Rotationally Invariant Image Representation for Viewing Direction Classification in Cryo-EM

Zhizhen Zhao[a,*], Amit Singer[b]

[a]*Courant Institute of Mathematical Sciences, New York University, Warren Weaver Hall, 251 Mercer street, New York NY 10012 USA.*
[b]*Department of Mathematics and PACM, Princeton University, Fine Hall, Washington Road, Princeton NJ 08544-1000 USA.*

Abstract

We introduce a new rotationally invariant viewing angle classification method for identifying, among a large number of cryo-EM projection images, similar views without prior knowledge of the molecule. Our rotationally invariant features are based on the bispectrum. Each image is denoised and compressed using steerable principal component analysis (PCA) such that rotating an image is equivalent to phase shifting the expansion coefficients. Thus we are able to extend the theory of bispectrum of 1D periodic signals to 2D images. The randomized PCA algorithm is then used to efficiently reduce the dimensionality of the bispectrum coefficients, enabling fast computation of the similarity between any pair of images. The nearest neighbors provide an initial classification of similar viewing angles. In this way, rotational alignment is only performed for images with their nearest neighbors. The initial nearest neighbor classification and alignment are further improved by a new classification method called vector diffusion maps. Our pipeline for viewing angle classification and alignment is experimentally shown to be faster and more accurate than reference-free alignment with rotationally invariant K-means clustering, MSA/MRA 2D classification, and their modern approximations.

---

*Corresponding author. Address: Courant Institute of Mathematical Sciences, New York University, Warren Weaver Hall, 251 Mercer street, New York NY 10012 USA. Tel: +1 (443)-791-4306.

Email addresses: `jzhao@cims.nyu.edu` (Zhizhen Zhao), `amits@math.princeton.edu` (Amit Singer)

1. Introduction

Single particle reconstruction (SPR) from cryo-electron microscopy (EM) images is an entirely general technique for determining the 3D structures of macromolecular complexes [1, 2, 3, 4], which does not require crystallization or other special preparation of the complexes to be imaged. In cryo-EM, the functionally active macromolecular complexes are prepared in vitro, stalled by chemical means, and rapidly frozen by immersion into liquid ethane at liquid-nitrogen temperature. The randomly oriented and positioned macromolecular "particles", typically complexes 200 kDa or larger in size, are maintained at the liquid-nitrogen temperature throughout the image acquisition in the microscope. One of the challenges in SPR with cryo-EM images is the low signal to noise ratio (SNR), due to the lack of periodicity of the molecule frozen in thin vitreous ice layer.

Because of the low SNR, it is extremely hard to visualize individual particle. To improve the resolution, a crucial step is alignment and averaging of the 2D projection images, a procedure known as "class averaging". Images from the same projection angles should be identified, centered, rotationally aligned and averaged to achieve a higher SNR. Generating 2D class averages could be useful for common-lines based 3D *ab initio* reconstruction. They can also be used for direct observation to look for heterogeneity or discover symmetry as well as for separating particles into subgroups for additional analysis. Therefore, it is important to have fast and accurate algorithms for computing class averages.

There are two main approaches for generating 2D class averages. IMAGIC [5] uses multivariate statistical analysis (MSA) and multi-reference alignment (MRA) for 2D image classification. The MSA compresses and denoises large image data sets to achieve efficient classification using hierachical ascending classification method. The clustered images produce references for the MRA class averaging step. Since projection images can be similar up to rotation and small translations, several invariant features were proposed as a preprocessing step for viewing angle classification, for example, autocorrelation functions (ACF) and double autocorrelation function (DACF) [6]. SPIDER [7] uses reference-free alignment (RFA) [8] followed by rotationally invariant K-means clustering [9] for 2D class averaging. Reference-free alignment tries to globally align images. The optimization method aims at finding

2

alignment parameters of rotations and shifts for all images that minimize the sum of squared deviations from their mean (i.e., minimum variance).

Modern software packages for SPR also include procedures for 2D class averaging. EMAN2 [10] 2D class averaging method uses invariant features for initial classification. The calculation of invariants is a 2-stage process. It first computes the self correlation function (SCF) [11] of an image to make it translational invariant, which is followed by a polar transformation and a sequence of 1-D autocorrelations on each ring to generate rotatioanlly invariant SCF images. The invariants are only used to bootstrap the process and the classification after this point is MSA/MRA based. The procedure for 2D class averaging in Xmipp [12] is CL2D, which is based on the algorithm proposed by Sorzano et al [13]. Their algorithm for 2D multireference alignment and classification is based on a hierachical clustering approach using correntropy instead of the traditional correlation. Computing the correntropy between each image and the class reference gives classification results that are less sensitive to noise. They also proposed a new clustering criterion so as to avoid the situation that the cleaneast class "attracts" many experimental images even if they belong to some other classes. This modified criterion for the definition of the clusters was shown to be especially suited for images with low SNR. SPARX [14] uses a 2D class averaging method called iterative stable alignment and clustering (ISAC) [15], which relies on the concepts of stability and reproducibility of clusters. Relion [16] uses a Bayesian approach to infer parameters for a statistical model from the data. This method is used in both reference-free 2D class averaging and unsupervised 3D classification. The class averages can be deblurred and refined by using algorithms proposed in [17, 18, 19].

We notice that RFA produces significantly large errors when the images have many different views. The reason for this failure is mathematical: there does not exist an assignment of in-plane rotational angles that can align all images simultaneously. The underlying theorem is known as the hairy ball theorem, and we will elaborate on this issue in the following section. While global alignment is impossible, one can always determine the rotationally invariant distances between all pairs of images by optimally aligning each pair of them. In this way, we have to perform $\binom{n}{2}$ alignments for n images. This is computationally intensive and unnecessary, because most of the time is spent on aligning images from very different views. It would be more efficient to use a rotationally invariant representation for the images, then find neighboring images, and finally align and average only neighboring images.

3

We introduce a new rotationally invariant representation for computing the rotationally invariant distance between all pairs of cryo-EM images. Our invariant representation is based on expanding the images in a steerable basis and deriving a bispectrum for this expansion [20, 21]. Unlike ACF, DACF and SCF, the new rotationally invariant representation maintains phase information and is complete, in the sense of uniquely specifying the original image up to an arbitrary rotation. In signal and image processing, a wide variety of invariants were devised for pattern recognition [22]. A common feature of most invariants is that they are lossy, in the sense that they do not uniquely specify the original signal. Among invariant features, the bispectrum and the triple-correlation function provide a lossless shift-invariant representation, and various algorithms have been devised to retrieve a signal (up to translation) from its (possibly noisy) bispectrum [23]. We therefore find this representation useful in determining the rotationally invariant distances between any pair of images. Bispectrum and triple-correlation function have been considered before for generating translational or rotational invariant features for cryo-EM images [6, 24, 25]. However, because the number of such features is extremely large, it was regarded impractical for computations. We reduce the number of bispectrum-features in two steps. We first perform principal component analysis (PCA) for all the images and their in-plane rotations efficiently to produce a steerable basis, where the eigen-images are separable to angular Fourier modes and radial functions [21]. The projection images are expanded and compressed in the leading M steerable eigen-images. Different triplets of these expansion coefficients are multiplied together to produce the invariant image representation. The resulting invariant representation is still high-dimensional, consisting of $O(M^3/k_{max})$ features, where k_{max} is the maximum angular frequency. Marabini and Carazo [25] suggested projecting the bispectrum onto a lower dimensional subspace as a pattern classification method. However, their method consists of using a predetermined subset of bispectrum coefficients and does not preserve the information content well enough to discriminate images of many different views. Instead, in the second step, we reduce the dimensionality of the invariant feature vectors by PCA. We use a randomized algorithm for low rank matrix approximation [26, 27, 28] to efficiently compute the principal components, overcoming the difficulties imposed by the large number of images and the high dimensionality of the input feature vectors. The top principal components provide the reduced invariant image representation. We then efficiently compute the rotationally invariant distance between images as the Euclidean

distance between their reduced invariant representations without performing any in-plane alignment. A predetermined number of nearest neighbors for each image are identified as those images with the smallest invariant distances. For a large number of input images, a randomized nearest neighbor algorithm [29] can avoid computing the distances between all pairs of images and effectively find the nearest neighbors in time nearly linear with the number of images. Either ordinary or randomized nearest neighbor search with reduced invariant image representation gives the initial classification result. The rotational alignment angles are then computed only for nearest neighbor pairs. With the techniques we propose here, a substantial gain in computation time is obtained by reversing the order of alignment and classification.

The initial nearest neighbors classification can be improved by a clustering algorithm, such as K-means, that takes into account all pairwise distances between images within the neighborhood. But it is usually very difficult to get good clustering for a large number of clusters and the cluster size varies considerably. Nearest neighbor classification is a natural algorithmic framework for averaging an image with a predetermined number of similar images. The initial classification can be further improved by taking into account the consistency of in-plane rotations along multiple paths that connect neighboring images through their common neighbors. This classification method is called Vector Diffusion Maps (VDM) [30, 31].

This paper is organized as follows. In Section 2, we put forward two problems with the reference-free alignment and rotationally invariant K-means clustering. In Section 3, we present our algorithms for generating rotationally invariant image representations for the purpose of viewing angle classification. Also in that section, we show how to improve the initial nearest neighbor classification and rotational alignment using VDM. The nearest neighbor pairs and their relative alignment are used to generate class means. In Section 4, we detail the results of numerical experiments for simulated projection images of the 70S ribosome with the purpose of benchmarking the efficiency and accuracy of the algorithm. Our algorithm is shown to be more accurate than other existing 2D class averaging procedures and it is also faster. We conclude that section by detailing the results of our class averaging method for three experimental data sets of the 70S ribosome, 50S ribosomal subunit, and IP$_3$R1. Our class averaging method is available in the
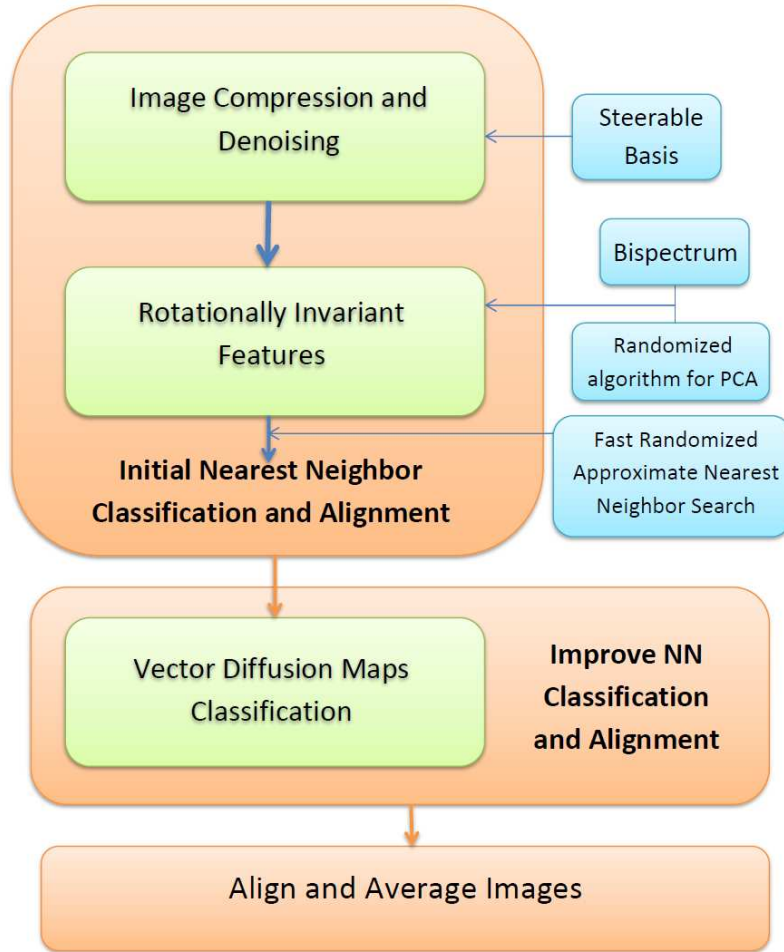
Figure 1: Schematic diagram of our class averaging procedure for single particle reconstruction.

