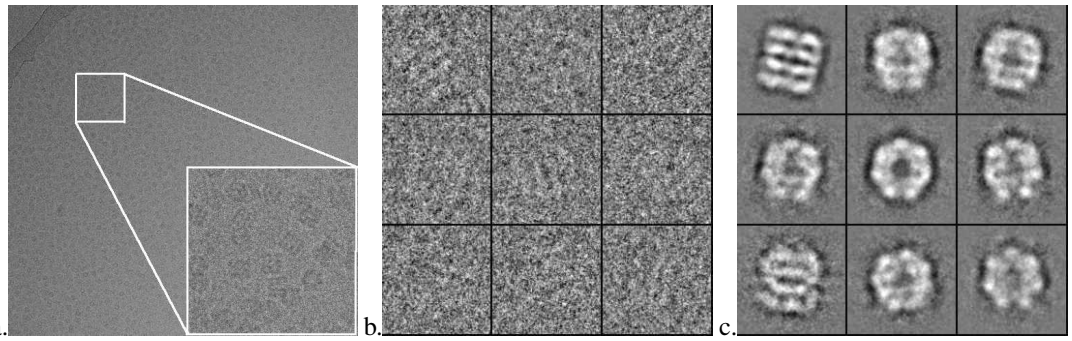

Figure 2. In (a) a typical cryo-EM micrograph containing several images of a macro-molecule called GroEL is shown. The inset shows a zoomed portion of the micrograph. (b) shows nine projections selected from a micrograph. Many such projections (≈ 10000) are clustered into $\approx 50 - 100$ classes. (c) shows the class averages of nine arbitrarily chosen classes. The class averages have significantly better signal to noise ratio at the expense of finer details (high resolution information) contained in raw projections.

EM micrograph is shown in Fig. 2 (a). A single micrograph contains noisy projections of several identical particles oriented randomly. The individual particles are selected and cropped from the micrograph. As can be seen in Fig. 2 (b), individual projections are extremely noisy. The signal to noise ratio can be improved by clustering a large number ($\sim 10,000$) of projections into a few classes ($\sim 50 - 100$) and averaging within each class; see Fig. 2 (c). Averaging within a class leads to smoothing of high resolution information contained in the projections. However the detail in class averages is sufficient for the purpose of reconstructing an initial model at a resolution of about $30 - 40$Å.

The different approaches for initial model reconstruction can be broadly classified on the basis of the imaging geometry used. In the untilted configuration, the carbon film is placed orthogonal to the direction of the electron beam and a single image of the specimen is obtained. On the other hand, in the tilted configuration, several images of the specimen are acquired by rotating the stage supporting the carbon film about a known axis by known angular increments. In this paper we focus our attention to the reconstruction of 3D density using a single exposure at zero tilt [5, 6, 10, 15, 19, 20, 22].

In modalities like electron microscopy where the obtained image is an integral of a 3D density along a particular direction, there arises a line correspondence between a pair of views; see Fig. 3 for an illustration. Typically, the planes of projection corresponding to two images i and j intersect in a line $c_{ij} (= c_{ji})$ called the *common line*; see Fig. 3 (a). The entire density can be projected onto the common line by integrating the intensities of either image in a direction orthogonal to the orientation of the common line. In other words, if ϕ_{ij} and ϕ_{ji} are the orientations of the common line in the local coordinate system of image i and j respectively, then

$$r_i(\phi_{ij}) = r_j(\phi_{ji}), \tag{1}$$

where r_i is the Radon transform of image i. We call this constraint the *common lines constraint*. Fig. 3 (b) shows a graphical illustration of the common lines constraint. It suggests that we can obtain the orientation of the common line in each image by performing a brute force comparison of their Radon transforms at all orientations. Unfortunately these estimates are very noisy because the error surface obtained by matching two Radon transforms typical contains several minima; see for example Fig. 3 (c).

