

Model-Based Particle Picking for Cryo-Electron Microscopy

H. Chi Wong ^{*} Jindong (JD) Chen ^{*} Fabrice Mouche [†] Isabelle Rouiller ^{† ‡}
Marshall Bern ^{* §}

Abstract

We describe an algorithm for finding particle images in cryo-EM micrographs. The algorithm starts from a crude 3D map of the target particle, computed from a relatively small number of manually picked images, and then projects the map in many different directions to give synthetic 2D templates. The templates are clustered and averaged and then cross-correlated with the micrographs. A probabilistic model of the imaging process then scores cross-correlation peaks to produce the final picks. We give quantitative results on two quite different target particles: keyhole limpet hemocyanin and p97 AAA ATPase. On these particles our automatic particle picker shows human performance level, as measured by the Fourier shell correlations of 3D reconstructions.

Keywords: 3D template, cross-correlation, maximum likelihood, p97 AAA ATPase.

1 Introduction

Cryo-electron microscopy (cryo-EM) uses a transmission electron microscope to acquire 2D projections of a specimen preserved in vitreous ice [Dubochet, 1988]. A 3D electron density map can then be reconstructed from the 2D projections computationally. In “single-particle” cryo-EM, the specimen consists of many randomly oriented copies of a single type of particle; the 3D reconstruction process estimates the unknown orientations at the same time that it estimates the 3D structure.

Cryo-EM is already established as an important technique in structural biology [Abbott, 2002], because it can determine low- to medium-resolution 3D structures for membrane-bound proteins and large molecular assemblies, which are very difficult to crystallize. Single-particle cryo-EM has been used quite extensively to study the ribosome [Frank, 1995, Gabashvili, 2000], recently achieving about 10 Å resolution of the *E. Coli* 50S and 70S ribosomal subunits [Gabashvili, 2000, van Heel, 2000]. Cryo-EM can also be used in conjunction with x-ray crystallography, with x-ray data supplying atomic resolution (3 Å or better) for subunits whose overall arrangement is determined by cryo-EM.

A critical bottleneck in cryo-EM processing is particle picking—the identification of particle images within micrographs. Upwards of 100,000 particle images [Glaeser, 1999, Henderson, 1995] will be needed to achieve resolution better than 7 Å, primarily because the electron dose (and hence

^{*}Palo Alto Research Center, 3333 Coyote Hill Rd., Palo Alto, CA 94070 {bern,hcwong,jchen}@parc.com

[†]CB 129, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037 fmouche@scripps.edu

[‡]Current address: Burnham Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037 rouiller@burnham.org

[§]Contact author. Phone: (650) 812-4443 Fax: (650) 812-4471

the magnification and resolution) used to acquire images is limited by radiation damage to the sample. Manual or semi-automatic picking of 100,000 particle images would be a large and unpleasant task, requiring several man-months. Even for more modest numbers of particle images, say 5000–10,000, automatic picking offers faster and less subjective picks. In this paper, we describe a system with an initial manual set-up phase, followed by fully automatic particle picking. In the set-up phase, a person picks a relatively small number of particle images (say 500) and chooses a few algorithm parameters. All the remaining particle picking is automatic, without the need for a person to even view the micrographs.

The particle picking algorithm is based on a probabilistic model of cryo-EM imaging. The manual picks are used to compute an initial 3D map, which is then projected from various directions to form 2D templates. Templates are clustered and cluster representatives are cross-correlated with the micrographs. Probabilistic scoring then screens the cross-correlation peaks based on likelihood ratios: the probability that a peak was produced by a projection of the 3D map, compared to the probability that it was produced by noise alone.

1.1 Previous Work

Previous work on particle picking falls roughly into three categories: local statistical measures, cross-correlation, and edge detection. A recent article by Nicholson and Glaeser [Nicholson, 2001] surveyed more than 40 papers on the subject. It is fair to say that none of the current methods can be used generically without extensive manual correction [Nicholson, 2001].

Local statistical measures include mean [Kivioja, 2000], variance [van Heel, 1982] and kurtosis [Frank, 1996] computed in particle-sized regions. Manually selected positive and negative examples can be used to set thresholds or train a classifier for particle picking. These methods can give useful results for small, roughly spherical, particles, but are error-prone for large particles or particles with a wide variety of projections.

Cross-correlation is probably the dominant approach [Nicholson, 2001]. In the “traditional” realization, manually selected particles are clustered and averaged to form one or more 2D templates. Templates may be further processed by rotational averaging [Thuman, 1995] or filtering to accentuate salient features [Stoschek, 1997]. Cross-correlation of the micrographs with the templates suggests candidate picks, which are then screened for false positives using features based on size, texture [Lata, 1995], and statistical moments. Correlation coefficient (“local correlation” [Roseman, 2003]) is sometimes used instead of cross-correlation.