SPR toolbox ASPIRE[1]. The toolbox includes three main functions written in MATLAB "Initial_classification.m", "VDM.m", and "align_main.m" that correspond to the three major components in the pipeline of our 2D class averaging method (see Figure 1).

2. Motivation

2.1. No global rotational alignment

To each projection image I there corresponds a 3×3 unknown rotation matrix R ($RR^T = R^T R = I_{3 \times 3}$ and $\det R = 1$), describing its orientation

$$R = \begin{pmatrix} | & | & | \\ R^1 & R^2 & R^3 \\ | & | & | \end{pmatrix}.$$

The projection image can be viewed as a tangent plane to the two dimensional unit sphere S^2 at the viewing direction $v = v(R) = R^3$. The first two columns of R, namely, R^1 and R^2, are vectors in $\mathbb{R}^3$ that form an orthonormal basis for the tangent plane and are identified with the coordinate axes of the image (see Figure 2). Together with the imaging direction v they make an orthonormal basis of $\mathbb{R}^3$. An in-plane rotation of the projection image can thus be viewed as changing the basis vectors R^1 and R^2 while keeping v fixed.
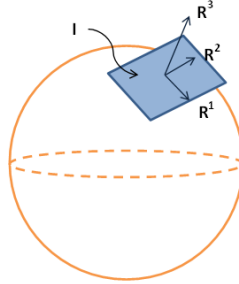

Figure 2: The image I is identified with the tangent plane to the sphere at the viewing direction R^3 which is the third column of the rotation matrix R.

The similarity of images can be measured by the Euclidean distance between the images when they are optimally aligned with respect to in-plane rotations (assuming the images are centered):

$$d_{ij} = \min_{\alpha \in [0,2\pi)} \|I_i - R(\alpha)I_j\|, \quad i,j = 1, ..., n, \tag{1}$$

where $R(\alpha)$ stands for rotating image I_j counter-clockwise by α. The optimal alignment angle is

$$\alpha_{ij} = \operatorname*{argmin}_{\alpha \in [0,2\pi)} \|I_i - R(\alpha)I_j\|, \quad i,j = 1, ..., n. \tag{2}$$

When two images I_i and I_j are of the same viewing angle ($v_i = v_j$), the matrix $R_i^{-1}R_j$ is of the form

$$R_i^{-1}R_j = \begin{pmatrix} \cos\alpha_{ij} & -\sin\alpha_{ij} & 0 \\ \sin\alpha_{ij} & \cos\alpha_{ij} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

given by $\cos(\alpha_{ij}) = (R_i^{-1}R_j)_{11}$ and $\sin(\alpha_{ij}) = (R_i^{-1}R_j)_{21}$. In practice, however, we cannot expect two projection images to have exactly the same viewing angle.

For clean images, it is expected that a small discrepancy between v_i and v_j would imply that α_{ij}, obtained from optimal rotational alignment, approximates the angle $\tilde{\alpha}_{ij}$ given by

$$\tilde{\alpha}_{ij} = \operatorname*{argmin}_{\alpha \in [0,2\pi)} \|\rho(\alpha) - R_i^{-1}R_j\|_F^2, \tag{3}$$

where

$$\rho(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and $\|A\|_F^2 = \operatorname{Tr}(AA^T)$ for any real valued $m \times n$ matrix A (i.e., it is the squared Frobenius norm). It can be verified that $\tilde{\alpha}_{ij}$ satisfies [30]

$$\cos(\tilde{\alpha}_{ij}) = \frac{(R_i^{-1}R_j)_{11} + (R_i^{-1}R_j)_{22}}{\sqrt{[(R_i^{-1}R_j)_{11} + (R_i^{-1}R_j)_{22}]^2 + [(R_i^{-1}R_j)_{21} - (R_i^{-1}R_j)_{12}]^2}}, \tag{4}$$

$$\sin(\tilde{\alpha}_{ij}) = \frac{(R_i^{-1}R_j)_{21} - (R_i^{-1}R_j)_{12}}{\sqrt{[(R_i^{-1}R_j)_{11} + (R_i^{-1}R_j)_{22}]^2 + [(R_i^{-1}R_j)_{21} - (R_i^{-1}R_j)_{12}]^2}}. \tag{5}$$

During our simulations, the true relative in-plane rotation is defined through equations (4) and (5).

Penczek *et al.* [8] introduced reference-free alignment that first globally aligns all the images and then the rotationally invariant distance is the Euclidean distance between the pre-aligned images. What we are about to elucidate is that such global alignment does not exist when there is a great variety of viewing angles. In such cases, the estimation of the in-plane rotations between images from similar views by RFA is not accurate.

We used a data set composed of clean images corresponding to many different views in order to numerically test the performance of RFA algorithm [8] for viewing angle classification and for rotational alignment of in-class images. Specifically, 10^4 centered clean projection images were simulated from the 3D model of *E.Coli* 70S ribosome with viewing directions that are sampled from the uniform distribution over the sphere. We used SPIDER AP RA program to run RFA on different subsets of the simulated data to test the rotational alignment results. Since we know the underlying rotations, we can compute $\tilde{\alpha}_{ij}$ for pairs of images that satisfy $\langle v_i, v_j \rangle \geq \cos(5°)$, that is, for viewing angles that are less than 5° apart. This list of true in-plane rotational angles are compared with the estimation from the reference free alignment. Firstly we ran RFA on the whole data set whose viewing directions are uniformly distributed over the sphere. The algorithm produces large errors when all views are included (see Figure 3a). As we decrease the size of the spherical cap to 80° , 60° and 40°, the errors in in-plane rotational alignment become smaller (see Figure 3).

The (perhaps surprising) failure of RFA to globally align all images is a consequence of a mathematical theorem called the hairy ball theorem [32]. The theorem says that a continuous tangent vector field to the two dimensional sphere S^2 must vanish at some point on the sphere. In other words, if f is a continuous function that assigns a vector in $\mathbb{R}^3$ to every point v on the sphere such that $f(v)$ is tangent to the sphere at v, then there is at least one $v \in S^2$ such that $f(v) = \mathbf{0}$. The theorem attests to the fact that it is impossible to comb a hairy (spherical) cat without creating a cowlick. The hairy ball theorem implies that any attempt to find a non-vanishing continuous tangent vector field to the sphere would ultimately fail. A successful global rotational alignment of all projection images means that we can choose orthogonal bases to all tangent planes such that the basis vectors vary smoothly from one tangent plane to the other. However, this is a contradiction to the hairy ball theorem.

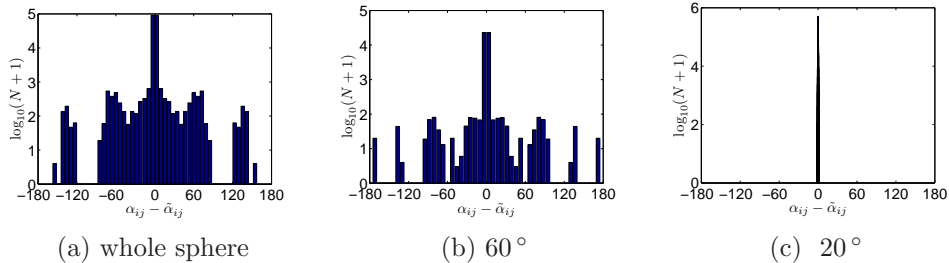| (a) whole sphere | (b) 60° | (c) 20° |

Figure 3: Error in degrees of in-plane rotational alignment between images with similar viewing angles that are less than five degrees apart for simulated clean projection images of the 70S ribosome, with viewing angles belonging to spherical caps of various opening angles (whole sphere, 60 degrees, 20 degrees). The y axis is in log scale, because the number of outliers is small. The fraction of pairs for which the error is larger than 2 degrees is $p_a = 0.13$, $p_b = 0.09$, and $p_c = 0$.

This implies that any classification algorithm that first attempts to globally align the images, such as K-means clustering after RFA, would ultimately fail whenever there are many different views that cover the sphere. We refer the reader to Appendix B of [30] for a discussion about the relevance of the hairy ball theorem in the discrete case of a finite number of images. For images that lie in a spherical cap, the error produced by global alignment is due to the curvature of the sphere.

Since we cannot align images from different views all at once, the distance computed between images after global alignment is not a truly rotationally invariant distance. In Section 3, we introduce a new rotationally invariant image representation $\tilde{\mathbf{b}}$ and replace the rotationally invariant distance (1) by

$$d_{ij} = \|\tilde{\mathbf{b}}_i - \tilde{\mathbf{b}}_j\|. \tag{6}$$

The new rotationally invariant feature vector $\tilde{\mathbf{b}}$ needs to be lower dimensional (so that (6) can be computed efficiently), and to retain the information in the image (so that (6) is meaningful). Using the rotationally invariant feature vectors, we are able to find images with similar views without performing rotational alignment.

2.2. Classification instead of clustering

Traditionally, the class averaging problem was considered as a clustering problem, in which a large data set of n images, $I_1, ..., I_n$ with unknown corresponding rotation matrices $R_1, ..., R_n$, is grouped into clusters, with the goal that images within a single cluster have similar viewing angles. In practice, however, the size of the cluster varies considerably from cluster to cluster (see Figure 4, where we tried to cluster 10^4 clean 70S ribosome projection images, whose viewing angles are uniformly distributed over the sphere), and therefore the resulting class averages will have different signal to noise ratio and resolution.
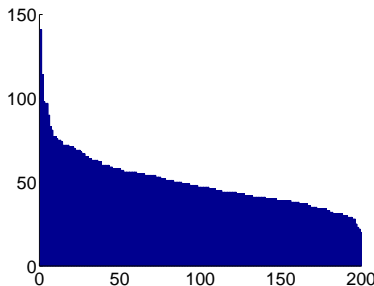

Figure 4: Number of particles in each cluster of the 200 clusters found by K-means clustering algorithm implemented in SPIDER. The data set has 10^4 clean centered images, whose viewing angles are uniformly distributed over the sphere.