Given noisy estimates of the common lines between N projections, the central problem is to find the relative orientation of these projections in 3-space. N projections of a density with known orientations can be assembled to obtain the 3D density using the Fourier Slice Theorem.

Theorem 1 (Fourier Slice Theorem)**.** *The 2D Fourier transform of a projection of a 3D density is a central slice through the 3D Fourier transform of the density. The orientation of the central slice is the same as the orientation of the plane of projection.*

The case when $N = 3$ is well studied and is also the minimal problem in terms of the number of images required. It was shown independently by Vainshtein and Goncharov [20] and Van Heel [22] (and later by Lauren and Nandhakumar [14]) that the relative orientation of three projections can be estimated up to a hand (chirality) ambiguity by using the common lines between the three projections. This method is called *Angular Reconstitution*.

Inspired by the work of Horn[12], Farrow and Ottensmeyer [5] used quaternions to obtain the relative orientation of a new projection in a least square sense. One of the criticisms of such an approach is that the solution is biased by the sequence in which the relative orientation of different projections are obtained. For example, if the common lines between the first three projections are noisy, the noise is propagated to the orientation estimates of all subsequent projections.

In [19] Penczek et al. try to obtain rotations corresponding to each projection simultaneously by minimizing an energy functional. Unfortunately there is no good way to minimize the functional except for a brute force search over all

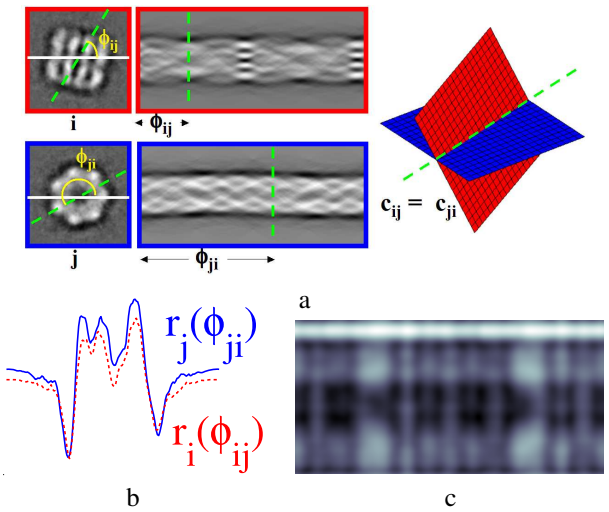

Figure 3. (a) Two projections i and j of a density (Left) with their Radon transforms r_i and r_j (Center), and their viewing planes (Right) are shown. The line of intersection of the viewing planes is called the common line and is shown using a dashed line. The common line is oriented at angles ϕ_{ij} and ϕ_{ji} in the local co-ordinate system of projection i and j respectively. (b) shows the Radon transforms $r_i(\phi_{ij})$ and $r_j(\phi_{ji})$ which match closely because of the common lines constraint (Eq. 1). The matching error surface $\mathcal{E}(\alpha, \beta) = \|r_i(\alpha) - r_j(\beta)\|$, $0 \leq \alpha < \pi$ and $0 \leq \beta < 2\pi$, between the Radon transforms of the two images is shown in (c). Since $\mathcal{E}(\alpha, \beta)$ has multiple minima, the estimate of the common lines is very noisy.

possible orientations for all projections. It is only expected to work well when the initial point is in the neighborhood of the optimal solution.

A similar problem has been studied in the field of tomography when the rotation corresponding to a projection is not known [14, 15]. In [15], the authors dealt with imaging noise using self consistency between four projections. Given four projections, it is possible to predict the location of common lines on one projection based on the other three projections.

3. Angular Reconstitution Revisited

In contrast to the traditional SFM problems, the minimal problem in uncalibrated tomography involves three views (projections). Given three projections of a 3D density, the viewing plane of each projection can be recovered up to a global rotation and chirality. Published derivations [20, 22] for uncalibrated three view tomography involve complex solid geometry which makes further analysis difficult. We present a novel derivation using simple linear algebra and vector calculus that enables a characterization of the necessary and sufficient conditions for a non-degenerate solution.