Faster computers have enabled a new realization (available in EMAN [Ludtke, 1999]), in which a large number of 2D templates are derived from a “3D template”, an initial, low-resolution model of the target particle. The advantage of a 3D template is that all of the manually selected images contribute to the 3D template and hence to each of the 2D templates, enabling the use of more—and less noisy—templates. EMAN screens candidate picks using lower and upper thresholds on contrast.

Edge detection has been applied successfully to filamentous particles [Zhu, 2001] and to the same hemocyanin micrographs used here [Zhu, 2002]. Edge detection, and its relatives such as contour finding and image segmentation, are probably not suitable for all particles, since many particles do not show strong edges. Moreover, edge detection is only a first step; the detected edges must be somehow grouped into particles. For hemocyanin, Zhu et al. [Zhu, 2002] used a particle-specific generalized Hough transform to group edge pixels into circles and rectangles, along with an additional special-purpose step to screen out the undesirable, long hemocyanin particles.

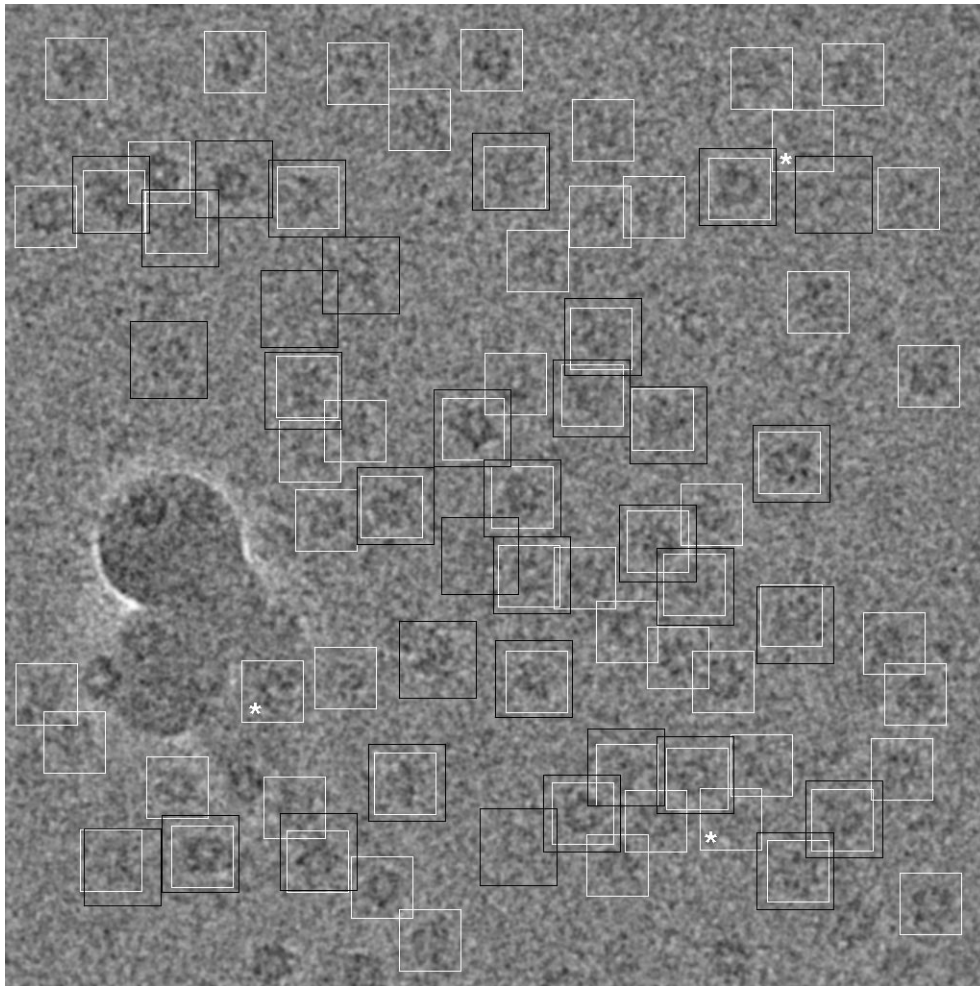


Figure 1: A 2 μm underfocused cryo-EM micrograph of p97 ATPase particles, filtered and gray-scale standardized as explained in the text. The white boxes show the 69 particle images picked by our fully automatic 3D template method, and the black boxes show the 35 semi-automatic picks made by an expert using SPIDER and manual correction. There are 27 picks in common. We think that our automatic picker has about 1–4 false positives and 8–12 false negatives. Three questionable picks are marked \star , the one on the left because it overlies a dark contaminant and the other two because we are not sure that they are real.

1.2 Our Contributions

We propose an explicit probabilistic model of the imaging process. The model includes a signal component, derived from a 3D template as in EMAN, and a noise component derived from the empirical distribution of gray values in a micrograph. Our current noise model is very simple—it assumes independent, identically distributed pixels—but in principle our approach could accomodate much more elaborate noise models.