Instead of K-means clustering and generating cluster means, we propose another classification method for generating class averages. For each image, we search for a fixed number (κ) of nearest neighbors. Each image is averaged with its aligned nearest neighbors to boost the signal to noise ratio. In this way, the resulting number of class averages is the same as the number of the original images and all class averages have the same SNR. It also prevents the situation that clustering reduces the full coverage of the viewing directions.

3. Methods

3.1. Fourier-Bessel Steerable PCA

We use Fourier-Bessel steerable PCA [21] to generate a data adaptive basis for compressing and de-noising images. Since the rotated copies of the

11

projection images are equally likely to appear in the data set, it is meaningful to perform PCA on the data set with all their rotated copies. However it is challenging to compute the steerable PCA efficiently and accurately, because the images are sampled on a Cartesian grid, while steering operations often require a polar grid. As the transformation from Cartesian to polar is not unitary, the eigenimages corresponding to images mapped to polar grid are not equivalent to transforming the original eigenimages from Cartesian to polar. In [21], we developed an accurate and efficient algorithm for steerable PCA. Its computational complexity is lower than that of traditional PCA (or MSA in cryo-EM 2D image processing). Since we incorporate more information from the data set, we can get better estimation of eigen-images that correspond to the clean projection images than the traditional PCA.

The steerable eigen-images have special separation of variables form,

$$u^{k,q}(r,\theta) = f^{k,q}(r)e^{\iota k\theta}, \tag{7}$$

where k and q in basis image $u^{k,q}$ are indices for angular frequency and radial frequency, respectively. $f^{k,q}$ can be computed from the Fourier-Bessel steerable PCA [21], which provides an optimal basis in the least-squares sense. Images are expanded on this steerable basis, $I(r,\theta) = \sum_{k,q} a_{k,q}u^{kq}(r,\theta)$, with expansion coefficients $a_{k,q}$. It is easy to "steer" the images. When image I is rotated counter-clockwise by angle α, the expansion coefficients of $I(r,\theta-\alpha)$ are given by $a_{k,q}^{\alpha} = a_{k,q}e^{-\iota k\alpha}$, because

$$\begin{aligned} I(r,\theta-\alpha) &= \sum_{k,q} a_{k,q}u^{kq}(r,\theta-\alpha) \\ &= \sum_{k,q} a_{k,q}e^{-\iota k\alpha}u^{kq}(r,\theta). \end{aligned} \tag{8}$$

The steerability of the basis allows us to define rotationally invariant features that are introduced in Section 3.2.

3.2. Bispectrum-like Rotationally Invariant Image Representation

Prior to introducing the rotationally invariant image representation, we quickly review here the bispectrum for 1D signals. Suppose we have a 1D periodic discrete signal $f(x), x = 1, ..., L$. The discrete Fourier transform of f is defined as

$$\hat{f}(k) = \sum_{x=1}^{L} f(x)e^{-i\frac{2\pi}{L}kx}. \tag{9}$$

The power spectrum $|\hat{f}|^2$ is the Fourier transform of the autocorrelation function

$$ACF(x) = \sum_{y=1}^{L} \overline{f(y)} f(y+x). \tag{10}$$

Both the power spectrum and the auto-correlation function are shift-invariant. However, the ACF loses the phase information in $\hat{f}$ and maintains only its amplitude. The idea behind bispectral invariants is to move from the auto-correlation function to the triple-correlation function

$$T(x_1, x_2) = \sum_{y=1}^{L} \overline{f(y)} f(y+x_1) f(y+x_2). \tag{11}$$

Again by the convolution theorem, Fourier transform of the triple-correlation function is

$$b(k_1, k_2) = \hat{f}(k_1) \hat{f}(k_2) \overline{\hat{f}(k_1 + k_2)}, \tag{12}$$

and is called the bispectrum of f. Under shift by z, the Fourier transform of $f^z = f(x - z)$ becomes

$$\hat{f}^z(k) = \sum_{x=1}^{L} f(x-z) \, e^{-i\frac{2\pi}{L}kx} = e^{-i\frac{2\pi}{L}kz} \sum_{x'=1}^{L} f(x') \, e^{-i\frac{2\pi}{L}kx'} = e^{-i\frac{2\pi}{L}kz} \hat{f}(k).$$
$$\tag{13}$$

Therefore, under translation by z, the bispectrum becomes

$$b^z(k_1, k_2) = e^{-i2\pi z k_1/L} \hat{f}(k_1) e^{-i2\pi z k_2/L} \hat{f}(k_2) e^{i2\pi z(k_1+k_2)/L} \overline{\hat{f}(k_1 + k_2)} = b(k_1, k_2),$$
$$\tag{14}$$

which shows that the bispectrum is shift-invariant. Unlike the power spectrum, the bispectrum does not lose the phase information and under mild conditions, the original signal can be reconstructed from its bispectrum (up to translation). The bispectrum is widely used in signal processing as a lossless shift-invariant representation, and various algorithms have been devised to reconstruct f from b [23]. Because of the symmetry properties of bispectrum coefficients, the knowledge of the bispectrum in the triangular region $k_1 \geq 0$, $k_2 \leq k_1$, $k_1 + k_2 \leq k_{max}$ is sufficient for a complete description of the bispectrum.

13

For 1D periodic signals of length L, there are $O(L^2)$ bispectrum coefficients. Therefore, the bispectrum is of very high dimensionality. The possibility of using the bispectrum as shift or rotational invariant image representation for classification of cryo-EM images has been previously mentioned in [6, 25]. Due to its high dimensionality, the full bispectrum has never been used for analyzing large cryo-EM data sets to generate class averages.

The bispectrum of 1D periodic signals for shift invariant features can be extended to generate rotationally invariant features for 2D images. We use Fourier-Bessel steerable PCA basis [21] described in Section 3.1 to expand images. Rotating the image is equivalent to phase shifting its expansion coefficients, which is similar to phase shifting the Fourier coefficients in (13).

Typically, most of the energy of the clean images is concentrated in a relatively small number M (a typical value of M is around 100 for noisy 2D images) of pricncipal components with low angular frequencies ($-k_{max} \leq k \leq k_{max}$), whereas the additive white Gaussian noise spreads over all components with low angular frequencies. Representing the images using only the leading M components can compress and denoise the images. Therefore, we use the truncated expansion coefficients with M terms instead of the total number of pixels.

We define the bispectrum for the steerable basis expansion coefficients as

$$b_{k_1,k_2,q_1,q_2,q_3} = a_{k_1,q_1} a_{k_2,q_2} \overline{a_{k_1+k_2,q_3}}, \tag{15}$$

where k_1 and k_2 are the angular indices and q_1, q_2 and q_3 are the radial indices.

A modification to the bispectrum is needed when treating noisy signals. Suppose the observed signal y is the true signal x contaminated with additive white Gausfdsian noise $\mathfrak{n} \sim \mathcal{N}(0, \sigma^2 I)$:

$$y = x + \mathfrak{n}. \tag{16}$$

Then the expansion coefficients are given by

$$a_{k,q}^y = a_{k,q}^x + a_{k,q}^{\mathfrak{n}}, \tag{17}$$

with $a_{k,q}^{\mathfrak{n}}$ satisfing $\mathbb{E} a_{k,q}^{\mathfrak{n}} = 0$ and $\mathbb{E}[a_{k_1,q_1}^{\mathfrak{n}} \overline{a_{k_2,q_2}^{\mathfrak{n}}}] = \sigma^2 \delta_{k_1 k_2} \delta_{q_1 q_2}$. Then the

expectation of the bispectrum of y,

$$
\begin{aligned}
\mathbb{E}b^y_{k_1,k_2,q_1,q_2,q_3} &= \mathbb{E}[(a^x_{k_1,q_1} + a^{\mathfrak{n}}_{k_1,q_1})(a^x_{k_2,q_2} + a^{\mathfrak{n}}_{k_2,q_2})(\overline{a^x_{k_1+k_2,q_3} + a^{\mathfrak{n}}_{k_1+k_2,q_3}})] \\
&= a^x_{k_1,q_1} a^x_{k_2,q_2} \overline{a^x_{k_1+k_2,q_3}} + \mathbb{E}\left[a^{\mathfrak{n}}_{k_1,q_1} a^{\mathfrak{n}}_{k_2,q_2} \overline{a^{\mathfrak{n}}_{k_1+k_2,q_3}}\right] \\
&\quad + a^x_{k_2,q_2} \overline{a^x_{k_1+k_2,q_3}} \mathbb{E}\left[a^{\mathfrak{n}}_{k_1,q_1}\right] + a^x_{k_1,q_1} \overline{a^x_{k_1+k_2,q_3}} \mathbb{E}\left[a^{\mathfrak{n}}_{k_2,q_2}\right] \\
&\quad + a^x_{k_1,q_1} a^x_{k_2,q_2} \mathbb{E}\left[\overline{a^{\mathfrak{n}}_{k_1+k_2,q_3}}\right] + a^x_{k_1,q_1} \mathbb{E}\left[a^{\mathfrak{n}}_{k_2,q_2} \overline{a^{\mathfrak{n}}_{k_1+k_2,q_3}}\right] \\
&\quad + a^x_{k_2,q_2} \mathbb{E}\left[a^{\mathfrak{n}}_{k_1,q_1} \overline{a^{\mathfrak{n}}_{k_1+k_2,q_3}}\right] + \overline{a^x_{k_1+k_2,q_3}} \mathbb{E}\left[a^{\mathfrak{n}}_{k_1,q_1} a^{\mathfrak{n}}_{k_2,q_2}\right].
\end{aligned} \tag{18}
$$

Hence,

$$
\mathbb{E}b^y_{k_1,k_2,q_1,q_2,q_3} = b^x_{k_1,k_2,q_1,q_2,q_3} + \sigma^2(\delta_{q_2,q_3} a^x_{0,q_1} + \delta_{q_1,q_3} a^x_{0,q_2} + \delta_{q_1,q_2} a^x_{0,q_3}). \tag{19}
$$

Therefore, if $a^x_{0,q} = 0$ for all q, then the bispectrum is unbiased, i.e., $\mathbb{E}\mathbf{b}^y = \mathbf{b}^x$. As a result, removing the zero-frequency part of the bispectrum makes it less sensitive to contamination by additive white Gaussian noise. The zero-frequency coefficients are rotational invariant and can be added as separate invariant features.

Van Heel *et al.* [6, 11] have previously noted that the ACF overweighs the already strong frequency components in the image due to the squaring of the Fourier components and therefore they defined a self correlation function (SCF) which under-emphasizes all amplitudes by replacing them by their square roots. The SCF was shown to perform better than the ACF. A similar situation occurs for the bispectrum, due to the multiplication of three frequency components. We therefore modify the expansion coefficients prior to computing the bispectrum such that the amplitude is the cubic root of the original:

$$
\tilde{a}^i_{k,q} = \frac{a^i_{k,q}}{|a^i_{k,q}|^{2/3}}. \tag{20}
$$

Notice that the phase information of the bispectrum is unaltered, as only the amplitudes are modified. It is natural to take the cubic root since in this way the bispectrum scales linearly with the intensity of the image (that is, multiplying an image I by the constant c results in multiplication of b by c, instead of c^3 for b).