In this derivation, the camera is assumed to be orthographic and the projections are assumed to be centered. Under the above assumptions, the pose of the camera associated with projection i is fully specified by a rotation matrix R_i. The line along which the viewing planes of images i and j intersect is denoted by $\mathbf{c}_{ij}$. Note that $\mathbf{c}_{ij} = \mathbf{c}_{ji}$ and we will use them interchangeably as the need arises. Let ϕ_{ij} denote the angle made by the vector $\mathbf{c}_{ij}$ with the local x-axis in image i. We refer to the matrix of these angles, $\Phi = [\phi_{ij}]$ as the *common lines matrix*. The direction of common line $\mathbf{c}_{ij}$ in the local co-ordinate of projection i is denoted by $\mathbf{b}_{ij} = [\cos\phi_{ij}, \sin\phi_{ij}, 0]^\top$.

Given two projections i and j, the angles ϕ_{ij} and ϕ_{ji} are determined by exhaustively comparing the Radon transforms (r_i and r_j) and identifying those orientations for which they match the best.

$$(\phi_{ij}, \phi_{ji}) = \arg \min_{\substack{0 \leq \alpha < \pi \\ 0 \leq \beta < 2\pi}} \|r_i(\alpha) - r_j(\beta)\|, \qquad i < j \quad (2)$$

Note that both (ϕ_{ij}, ϕ_{ji}) and $(\phi_{ij} + \pi, \phi_{ji} + \pi)$ are valid solutions of (2). Restricting the maximum value of α to π ensures a unique solution.
The vector $\mathbf{c}_{ij}$ and its projection $\mathbf{b}_{ij}$ are related by

$$\mathrm{R}_i \mathbf{c}_{ij} = \mathbf{b}_{ij} \quad (3)$$

Considering the inner product of $\mathbf{b}_{ij}$ and $\mathbf{b}_{ik}$ we obtain

$$\begin{aligned} \langle \mathrm{R}_i \mathbf{c}_{ij}, \mathrm{R}_i \mathbf{c}_{ik} \rangle &= \langle \mathbf{b}_{ij}, \mathbf{b}_{ik} \rangle \\ \langle \mathbf{c}_{ij}, \mathbf{c}_{ik} \rangle &= \langle \mathbf{b}_{ij}, \mathbf{b}_{ik} \rangle \end{aligned}$$

Using the fact that $\langle \mathbf{b}_{ij}, \mathbf{b}_{ik} \rangle = \cos(\phi_{ij} - \phi_{ik})$, we obtain the fundamental geometric constraint

$$\langle \mathbf{c}_{ij}, \mathbf{c}_{ik} \rangle = \cos(\phi_{ij} - \phi_{ik}) \quad (4)$$

The above constraint is essentially a restatement of the simple fact that the angle between two vectors is preserved under a rigid transformation.

In the three view case, three common lines $\mathbf{c}_{12}$, $\mathbf{c}_{23}$, and $\mathbf{c}_{31}$ are shared between the projections. Let us define matrices $\mathrm{C} = [\mathbf{c}_{12}, \mathbf{c}_{23}, \mathbf{c}_{31}]^\top$, and $\mathrm{M} = \mathrm{C}\mathrm{C}^\top$. Then using Eq (4), we obtain the following relation

$$\mathrm{M} = \begin{bmatrix} 1 & \cos(\phi_{31} - \phi_{32}) & \cos(\phi_{23} - \phi_{21}) \\ \cos(\phi_{31} - \phi_{32}) & 1 & \cos(\phi_{12} - \phi_{13}) \\ \cos(\phi_{23} - \phi_{21}) & \cos(\phi_{12} - \phi_{13}) & 1 \end{bmatrix}.$$

Note that if C is a rank 3 matrix, i.e. the three common lines are not co-planar, then the matrix M is symmetric positive definite with unit diagonal entries. Given a matrix M of this kind with eigenvalue decomposition $\mathrm{M} = \mathrm{U}\mathrm{D}\mathrm{U}^\top$, we can determine C up to a rotation and reflection as

$$\mathrm{C} = \mathrm{U}\mathrm{D}^{1/2}$$

The solution is ambiguous up to a rotation because a global rotation preserves the angles between the common lines. The reflection ambiguity is a result of the fact that the pair

of entries M_{ij} and M_{ji} are only known up to a sign. This is so because the intersection of viewing planes of projections i and j can be represented equally well by vectors $\mathbf{c}_{ij}$ and $-\mathbf{c}_{ij}$. This ambiguity is reflected in the matrix Φ by the fact that we can change the entries ϕ_{ij} and ϕ_{ji} by π without changing the common lines. This ambiguity is not just a mathematical artifact, it manifests in nature in the form of chirality and the phenomenon of optical isomerism. Two molecules that are reflections of each other give rise to the same set of projections, and while performing a reconstruction, a choice of either a left handed or a right handed co-ordinate system must be made to recover a unique solution.

The above analysis is valid when the matrix M is positive definite. Is it possible to obtain a full rank matrix C in a least squares sense even when the matrix M is not positive definite? This problem is equivalent to finding the closest symmetric positive semidefinite matrix $\hat{M}$ with unit diagonal. The resulting matrix can then be exactly factorized to estimate C. However, the following theorem by Higham [11] shows that such an attempt is not useful as $\hat{M}$ is always rank deficient.

Theorem 2 (Higham)**.** *If a symmetric matrix M with unit diagonal has k non-positive eigenvalues, then the nearest positive semidefinite matrix to M with unit diagonal has at least k zero eigenvalues.*

Thus if M is not positive definite, then it is either rank deficient or the closest positive semidefinite matrix to it is. In either case this would result in a matrix C that is rank deficient, i.e. the three common lines lie in the same plane. Co-planar common lines is a degenerate configuration for tomographic reconstruction.

3.1. Rotation Estimation

We now consider the problem of estimating the rotation matrix that relates a set of common lines with their projections. Let $C_i = [\mathbf{c}_{ik}]$ be a matrix with columns consisting of common lines formed by view i with other viewing planes and B_i be the matrix formed by collecting the corresponding image of the common lines in projection i as column vectors; then from Eq. (3) we have the relation

$$R_i C_i = B_i \tag{5}$$

For matrices C_i and B_i with 3 or more columns we have an over-constrained problem, one which can be solved in the least squares sense as

$$\min_{R_i} \|R_i C_i - B_i\|_F \qquad \text{subject to } R_i^\top R_i = 1 \tag{6}$$

The above is a well studied problem in computer vision and linear algebra [1, 13]. We use the solution proposed by Arun et al. [1] Let

$$U D V^\top = B_i C_i^\top$$

be the singular value decomposition of $B_i C_i^\top$, then Arun et al have shown that the optimal solution to the optimization problem above is given by

$$R_i = U V^\top$$

In the case of three projections, the matrices C_i and B_i are both 3×2 and their outer product is rank deficient and generally of rank 2. Thus while the first two pairs of singular vectors of $B_i C_i^\top$ are well defined, the two vectors spanning the left and right null spaces have a sign ambiguity. This results in two estimates of R_i,

$$R_i^+ = U V^\top, \qquad R_i^- = U I^- V^\top, \tag{7}$$

Here, I^- is an identity matrix with its third diagonal entry set to -1. The ambiguity between these two solutions can be resolved by observing that rotation matrices have a positive determinant of 1.