The rotationally invariant image representation derived in (15) is of very high dimensionality. Suppose that the truncated expansion coefficients have M components and that the corresponding maximum angular frequency is

k_{max}, then the resulting invariant feature vector is of length $O\left(\frac{M^3}{k_{max}}\right)$. Computing the inner product of vectors of length $10^4 - 10^5$ can be quite expensive. It is therefore required to reduce the dimensionality of the invariant feature vectors. While this reduction can be achieved by PCA, the typically large number of images and the high dimensionality of the feature vectors make the computational cost of classical PCA quite demanding. Instead, we use the recently proposed randomized algorithm for low rank matrix approximation [26, 27, 28]. We denote by M' the reduced dimension, that is, the number of principal components chosen in this step.

We define the rotationally invariant affinity between image I_i and image I_j as the normalized cross-correlation C_{ij} between their corresponding low dimensional feature vectors of length M', where M' is about 200 in application.

A fixed number of nearest neighbors with the largest normalized cross-correlation C_{ij} with image i are determined, with computational complexity $O(n^2 M')$. For large data sets, consisting of 10^5 images or more, the randomized approximate nearest neighbor (RANN) search algorithm [29] is an efficient way for finding the nearest neighbors without computing C_{ij} for all pairs of i and j. RANN is an iterative algorithm. It first randomly rotates the data points (in our case, complex valued vectors of length M') and subdivides them into smaller boxes by looking at 1, 2, 3, 4... coordinates, until each box contains about κ points. Then the suspected nearest neighbors are determined locally as those in same boxes. The process is repeated through independent iterations, and the list of suspected neighbors is refined. In practice only a small number of iterations is needed in order to find the true nearest neighbors with very high probability. The computational complexity for this randomized algorithm is $O(Tn(M'\log M' + \kappa\log\kappa\log n) + n\kappa^2(M' + \log\kappa))$, where T is the number of iterations and κ is the number of nearest neighbors.

After classifying images of similar views, we rotationally align images with their nearest neighbors. The in-plane rotation angle α_{ij}^* for a pair of neighboring images I_i and I_j is determined by aligning their denoised versions.

3.3. Vector Diffusion Maps Classification and Rotational Alignment

When the SNR is very low, the initial rotationally invariant classification based on just nearest neighbors might still give some outliers. Further improvement can be obtained by taking into account the consistency of pairwise distances and rotational transformations among the images in the neighbor-

hood. This can be achieved by using a classification method called Vector Diffusion Maps (VDM) [31, 30], which is a generalization of Diffusion Maps, a popular method in manifold learning [33]. This method takes into account the consistency of in-plane rotational transformations (see Figure 5). The affinity between images I_i and I_j (shown as nodes i and j) is defined as the consistency of the transformations summed over all different paths of a fixed length connecting i and j. To quantify this, we build a sparse $n \times n$ Hermitian matrix H (21) using the union rule that i and j are neighbors if either i is one of j's κ nearest neighbors or j is one of i's κ nearest neighbors,

$$H_{ij} = \begin{cases} e^{\iota \alpha_{ij}^*} & \{i, j\} \in E, \\ 0, & \{i, j\} \notin E, \end{cases} \tag{21}$$

where E denotes the set of neighboring pairs and α_{ij}^* is the optimal in-plane rotation of images I_i and I_j. The fact that H is Hermitian follows from $\alpha_{ij}^* = -\alpha_{ji}^* \mod 2\pi$. Moreover, since only neighboring images contribute non-zero entries in H, it follows that H is a sparse matrix whose storage requires only $O(n\kappa)$ space. Each row of H is divided by the degree of the corresponding image, yielding the matrix S that is given by

$$S = D^{-1}H, \tag{22}$$

where D is an $n \times n$ diagonal matrix with

$$D(i, i) = \deg(i) = \sum_j |H_{ij}|. \tag{23}$$

The matrix S (22) is similar to the Hermitian matrix

$$\tilde{S} = D^{-1/2}HD^{-1/2} \tag{24}$$

through $S = D^{-1/2}\tilde{S}D^{1/2}$. We can define the affinity between i and j as $|\tilde{S}^{2t}(i, j)|^2$, that is, as the squared absolute value of $\tilde{S}^{2t}(i, j)$, which takes into account all paths of length $2t$, where t is a positive integer. In a sense, $|\tilde{S}^{2t}(i, j)|^2$ measures not only the number of paths of length $2t$ connecting i and j but also the amount of agreement between their transformations. That is, for a fixed number of paths, $|\tilde{S}^{2t}(i, j)|^2$ is larger when the path transformations are in agreement, and is smaller when they differ. We define the normalized affinity between i and j as

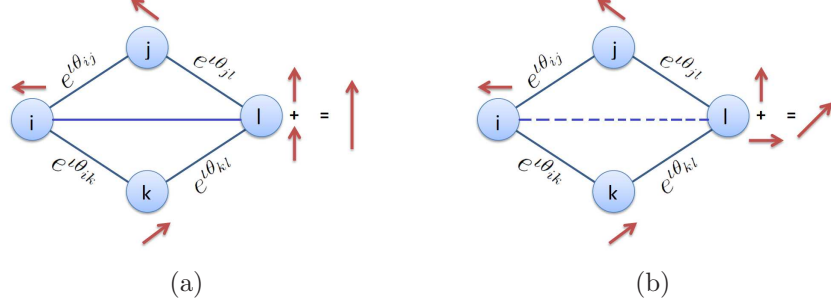$$\frac{|\tilde{S}^{2t}(i, j)|^2}{\sqrt{|\tilde{S}^{2t}(i, i)|^2 |\tilde{S}^{2t}(j, j)|^2}}. \tag{25}$$

Figure 5: Illustration for Vector Diffusion Map (VDM) affinity. Pick an arbitrary planar vector for node i (realized as a complex number $e^{\iota\phi}$). Consider two different paths from i to l of length 2: $i \to j \to l$ and $i \to k \to l$. The arrow is rotated according to the edges $i \to j$ and $i \to k$, respectively (by multiplying it by the phase factors, $e^{\iota\theta_{ij}}$ and $e^{\iota\theta_{ik}}$, respectively), and then rotated according to the edges $j \to l$ and $k \to l$, respectively (by multiplying it by the phase factors, $e^{\iota\theta_{jl}}$ and $e^{\iota\theta_{kl}}$, respectively). Different paths may be consistent as in (a) or inconsistent as in (b). When vectors from different paths are added together the amplitude of the resulting vector can be as large as the number of paths if they are all consistent (a), or much smaller due to inconsistencies (b). Node i and node l have higher affinity in (a) than in (b).

Since $\tilde{S}$ is Hermitian, it has a complete set of eigenvectors $v_1, v_2, ..., v_n$ and real eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$. We order the eigenvalues in decreasing order of magnitude. The spectral decomposition of $\tilde{S}$ and $\tilde{S}^{2t}$ are given by

$$\tilde{S}(i,j) = \sum_{l=1}^{n} \lambda_l v_l(i)\overline{v_l(j)}, \quad \text{and} \quad \tilde{S}^{2t}(i,j) = \sum_{l=1}^{n} \lambda_l^{2t} v_l(i)\overline{v_l(j)}. \tag{26}$$

It follows that the affinity $|\tilde{S}^{2t}(i,j)|^2$ is an inner product for the finite dimensional Hilbert space $\mathbb{C}^{n^2}$ via the mapping V_t :

$$V_t : i \mapsto ((\lambda_l \lambda_r)^t v_l(i)\overline{v_r(i)})_{l,r=1}^n. \tag{27}$$

That is,

$$|\tilde{S}^{2t}(i,j)|^2 = \langle V_t(i), V_t(j) \rangle. \tag{28}$$

18

Then the normalized affinity (25) can be expressed using the mapping V_t as

$$\frac{|\tilde{S}^{2t}(i,j)|^2}{\sqrt{|\tilde{S}^{2t}(i,i)|^2|\tilde{S}^{2t}(j,j)|^2}} = \left\langle \frac{V_t(i)}{|V_t(i)|}, \frac{V_t(j)}{|V_t(j)|} \right\rangle. \tag{29}$$

The matrix $\tilde{S}^{2t}$ may be too dense to be computed efficiently. Instead, we can approximate the normalized affinity (29) by truncating the mapping V_t to its leading m^2 coordinates (instead of n^2) as

$$V_t^m : i \mapsto ((\lambda_l \lambda_r)^t v_l(i)\overline{v_r(i)})_{l,r=1}^m. \tag{30}$$

where m is the largest integer satisfying $\lambda_m^{2t} > \delta$ for some δ much smaller than 1. The approximate normalized affinity becomes

$$\left\langle \frac{V_t^m(i)}{|V_t^m(i)|}, \frac{V_t^m(j)}{|V_t^m(j)|} \right\rangle. \tag{31}$$

We use (25) as the measure of closeness between two images to improve our estimation of the κ nearest neighbors for each image. This measure of affinity can be approximated using the eigenvectors of the matrix $\tilde{S}$ as shown in (30) and (31) [31]. The algorithm is very efficient in terms of running time and memory requirements, because it is based on the computation of the top eigenvectors of a sparse Hermitian matrix.

The eigenvectors of $\tilde{S}$ encode the information for in-plane rotational alignment between neighboring images. For clean images, if i and j are of the same viewing directions and their in-plane alignment angle is α_{ij}, the following holds

$$v_l(i) = e^{\iota \alpha_{ij}} v_l(j), \quad \forall l = 1, ..., n. \tag{32}$$

This is illustrated in Figure 6. When the viewing directions are close (though not identical), then (32) holds approximately. The level of approximation deteriorates as the eigenvalues become smaller, because their corresponding eigenvectors are more oscillatory and more sensitive to noise. Therefore, we use λ_l^{2t} to give more weight to the leading eigenvectors. We estimate the rotational angle using the top m eigenvectors:

$$\alpha_{ij}^* = \underset{\alpha_{ij}}{\operatorname{argmin}} \sum_{l=1}^m \lambda_l^{2t} |v_l(i) - e^{\iota \alpha_{ij}} v_l(j)|^2, \tag{33}$$
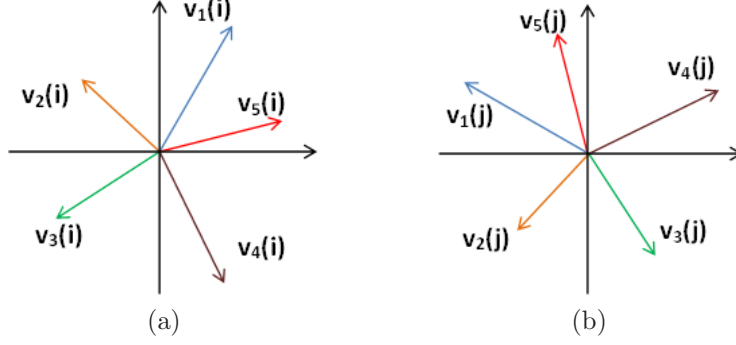
19

Figure 6: When i and j are of the same viewing angle, the tangent plane at point i coincides with the tangent plane at point j and the eigenvectors satisfy equation (32).

given by

$$e^{\iota\alpha_{ij}^*} = \frac{\sum_{l=1}^m \lambda_l^{2t} v_l(i)\overline{v_l(j)}}{|\sum_{l=1}^m \lambda_l^{2t} v_l(i)\overline{v_l(j)}|}. \tag{34}$$

In this way, we improve the estimation of the in-plane rotational alignment between nearest neighbors.

3.4. Shift Alignment

The experimental particle images are cropped from the micrographs through a particle selection procedure, and therefore they are not centered. Shift alignment is needed for generating class averages. Ideally we would like to center all images before performing rotational alignment and classification. What we are going to elucidate is that it is hard to center all projection images at the class averaging stage.

There are three degrees of freedom in defining the centers of all images. The three degrees of freedom correspond to the definition of the center of the three-dimensional molecule. We can fix the three degrees of freedom by choosing the center of mass of the volume as the origin. Then the center of mass of the clean projection images should also be at the origin. Therefore, for clean images with the same CTF function, we can center the images by finding the center of mass of the projection images. However, this method performs poorly at low SNR and when images are pooled together from different defocus groups. The practical procedure in the field is to shift-align

the images iteratively by correlating them with the mean of the data set or with a circular reference image. The estimation error for this procedure is typically of the order of 5 pixels in each direction.

To align images, we have to perform brute force shift search for the rudimentarily shift-aligned images. For image i and image j of the same view, with relative in-plane rotation angle α_{ij}^* and relative shift $(s_{ij,x}, s_{ij,y})$, the following equation holds,

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} - \begin{pmatrix} \cos\alpha_{ij}^* & -\sin\alpha_{ij}^* \\ \sin\alpha_{ij}^* & \cos\alpha_{ij}^* \end{pmatrix} \begin{pmatrix} x_j \\ y_j \end{pmatrix} = \begin{pmatrix} s_{ij,x} \\ s_{ij,y} \end{pmatrix}, \qquad (35)$$

where (x_i, y_i) and (x_j, y_j) are the location of the center of the projection images. Equation (35) is exact only when i and j share exactly the same viewing direction. When they are slightly different, this equation is not exact anymore. Therefore, the least squares solution to (35) does not produce the true global shifts (x_i, y_i). The least squares solution would perform well in aligning neighboring images, but it is not expected to find the shifts between different classes.

The rotationally invariant features described in Section 3.2 are not shift invariant. Therefore, we would like the images to be centered. However, as we have shown above, centering the images at the stage of class averaging is hard to achieve. As a result, in practice we use low pass filtering to make the images approximately shift invariant. During the classification by VDM we only use the consistency of the rotations. Once we identify the nearest neighbors and rotational alignment, we search for shift alignment in the small neighborhood. Although we cannot globally center the projection images in class averaging step, the centers can be estimated later on using common-lines [34].

4. Experimental Results

We performed numerical experiments to test the speed and accuracy of our algorithm on a machine with 2 Intel(R) Xeon(R) CPUs X7542, each with 6 cores, running at 2.67 GHz with 256GB RAM in total. These experiments were performed in MATLAB in UNIX environment.

4.1. Simulated noisy data

We compared our algorithms on simulated data against five 2D classification methods: RFA with K-means clustering implemented in SPIDER [7],

MSA/MRA implemented in IMAGIC [5], e2refine2d in EMAN2 [10], Relion 2D classification [16], and Xmipp CL2D [13]. The volume of *Escherichia coli* 70S ribosome-elongation factor G(EF-G) [7] was used to simulate projections. The image size is 129×129 pixels with 2.82Å/pixel. Images observed by an

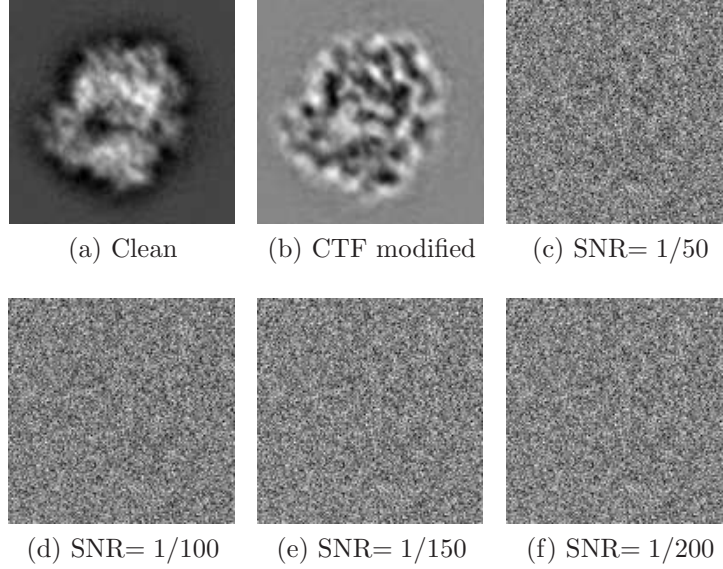(a) Clean   (b) CTF modified   (c) SNR= 1/50

(d) SNR= 1/100   (e) SNR= 1/150   (f) SNR= 1/200

Figure 7: Simulated 70S ribosome projection images. (a) Simulated clean centered projection image. (b) Clean projection image modified by Gaussian envelope function and Contrast Transfer Function (CTF). (c), (d), (e), and (f) Slightly shifted (randomly shifted within the range of ± 4 pixels in x and y directions) projection images with CTF contaminated with white Gaussian noise at SNR=$1/50, 1/100, 1/150,$ and $1/200$.

electron microscope are not true projections of the specimen. Imaging modifications include the effects of the contrast transfer function (CTF), which is introduced through electron lens aberrations and defocusing [35], and also the envelope function of the microscope, which contains contributions from a number of effects, such as spatial and temporal coherence, specimen motion, etc. [36]. In addition, background noise is present from a variety of sources. Therefore, we attempted to closely emulate the image formation process in the electron microscope including the effects of CTF, envelope function and noise. We projected 10^4 clean images at directions sampled uniformly over

the sphere (see Figure 7a). Then a Gaussian low-pass filter with half-width $1/10\text{Å}^{-1}$ was applied to simulate the effect of the envelope function. CTFs with different defocus values were applied to the images (see Figure 7b). The contrast transfer functions are generated according to the formula,

$$\text{CTF}(f) = \sin(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s)) + B\cos(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s)), \quad (36)$$

where the variable f is the spatial frequency, Δz is the defocus, c_s is the spherical abberation, λ is the electron wavelength, and B is the fraction of amplitude contrast. The imaging parameters were taken from the simulative data in SPIDER protocol [7]: electron beam energy $E = 200\text{KeV}$ with wavelength $\lambda = 0.025\text{Å}$ and spherical abberation is $c_s = 2.26\text{mm}$. The images were divided into 20 different defocus groups, with minimum defocus $1.5\mu\text{m}$ and maximum defocus $4\mu\text{m}$.

The centered projection images are randomly shifted within the range of ± 4 pixels in x and y directions. The images are then contaminated with additive white Gaussian noise at different signal to noise ratios, SNR= $1/50, 1/100, 1/150$, and $1/200$ (see Figure 7). The SNR in all our experiments is defined by

$$\text{SNR} = \frac{\text{Var}(Signal)}{\text{Var}(Noise)}. \quad (37)$$

The input images to our algorithm are first CTF corrected by phase flipping. More sophisticated CTF corrections are possible, but we find that phase flipping already produces satisfactory results.

In our simulation we know the original viewing angles, so for each image we compute the angles (in degrees) between the viewing angle of the image and the viewing angles of its 50 nearest neighbors. Small angles indicate successful identification of "true" neighbors that belong to a small spherical cap, while large angles correspond to outliers. We compute the percentage of nearest neighbor pairs whose viewing angles are within $18.2°$ spherical cap $(\cos(18.2°) = 0.95)$ as a measure of the quality of 2D image classification (see Table 1).

For experiments performed in SPIDER, all phase-flipped noisy images were filtered with a low-pass Butterworth filter, with the pass band and stop band at 0.08 and 0.12 respectively, given in reciprocal pixels, as described in [7]. To convert these values to Angstroms, divide the pixel size by the spatial frequency, i.e., in our case, $2.82/0.12\text{Å}^{-1} = 23.5\text{Å}$. We used a program in SPIDER (AP SR) to perform RFA on band-pass filtered projection images.

	RFA/K-means	MSA/MRA	Relion	EMAN2	Xmipp	ASPIRE
SNR= 1/50	0.45	0.97	0.79	0.74	0.83	**1.00**
SNR= 1/100	0.09	0.87	0.70	0.45	0.68	**0.99**
SNR= 1/150	0.07	0.67	0.52	0.13	0.48	**0.90**
Timing (hrs)	1.5	7.5	16	12	42	**0.5**

Table 1: Proportion of viewing angles of nearest neighbors that lie within 18.2°. Experiments are performed with 10^4 projection images of 70S ribosome at different noise levels. RFA was performed with AP SR program in SPIDER, the aligned particles were then classified into 200 groups using K-means algorithm. MSA/MRA was implemented in IMAGIC and was iterated for 5 times. We performed 25 iterations of Relion 2D class averaging, 10 iterations of e2refine2d in EMAN2, and 60 iterations of CL2D in Xmipp. The particles were classified in 200 classes, so that on average there were 50 particles in each class. In our algorithm, we found 50 nearest neighbors for each particle. The running time is measured for data with SNR= 1/100.

K-means clustering was used to classify the aligned and filtered images into $K = 200$ groups. Software description and details for performing the 2D image classification in SPIDER are available in [7]. The running time for generating 200 class averages is 1.5 hours (see Table 1).

For the experiments performed in IMAGIC, images were crudely centered by correlating the images with the data mean iteratively. The crudely centered images were first classified into 50 classes using MSA. Then 50 reference images were generated and the projection images were aligned with the references using multi-reference alignment. The aligned images were classified into 200 groups. The multi-reference alignment and MSA classification into 200 classes were iterated 3 more times to get the final alignment and classification results. More iterations of the MSA/MRA classification can improve the classification result. However each iteration took about 2 hours to finish for this data set.