4. Robust Rotation Estimation

The matrix Φ contains the orientation of the common lines in the local co-ordinate system of images and its entries are directly measured using Eq. 2. If the entries of the matrix Φ are relatively noise free, a simple greedy strategy of starting with a triplet of views, and adding one view of the molecule at a time, using the least squares solution explained in sub-section 3.1, would be sufficient for generating an initial model. However, as shown in Fig. 3 (c), the entries of the matrix Φ often contain gross errors. This makes the use of a greedy least squares based strategy unsuitable for producing an initial model. In this section we present a Bayesian Maximum A Posteriori (MAP) estimation procedure that denoises the common lines matrix, which can then be used for estimating the initial model. Before we discuss the details of the denoising algorithm, we illustrate key ideas of our approach using a simpler problem.

4.1. Denoising Pairwise Distances

Consider the classical Multidimensional Scaling (MDS) [2] problem of embedding m points in a plane given pairwise distance matrix $S = [s_{ij}]$. As classical MDS is based on minimizing the L_2 norm, a single outlier can result in an arbitrarily bad embedding. Therefore, it would be useful to denoise the distance matrix S to correct for gross errors before MDS is performed.

The key idea in the denoising process is that of consistency. To denoise the distance between points i and j we choose a triple of points (p, q, r) from the remaining points. Unless the pairwise distances between the points p, q, and r do not obey the triangle inequality, they can be used to define a triangle in a plane. This triangle provides a fixed coordinate system in which we can embed the points i and

j using their distances from points p, q, and r. Notice that the points i and j are embedded in the plane defined by the triple (p, q, r) without using the distance between i and j. The coordinates of the points i and j in the embedding can be used to measure the distance between them. This is an estimate of the distance between i and j via the triplet. This process can be repeated over various choices of the triplets sampled from ${}^{m-2}C_3$ possibilities, and the resulting distance estimates collected. If the corruption in S is random and it does not have a systematic bias, one expects that histogram of these distances will exhibit a peak near the correct value of s_{ij} and the errors would be distributed randomly. Of course, one cannot rule out a systematic corruption in the matrix, and one cannot just trust the peak in the histogram; some weight should be given to the original estimate of s_{ij}. These ideas can be formalized in a Maximum A Posteriori (MAP) framework as follows.

The Maximum A Posteriori estimate of a random variable is the value for which its posterior given some observation achieves the maximum value. More formally, let us assume that we are interested in estimating a parameter θ and let $p_0(\theta)$ be a prior distribution on the domain of θ. If we now observe a random variable X that has conditional probability density $p(X|\theta)$, then by Bayes' rule, the probability of observing θ given X is written as

$$p(\theta|X) = \frac{p(X|\theta)p_0(\theta)}{p(X)},$$

where $p(\theta|X)$ is the posterior of θ. The Maximum A Posteriori (MAP) estimate of θ is then written as

$$\begin{aligned}\theta^* = \arg\max_\theta p(\theta|X) &= \arg\max_\theta \frac{p(X|\theta)p_0(\theta)}{p(X)} \\ &= \arg\max_\theta p(X|\theta)p_0(\theta) \qquad (8)\end{aligned}$$

In the context of the distance estimation problem, $p_0(\theta)$ is a distribution centered around the initial estimate s_{ij} and $p(X|\theta)$ is the empirical density of the observed distances. As we know very little about the behavior of the observed data density, but we have plenty of data available as triplets are sampled from ${}^{m-2}C_3$ possibilities, we will use a kernel density estimate to model $p(X|\theta)$. In particular, if a Gaussian distribution is used for the prior as well as the observations, then the MAP estimator is given by

$$s_{ij}^{MAP} = \arg\min_x e^{-\|s_{ij}-x\|^2/\sigma_p^2} \sum_k e^{-\|x_k-x\|^2/\sigma_o^2}. \qquad (9)$$

Here σ_p and σ_o are the standard deviation of the kernel for the prior and the observations respectively and express our confidence in each of these quantities.

4.2. Denoising the Common Lines Matrix

We now consider the problem of denoising the common lines matrix Φ. As in the case of denoising the distance matrix of points in a plane, we use geometric consistency to indirectly obtain multiple estimates for each entry of Φ in addition to the noisy direct measurement.

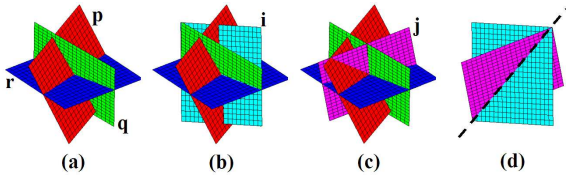

Figure 4. Indirect estimate of common line c_{ij}. (a) Three projections p, q and r are first assembled in three space using Angular Reconstitution. (b) The common lines between projection i and triplet (p, q, r) are used to find the orientation of projection i in the co-ordinate system defined by (p, q, r). (c) Similarly, the orientation of projection j is found. (d) The common line between projections i and j estimated via the triplet (p, q, r).

Fig. 4 shows indirect estimation of the common line c_{ij} via three projections p, q, and r. A triplet (p, q, r) of views establishes a coordinate system in which we can use the common line angles between the views (p, q, r) and the views i and j to estimate the rotations R_i and R_j in space. The method for this was described in sub-section 3.1. Given the two rotations, the common line between the two views is the cross product of their third rows.

$$\mathbf{c}_{ij} = \mathrm{R}_i^3 \times \mathrm{R}_j^3$$

This common line can then be back projected to images i and j using Eq. (3) to obtain estimates of ϕ_{ij} and ϕ_{ji}. Figure 5 shows typical histograms corresponding to entries of Φ. Notice the sharp peaks which allow us to estimate the common line orientations robustly.

There are three major differences between the problem of denoising a distance matrix and the problem of denoising a common lines matrix. First we have to estimate two scalar parameters (ϕ_{ij}, ϕ_{ji}) simultaneously as compared to a single distance $(s_{ij} = s_{ji})$. Further, to ensure uniqueness in common line matching, we have to enforce the constraints $0 \leq \phi_{ij} < \pi$ and $0 \leq \phi_{ji} < 2\pi$ for $i < j$. While this representation enforces uniqueness, it destroys the topology of the manifold of common lines angles. For example, the common line with angle $\phi_{ij} = 0 + \epsilon$ and the common line with angle $\phi_{ij} = \pi - \epsilon$ are very close in space and yet they are mapped far apart in this representation. This is a standard issue in analyzing axial data, i.e. data which is represented using a line segment and, as opposed to a vector, has ambiguous orientation [7, 18]. A standard trick is to map the data back onto the circle by doubling it. This maps the point π to 2π, thereby fixing the topological problem. Finally, now that the data lies on $S^1 \times S^1$, i.e. the product of two circles, which unlike the real line wraps around, and the kernel density function must take this into account. We use the analog of the Gaussian distribution for circular data, the von Mises distribution, which has the correct wraparound

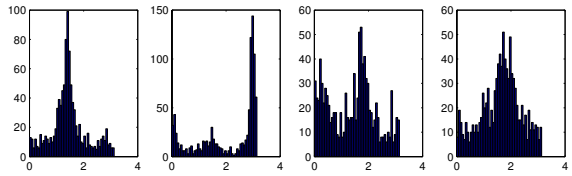

Figure 5. Typical histograms corresponding to entries of Φ obtained during the reconstruction of GroEL.

behavior for circles and spheres. In the univariate case it is written as

$$K(\phi) = \kappa(\sigma) \exp(\cos(\phi_0 - \phi)/\sigma)$$

where, κ is the normalization constant. For the bi-variate case we use a product of two univariate kernels.

While there is a substantial literature on finding the maximum of a probability density from samples [4] which work well in $\mathbb{R}^n$, generalizations to more complex manifolds like a product of circles are not known to the best of our knowledge. Thus we evaluate the density at all points in our sample and select the point at which the maximum is attained.