We also tested the more modern cryo-EM SPR packages EMAN2, Xmipp, and Relion. The program e2refine2d in EMAN2 is very similar to the MSA/MRA algorithm in IMAGIC. The difference is that the initial classification is done on translational and rotationally invariant features. We used 10 iterations of the 2D class averaging in EMAN2. For experiments performed in Xmipp, we used CL2D algorithm for generating 2D class averages. The images were
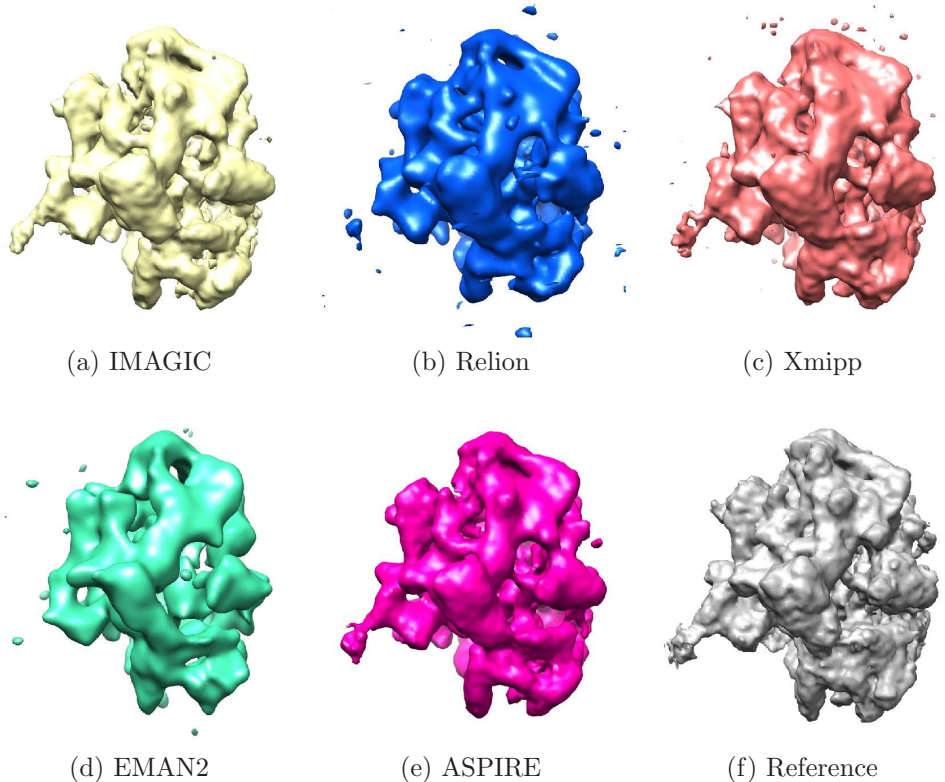
Figure 8: Ab initio models of 70S obtained from 10^4 simulated noisy projection images (SNR=1/100) with 20 defocus groups. The ab initio models are obtained by assigning orientations to the class averages using the common-lines based LUD method [37]. Reconstructed volumes from class averages generated by (a) MSA/MRA 2D image classification implemented in IMAGIC with 5 iterations, (b) 2D class averaging in Relion with 25 iterations, (c) CL2D in Xmipp with 60 iterations, (d) e2refine2d in EMAN2 with 10 iterations, and (e) 2D class averaging in ASPIRE (described in this paper). (f) Reference volume. The reconstructed volumes are Gaussian filtered.

classified into 8 classes initially and then refined into, 16, 32, 64, 128, and finally 200 classes. In each level, there were 10 iterations to refine classification and alignment. Relion employs an empirical Bayesian approach for 2D classification. We ran 25 iterations of 2D class classification in Relion. The

Step	Time (sec)
Fourier-Bessel sPCA	537.7
Rotationally Invariant Features	28.2
Initial Nearest Neighbor Search	13.9
VDM Classification	57.4
Local Alignment and Class Average	1081
Total	**1718.3 (28.6 min)**

Table 2: Timing for different steps of our 2D class averaging algorithm.

	no FBsPCA denoising	FBsPCA denoising
RFA/K-means	0.09	**0.48**
MSA/MRA	0.87	**0.95**
EMAN2	0.45	**0.76**
Xmipp	0.68	**0.96**

Table 3: Denoising using FBsPCA improves the classification results in RFA/K-means, MSA/MRA, EMAN2 and Xmipp 2D image classification (SNR= 1/100). Values in the table are the proportion of the viewing angles of particles in the same class that are within 18.2°.

accuracy and running time for 2D classification are detailed in Table 1.

We applied our rotational invariant viewing angle classification on the phase-flipped images. Our rotational invariant classification achieves better classification results in finding particles of similar views than the other five methods (see Table 1). Each image was aligned and averaged with its 50 nearest neighbors. It took about half an hour to generate 10^4 class averages. Table 2 summarizes the timing for each step of our algorithm.

In another set of experiments, we used Fourier-Bessel steerable PCA denoised images (SNR= 1/100) as the input for both SPIDER, IMAGIC, EMAN2 and Xmipp 2D classification programs. The classification results are greatly improved (see Table 3). This demonstrates that the denoising scheme we used in our pipeline is very useful for 2D image classification.

The resulting class averages were used to find common-lines. An *ab initio* estimate of the 3D orientations was determined by the least unsquared
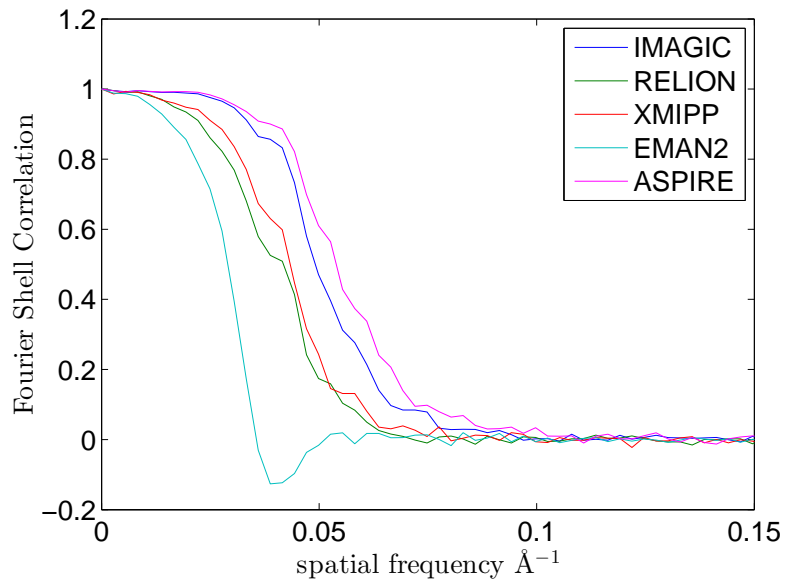
Figure 9: Fourier shell correlation of the reference volume with the ab initio models from different class averages (IMAGIC, Relion, Xmipp, EMAN2, and ASPIRE).
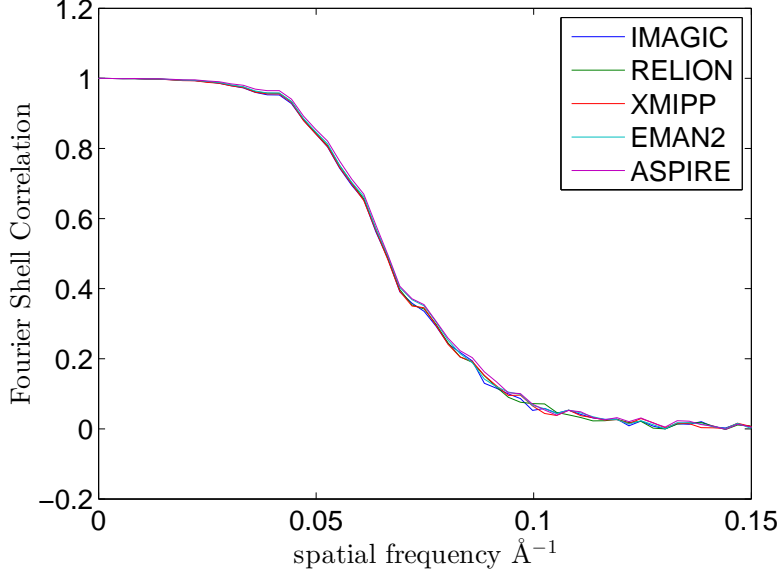
Figure 10: Fourier shell correlation of the reference volume with the refined models from different ab initio models (IMAGIC, Relion, Xmipp, EMAN2, and ASPIRE).

deviation (LUD) method [37], which is also available in the ASPIRE toolbox under "est_orientations_LUD.m". The reconstructed volumes from the class averages are shown in Figure 8. We were unable to reconstruct a meaningful model from the class averages generated by RFA/K-means procedure due to the large error in classification. The reconstructed volumes from the class averages produced by IMAGIC, Relion, Xmipp and EMAN2 and this paper were compared with the reference volume (Figure 8f). The *ab initio* model built from the class averages with this paper's methods agrees best with the reference volume (see Figure 9).

After *ab initio* reconstruction, we used Relion 3D auto-refine [38] to refine those five different *ab initio* models (IMAGIC, Relion, Xmipp, EMAN2, and ASPIRE) with simulated projection images whose SNR is 1/100. The FSC curves look very similar for the refined models (see Figure 10). However it takes different number of iterations to reach convergence (see Table 4). Refinement starting from ASPIRE *ab initio* model converged most quickly

IMAGIC	Relion	EMAN2	Xmipp	ASPIRE
17	18	20	18	**14**

Table 4: Number of refinement iterations needed for convergence starting from different ab initio models (IMAGIC, Relion, Xmipp, EMAN, and AS-PIRE). We used Relion 3D auto-refine for refinement.

and it took 14 iterations. The FSC curves (in Figure 9) and the number of iterations (in Table 4) show that the quality of the *ab inito* volume affects the refinement's convergence rate.

4.2. Experimental data: 70S ribosome

We applied the pipeline of image denoising, classification and alignment to an experimental data set provided by Dr. Joachim Frank's [39]. This data set comes from a larger heterogeneous data set with $216,517$ particles. ML3D [40] was used to separate the data into 6 more homogeneous subsets. The data used here is class number 6 and contains $40,778$ projection images of 70S ribosome (see top row of Figure 11). The images are of size 250×250 pixels with 1.5Å/pixel and the electron beam wavelength $\lambda = 0.0197$Å. They were pooled together from 77 different defocus groups and CTF corrected by phase-flipping. We split the data set randomly into two equally sized groups, each containing $20,389$ images. 50 nearest neighbors and the corresponding rotational and shift alignment were identified for each image. The second row of Figure 11 shows the averaged images. 1500 class averages were used to build a *ab initio* model for each group, with the common-lines based method [41, 37] for orientation determination.

The *ab initio* volumes (see Figure 12a and 12b) are consistent with each other up to 11.53Å. Below the corresponding frequency, the Fourier shell correlation (blue line in Figure 13) between the two volumes is above 0.143. The *ab initio* model was refined in Relion 3D auto-refine [38]. The refined model achieves 8.58Å resolution with 0.143 cutoff and 10.25Å with 0.5 cutoff (see red dot-dash line in Figure 13). Our refined model achieves higher resolution than the previously reported resolution 11.5Å, with 0.5 cutoff criterion for FSC [39]. Note that in our refinement process, two volumes were refined independently until the refinement converges whereas in the previous work [39], the refinement was not done independently with the gold-standard FSC. With our *ab initio* model, the refined model achieves higher resolution.
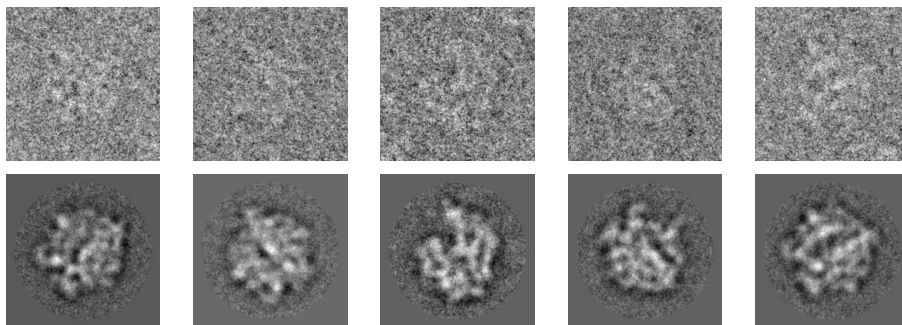
Figure 11: Top row: Samples of experimental images for 70S ribosome. Bottom row: Class averages by averaging the raw images of the top row with their 20 aligned nearest neighbors. Courtesy of Dr. Joachim Frank.