5. Results

Experiments on real data were performed on a protein macro-molecule called GroEL and results were compared with a widely used cyro-EM reconstruction tool called Electron Micrograph Analysis (EMAN) [16]. 15839 projections of GroEL were clustered into 40 classes and the corresponding 40 class averages were generated. A few examples of the class averages are shown in Fig. 2 (c). A comparison between the initial model obtained using our method and the initial model obtained using EMAN is shown in Fig. 1. Our initial model clearly captures the gross structure of published high resolution structure [17], and appears to be better than EMAN's initial model. The two initial models were refined using standard refinement routines in EMAN using all 15839 projections. Fig. 6 shows the progress of the refinement stage. Our initial model refined to a reasonable model by the end of two iterations (18 hours of compute time), while the convergence of EMAN's initial model is much slower. It is worth noting that with an unreliable initial model, not only is the rate of convergence slow, the probability of the initial model converging to the true solution is very low.

The last iteration shown in Fig. 6 (b) and 6 (d) also illustrate the hand ambiguity inherent in the solution. A close look at the two solutions shows that one is the mirror reflection of the other. As mentioned earlier, the hand ambiguity is resolved by other means after the reconstruction.

6. Discussion

In this paper we considered the initial model problem in uncalibrated computed tomography. We proposed a Bayesian solution to the problem which improves the practical applicability of a theoretical result known for two decades. In addition, we presented a novel and much simpler algebraic derivation and analysis of uncalibrated three view tomography.

An interesting aspect of our MAP formulation is that it allows for the use of more sophisticated priors on the estimates of the common lines than the one we have used in the paper. In particular it allows us to use the entire matching error surface of the Radon transform of two projections as the basis of a prior. In addition, we have not addressed the question of optimal kernel bandwidth for the MAP estimator. We hope to address these issues in future work.

7. Acknowledgments

The authors would like to thank Scott Stagg for providing the GroEL dataset and help with software package EMAN. Part of this work was conducted at the National Resource for Automated Molecular Microscopy which is supported by the National Institutes of Health through the National Center for Research Resources' P41 program (RR17573). David Kriegman and Satya Mallick were supported under grant NSF EIA-03-03622, Sameer Agarwal was supported under NSF CCF-04-26858, and Serge Belongie was supported under NSF CAREER 0448615, the Alfred P. Sloan Research Fellowship, and the Department of Energy under contract No. W-7405-ENG-48.

References

[1] K. Arun, T. Huang, and S. Bolstein. Least-Squares Fitting of Two 3-D Point Sets. *Pattern Analysis and Machine Intelligence*, 9:698–700, 1987.

[2] Ingwer Borg and Patrik Groenen. *Modern multidimensional scaling: Theory and Applications*. Springer Verlag, 1997.

[3] R. Bracewell. Strip Integration in Radioastronomy. *Australian J. of Phy.*, 9:198–205, 1956.

[4] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, 2002.

[5] M. Farrow and P. Ottensmeyer. *A Posteriori* Determination Of Relative Projection Directions Of Arbitrarily Oriented Macrmolecules. *JOSA-A*, 9(10):1749–1760, October 1992.

[6] M. Farrow and P. Ottensmeyer. Automatic 3D Alignment of Projection Images of Randomly Oriented Objects. *Ultramicroscopy*, 52:141–156, 1993.