To compare with another 2D class averaging method, we used Relion 2D classification to generate 400 class averages for each group. About 60 good class averages in each group were chosen to generate *ab initio* models. The resolution for the *ab initio* model is 15.38Å with 0.143 cutoff criterion (see magenta line in Figure 13). The refined model achieves the same resolution as the refined model from ASPIRE (see Figure 13). The refinement took 20 iterations to converge, three more iterations than was needed for ASPIRE *ab initio* model. Therefore, our 2D class averaging method improved the resolution of the *ab initio* model of 70S ribosome and the refinement converged more quickly.

4.3. Experimental data: 50S ribosomal subunit

A set of micrographs of *E. coli* 50S ribosomal subunit was provided by Dr. Marin van Heel. We applied our algorithms to this data set, which contains $27,121$ projection images of the 50S ribosomal subunit. These micrographs were acquired by a Philips CM20 electron microscope at 9 different defocus values between 1.37 and 2.06μm. Each image (see top row of Figure 14) is of size 90×90 pixels with 3.36Å/pixel. The particles were picked using the automated particle picking algorithm in EMAN Boxer [42]. Then using the IMAGIC software package [5], the images were phase-flipped to remove the phase reversals in the CTF, bandpass filtered at $1/150$ and $1/8.4\text{Å}^{-1}$, and normalized by their variance. The images were initially crudely centered by correlating them with a fixed circularly-symmetric reference (rotationally
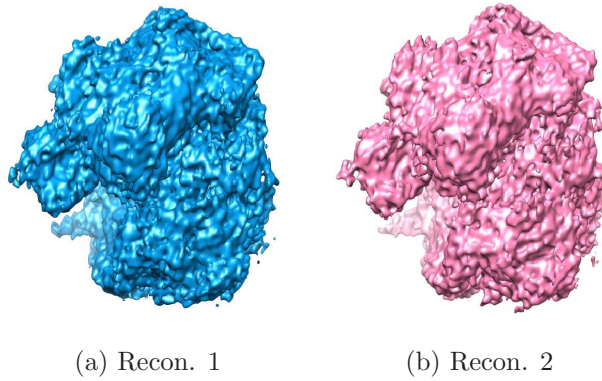
(a) Recon. 1   (b) Recon. 2

Figure 12: Ab initio reconstructions of 70S ribosome from two independent data sets. (a) Snapshot of ab initio volume 1. (b) Snapshot of ab initio volume 2.
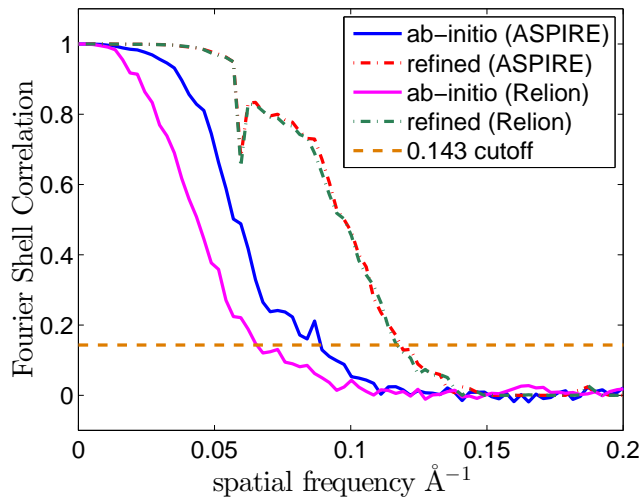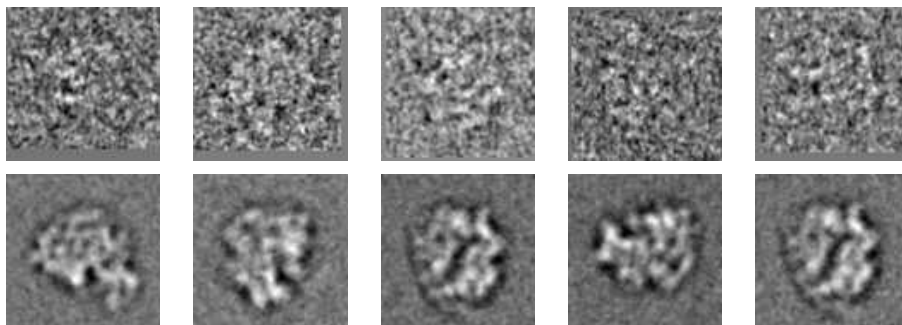

Figure 13: Fourier shell correlation curves for ab initio models and refined models. With 0.143 cutoff criterion, the resolution is 11.53Å for ASPIRE ab initio model (blue) and 15.38Å for Relion ab initio model (magenta). Both refined models achieve 8.58Å resolution according to gold-standard FSC (green and red dot-dash lines).

Figure 14: Top row: Samples of experimental images of 50S ribosomal subunit. Bottom row: Class averages by averaging the raw images of the top row with their 50 aligned nearest neighbors (including reflected images). Courtesy of Dr. Marin van Heel.

averaged total sum of the data).

We split the data set randomly into two groups of size 13, 560 to generate class averages and reconstructions separately. Each image was identified with 50 nearest neighbors (including reflection) and aligned to get class averaged images. We randomly chose 200 class averages in each group to build the *ab initio* models with the common-lines based method [41, 37] for orientation determination. Figure 14 shows 5 arbitrarily chosen class averaged images produced by our algorithm. The two volumes (see Figure 15) are consistent with each other up to 9.75Å with 0.143 cutoff criterion (see blue line in Figure 16). We refined the *ab initio* model using Relion 3D auto-refine [38], and it took 20 iterations to converge to the refined resolution 8.64Å with gold-standard FSC (see red dot-dash line in Figure 16).

We used Xmipp CL2D to generate class averages for comparison. CL2D computed 256 class averages for each group and all class averages were used to build *ab initio* models. The resolution for the *ab initio* model is 15.91Å with 0.143 cutoff criterion (see magenta line in Figure 16). The refined model achieves the same resolution as the refined model from ASPIRE (see Figure 16). The refinement took 19 iterations to converge, one less iteration than was needed for ASPIRE *ab initio* model. In this example, our class averaging method improved the resolution of the *ab initio* model. However the refinement starting from ASPIRE *ab initio* model did not converge faster than the refinement starting from Xmipp *ab initio* model.
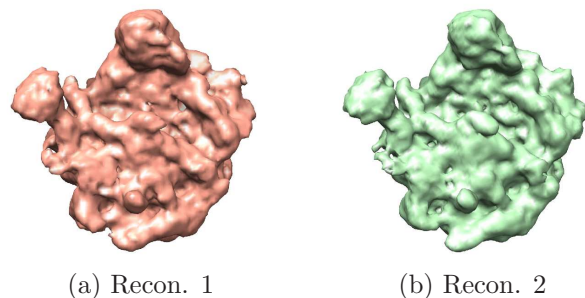
(a) Recon. 1   (b) Recon. 2

Figure 15: Reconstructions of 50S ribosomal subunit from two independent data sets. (a) Snapshot of reconstructed volume 1. (b) Snapshot of reconstructed volume 2.
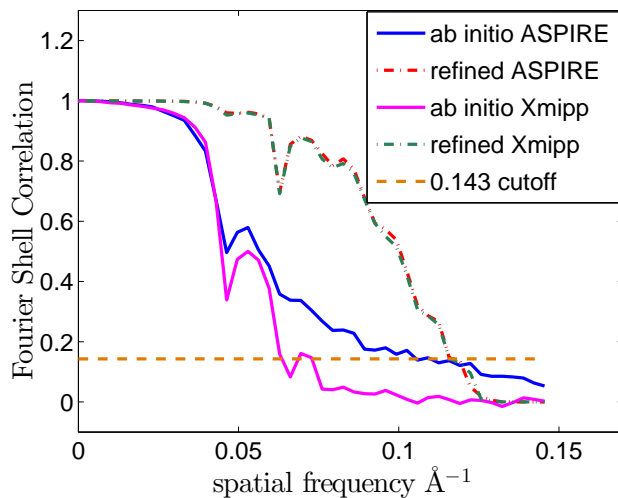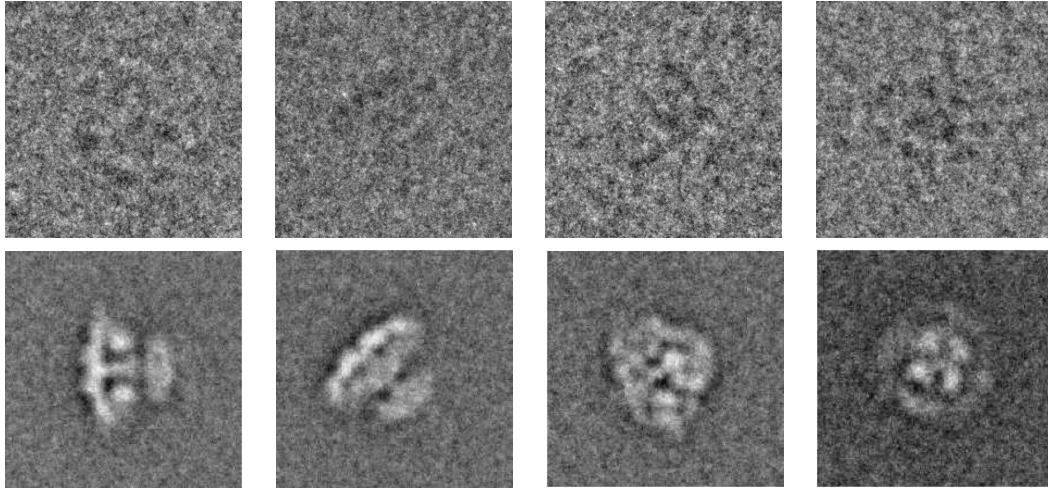

Figure 16: Fourier shell correlation curves for ab initio models and refined models. With 0.143 cutoff criterion, the resolution is 9.75Å for ASPIRE ab initio model (blue) and 15.91Å for Xmipp ab initio model (magenta). Both refined models achieve 8.64Å resolution according to gold-standard FSC (green and red dot-dash lines).

Figure 17: Top row: Samples of experimental images for IP$_3$R1. Bottom row: Class averages obtained by averaging the raw images of the top row with their 50 aligned nearest neighbors. Courtesy of Dr. Irina Serysheva.

4.4. Experimental data: IP$_3$R1

A set of Inositol $1, 4, 5$-triphosphate receptor 1 (IP$_3$R1) particle images were provided by Dr. Irina Serysheva. The protein has four-fold symmetry. We are able to generate class averages (the botton row of Figure 17) from the original data set (the top row of Figure 17), which contains $37,382$ images of size 256×256 pixels. We refer the readers to [43] for the details of the data set. The experiment shows that our 2D class averaging method, especially the vector diffusion maps classification, also works for particles with non-trivial point group symmetries. The common-lines based *ab initio* orientation determination procedures [41, 37] have yet to be modified for particles with non-trivial point group symmetry, therefore, we did not attempt to reconstruct the 3D model for this data set.

5. Summary and Discussion

Vitreous-ice-embedded biological macromolecules show a great randomness in orientation. This randomness is exactly what is desired for obtaining high quality 3D reconstructions. However the variety of viewing angles poses a problem for methods that attempt to rotationally align all images since it

is mathematically impossible to bring all images to global alignment. This means that in practice, the distance computed from allegedly globally aligned images is not a rotationally invariant distance.

In this paper, we introduced a new 2D class averaging procedure. The algorithm has three major components: Fourier-Bessel steerable PCA for image compression and de-noising, bispectrum-like rotational invariant features for classification, and Vector Diffusion Maps for more robust nearest neighbor search and rotational alignment.

Fourier-Bessel steerable PCA is a fast and accurate procedure for computing the eigen-images of a set of 2D images and their in-plane rotated copies. It is a viable alternative to MSA for compressing and de-noising of the raw 2D images. We demonstrated that this image de-noising method improves the classification results in RFA based classification, MSA/MRA classification, EMAN2 and Xmipp.

Our rotationally invariant representation of images is based on the bispectrum of their expansion coefficients in the steerable basis. Although the resulting invariant feature vectors are of very high dimensionality, we are able to efficiently project them into a lower dimensional space that captures most variability. Alignment parameters are searched only for nearest neighbors. Reversing the order of alignment and classification leads to a signifcantly faster viewing angle classification. The algorithm scales almost linearly with the number of images by using a randomized algorithm for nearest neighbor search.

For low SNR, the method that uses direct normalized cross-correlation of the rotationally invariant feature vectors can have many misidentified neighbors. For such situations, Vector Diffusion Maps, a classification method which takes into account the consistency of in-plane rotational transformations between images within the neighborhood, is used to boost the initial viewing angle classification. The eigenvectors of the VDM matrix contain the information of in-plane rotation for nearest neighbor pairs and lead to a much faster and more accurate estimation of the rotational alignments.

Through both simulated and experimental data sets, we demonstrated that the new 2D class averaging procedure proposed in this paper is not only fast, but also very robust to noise compared with the commonly used class averaging methods in the field, such as those implemented in SPIDER, IMAGIC, EMAN2, Relion, and Xmipp. The *ab initio* models we built from the experimental data sets are of high resolution and they need fewer iterations of refinement to reach convergence. The methods presented in this

paper are also applicable for molecules with non-trivial point group symmetries. The 2D class averaging method described in this paper is freely available as part of our ASPIRE toolbox.

Acknowledgements

References

[1] J. Frank, Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State, 2nd Edition, Oxford University Press, New York, 2006.

[2] M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, A. Patwardhan, Single-particle electron cryo-microscopy: towards atomic resolution., Quarterly Reviews of Biophysics 33 (4) (2000) 307–369.

[3] L. Wang, F. J. Sigworth, Cryo-EM and single particles, Physiology (Bethesda) 21 (2006) 13–18, review. PMID: 16443818.

[4] J. Frank, Single-particle reconstruction of biological macromolecules in electron microscopy-30 years., Quarterly Reviews of Biophysics 42 (3) (2009) 139–158.

[5] M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, M. Schatz, A new generation of the IMAGIC image processing system, Journal of Structural Biology 116 (1996) 17–24.

[6] M. Schatz, M. van Heel, Invariant classification of molecular views in electron micrographs, Ultramicroscopy 32 (1990) 255–264.

[7] T. Shaikh, H. Gao, W. T. Baxter, F. J. Asturias, N. Boisset, A. Leith, J. Frank, SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs, Nature Protocols 3 (12) (2008) 1941–1974.

[8] P. A. Penczek, M. Radermacher, J. Frank, Three-dimensional reconstruction of single particles embedded in ice, Ultramicroscopy 40 (1992) 33–53.

[9] P. A. Penczek, J. Zhu, J. Frank, A common-lines based method for determining orientations for $N > 3$ particle projections simultaneously, Ultramicroscopy 63 (3-4) (1996) 205–218.

[10] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, L. S. J., EMAN2: an extensible image processing suite for electron microscopy, J. Struct. Biol. 157 (1) (2007) 38–46.

[11] M. van Heel, M. Schatz, E. Orlova, Correlation functions revisited, Ultramicroscopy 46 (1-4) (1992) 307–316.

[12] J. M. de la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldívar, J. M. Vargas, J. Carazo, C. O. Sorzano, RELION: Implementation of a Bayesian approach to cryo-EM structure determination, J. Struct. Biol. 184 (2) (2013) 321–328.

[13] C. O. Sorzano, J. R. Bilbao-Castro, Y. Shkolnisky, M. Alcorlo, R. Melero, G. Caffarena-Fernández, M. Li, G. Xu, R. Marabini, J. M. Carazo, A clustering approach to multireference alignment of single-particle projections in electron microscopy, J. Struct. Biol. 171 (2) (2010) 197–206.

[14] M. Hohn, G. Tang, G. Goodyear, P. R. Baldwin, P. A. Huang, Z. Penczek, C. Yang, R. M. Glaeser, P. D. Adams, S. J. Ludtke, SPARX, a new environment for Cryo-EM image processing, J. Struct. Biol. 157 (1) (2007) 47–55.

[15] Z. Yang, J. Fang, J. Chittuluru, F. J. Asturias, P. A. Penczek, Iterative stable alignment and clustering of 2D transmission electron microscope images, Structure 20 (2) (2012) 237–247.

[16] S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination, J. Struct. Biol. 180 (3) (2012) 519–530.

[17] W. Park, D. R. Madden, D. N. Rockmore, D. R. Madden, G. S. Chirikjian, Deblurring of class-averaged images in single-particle electron microscopy, Inverse problems 26 (3) (2010) 035002.

[18] W. Park, C. R. Midgett, D. R. Madden, G. S. Chirikjian, A stochastic kinematic model of class averaging in single-particle electron microscopy, The International journal of robotics research 30 (6) (2011) 730–754.

[19] W. Park, G. S. Chirikjian, An assembly automation approach to alignment of noncircular projections in electron microscopy, IEEE Transactions on Automation Science and Engineering. Accepted (2014).

[20] C. Ponce, A. Singer, Computing steerable principal components of a large set of images and their rotations, IEEE Transactions on Image Processing 20 (11) (2011) 3051–3062.

[21] Z. Zhao, A. Singer, Fourier-Bessel rotational invariant eigenimages, J. Opt. Soc. Am. A 30 (5) (2013) 871–877.

[22] M. Michaelis, G. Sommer, A Lie group approach to steerable filters, Pattern Recognition Letters 16 (11) (1995) 1165 – 1174.

[23] B. M. Sadler, G. B. Giannakis, Shift- and rotation-invariant object recognition using the bispectrum, Jounal of Optical Society of America, A 9(1) (1992) 57–69.

[24] L. Joyeux, P. A. Penczek, Efficiency of 2D alignment methods, Ultramicroscopy 92 (2) (2002) 33–46.

[25] R. Marabini, J. M. Carazo, On a new computationally fast image invariant based on bispectral projections, Pattern Recognition Letters 17 (1996) 959–967.

[26] V. Rokhlin, A. Szlam, M. Tygert, A randomized algorithm for principal component analysis, SIAM J. Matrix Anal. Appl. 31 (2009) 1100–1124.

[27] N. Halko, P. G. Martinsson, J. A. Tropp, Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decomposition, SIAM Rev. 53(2) (2011) 217–288.

[28] N. Halko, P. G. Martinsson, Y. Shkolnisky, M. Tygert, An algorithm for the principal component analysis of large data sets., SIAM Journal on Scientific computing 33 (5) (2011) 2580–2594.

[29] P. W. Jones, A. Osipov, V. Rokhlin, A randomized approximate nearest neighbors algorithm, Proc. Natl. Acad. Sci. 108 (38) (2011) 15679–15686.

[30] A. Singer, Z. Zhao, Y. Shkolnisky, R. Hadani, Viewing angle classification of cryo-electron microscopy images using eigenvectors, SIAM Journal on Imaging Sciences 4(2) (2011) 543–572.

[31] A. Singer, H.-T. Wu, Vector diffusion maps and the connection Laplacian, Communications on Pure and Applied Mathematics (CPAM) 65 (8) (2012) 1067–1144.

[32] J. Milnor, Analytic proofs of the "hairy ball theorem" and the Brouwer fixed point theorem, The American Mathematical Monthly, 85 (7) (1978) 521–524.

[33] R. R. Coifman, S. Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (2006) 5–30.

[34] Y. Shkolnisky, A. Singer, Center of mass operators for cryo-EM–Theory and implementation, in: T. Vogt, W. Dahmen, P. Binev (Eds.), Modeling Nanoscale Imaging in Electron Microscopy, Nanostructure Science and Technology Series, Springer, 2012, pp. 147–177.

[35] J. Zhu, P. A. Penczek, R. Schröder, J. Frank, Three-Dimensional reconstruction with contrast transfer function correction from energy-filtered

cryoelectron micrographs: Procedure and application to the 70S *Escherichia coli* Ribosome, Journal of Structural Biology 118 (3) (1997) 197–219.

[36] K. J. Hanszen, The optical transfer theory of the electron microscope: fundamental principles and applications, In: Barer, R. and Cosslett, V. E. (Eds.), Advances in Optical and Electron Microscopy 4 (1971) 1–84.

[37] L. Wang, A. Singer, Z. Wen, Orientation determination from cryo-EM images using least unsquared deviation, SIAM Journal on Imaging Sciences 6 (4) (2013) 2450–2483.

[38] S. H. W. Scheres, Single-particle processing in RELION (2013).

[39] X. Agirrezabala, H. Y. Liao, E. Schreiner, J. Fu, R. F. Oritz-Meoz, K. Schulten, G. Green, J. Frank, Structural characterization of mRNA-tRNA translation intermediates, Proc. Natl. Acad. Sci. 109 (16) (2012) 6094–6099.

[40] S. H. W. Scheres, H. Gao, M. Valle, G. T. Herman, P. P. Eggermont, J. Frank, J. M. Carazo, Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization, Nature Methods 4 (2007) 27–29.

[41] A. Singer, Y. Shkolnisky, Three-Dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming, SIAM J. Imaging Sciences 4 (2011) 543–572.

[42] S. J. Ludtke, P. R. Baldwin, W. Chiu, EMAN: semiautomated software for high-resolution single-particle reconstructions., J. Struct. Biol. 128 (1) (1999) 82–97.

[43] S. J. Ludtke, T. P. Tran, Q. T. Ngo, V. Y. Moiseenkova-Bell, W. Chiu, I. I. Serysheva, Flexible architecture of IP$_3$R1 by cryo-EM, Structure 19 (8) (2011) 1192–1199.