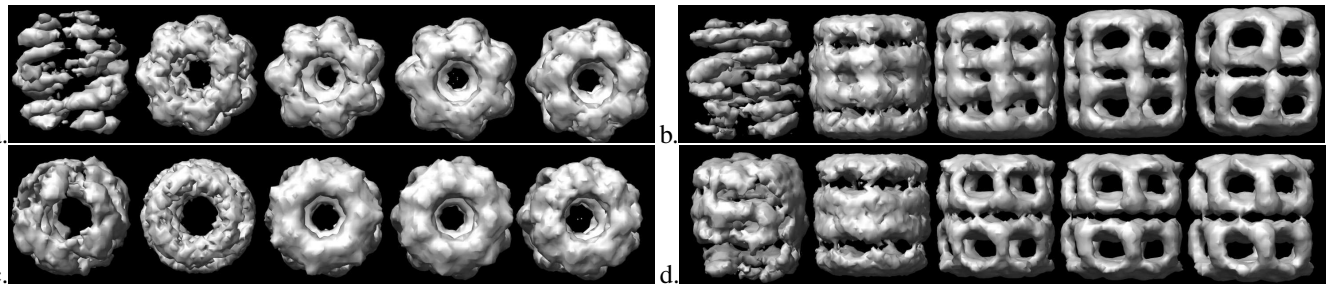

Figure 6. The figure shows a comparison of our method with EMAN – a widely used tool for cryo-EM reconstruction. The columns in the figure denote different stages of refinement starting with the initial model shown in the first column. (a) and (b) show the refinement of the top and side views of the initial model obtained using EMAN while (c) and (d) show the refinement of the top and side views of the initial model obtained using our method. The refinement is done using routines available in EMAN. The ground truth is shown in Fig. 1. Using the initial model obtained by EMAN, the refinement procedure starts to converge after the fourth iteration. In contrast, the gross shape of the GroEL is already visible in the second refinement iteration on our initial model.

[7] N. Fisher. *Statistical analysis of circular data*. Cam. Univ. Press, 1993.

[8] J. Frank. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press, 2006.

[9] J. Frank, M. Radermacher, P. Penczek, J. Zhu, et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields . *J. of Struct. Bio.*, 116:190–199, 1996.

[10] A. Goncharov and M. Gelfand. Determination of Mutual Orientation of Identical Particles from Their Projections by the Moments Method. *Ultramicroscopy*, 25:317–328, 1988.

[11] Nicholas J. Higham. Computing the Nearest Correlation Matrix a Problem From Finance. *IMA Journal of Nummerical Analysis*, pages 329–343, 2002.

[12] B. Horn. Closed-form solution of absolute orientation using unit quaternions. *JOSA-A*, 4:629–642, 1987.

[13] B. Horn, H. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A*, 5(7), 1988.

[14] P. Lauren and N. Nandhakumar. Computing the view orientations of random projections of asymmetricobjects. *CVPR*, pages 71–76, Jun 1992.

[15] P. Lauren and N. Nandhakumar. Estimating the Viewing Parameters of Random, Noisy Projections of assymetric objects for Tomorgraphic Reconstruction. *PAMI*, 19(5), May 1997.

[16] S. Ludtke, P. Baldwin, and W. Chiu. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struc. Bio.*, 122:82–97, 1999.

[17] S. Ludtke, J. Jakana, J. Song, D. Chuang, and W. Chiu. An 11.5 A Single Particle Reconstruction of GroEL Using EMAN. *J. of Mol. Bio.*, 314:241–250, 2001.

[18] K. Mardia, J. dKent, and J. Bibby. *Multivariate Analysis*. Acad. Press, 2000.

[19] P. Penczek, J. Zhu, and J. Frank. A Common-lines Based Method for Determining Orientations for N>3 Particle Projections Simultaneously. *Ultramicroscopy*, 63:205–218, 1996.

[20] B. Vainshtein and A. Goncharov. Determination of the spatial orientation of arbitrarily arranged identical particles of an unknown structure from their projections. *Proc. llth Intern. Congr. on Elec. Mirco.*, pages 459–460, 1986.

[21] M. Van Heel. Single-particle electron microscopy: towards atomic resolution. *Quart. Rev. of Bio. phy.*, 33:307–369, 2000.

[22] Marin van Heel. Angular Reconstitution: A Posteriori Assignment of Projection Directions for 3D Reconstruction. *Ultramicroscopy*, 21:111–124, 1987.