In implementation, our method resembles cross-correlation followed by statistical screening. Probabilistic scoring, however, has a number of advantages over more *ad hoc* screening methods: it applies to all “round” particles, there is no feature computation nor training, and it ranks candidates with different orientations on a common likelihood scale. For our two test particles our method outperformed EMAN’s 3D template method, even when we chose EMAN’s screening thresholds indi-

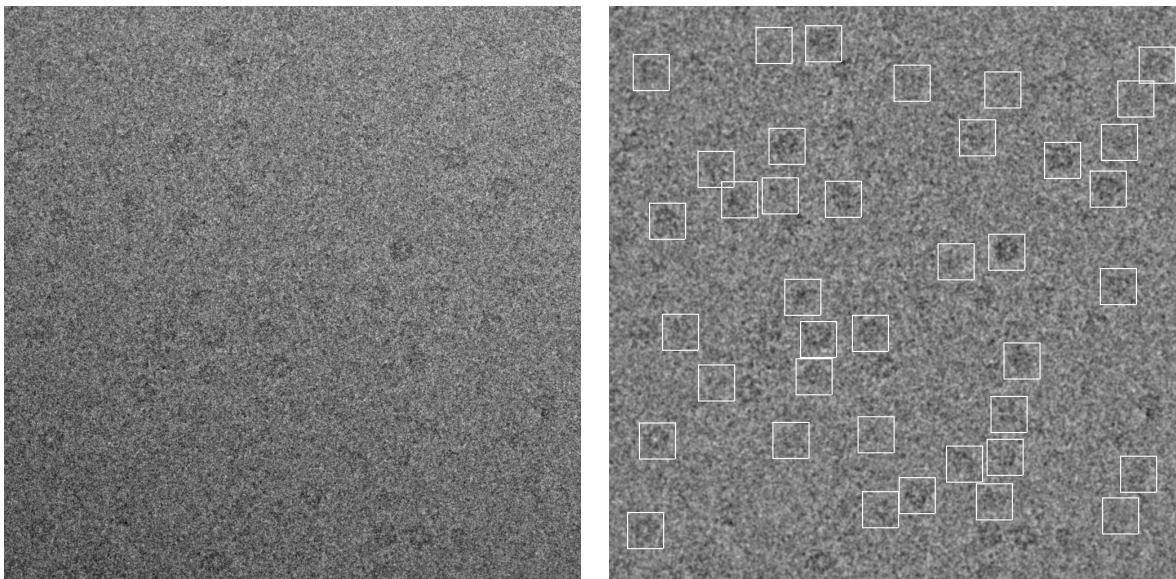


Figure 2: The image on the left shows a raw micrograph with some top-to-bottom shading due to varying ice thickness; the image on the right shows the filtered and standardized micrograph with picked p97 particles. The same thresholds were used for all the p97 micrographs.

vidually for each micrograph.

Finally, we raise the standards of evaluation for particle picking algorithms. Most previous papers simply give example micrographs. We give quantitative results, including resolution numbers for 3D reconstructions. The bottom line is that our automatic picks gave equally good resolution as data sets assembled manually and semi-automatically by experts.

2 Materials and Methods

The electron microscope images were acquired using the Leginon system [Carragher, 2000] at the Scripps Research Institute’s Center for Integrative Molecular Biosciences. Leginon collected pairs of digital micrographs with different levels of defocus. The locations of particles are most evident to the eye—and to computer vision—in the far-from-focus ($1\text{--}3\ \mu\text{m}$ underfocus) images; however, the high-spatial-frequency information crucial to 3D reconstruction is more easily obtained from the close-to-focus ($.2\text{--}1.2\ \mu\text{m}$ underfocus) images. Hence our cryo-EM particle picking algorithm uses the more defocused micrographs, but 3D reconstruction uses the less defocused micrographs.

Our first data set consisted of 17 defocus pairs (at about $-1\ \mu\text{m}$ and $-2\ \mu\text{m}$) of p97 AAA ATPase [Rouiller, 2000, Rouiller, 2001, Rouiller, 2002], collected with a Philips CM200 FEG microscope with a LaB6 filament, operating at 120 KV at a nominal magnification of $66,000\times$. Micrographs were recorded on a 1024×1024 Gatan MSC CCD camera (24 micron pixels) at low-dose conditions ($10\ e^-/\text{\AA}^2$), giving an image pixel size of $66/24 = 2.75\ \text{\AA}$. Figure 1 shows a far-from-focus micrograph containing many images of the p97 hexamer, which is shaped something like six short barrels standing in a hexagon, about $150\ \text{\AA}$ across.

Our second data set consisted of 82 defocus pairs (at about $-1\ \mu\text{m}$ and $-3\ \mu\text{m}$) of keyhole limpet hemocyanin [Orlova, 1997], collected with a Philips CM200 FEG microscope on a TVIPS 2048×2048 Peltier cooled CCD camera (24 micron pixels) with an image pixel size of about $2.2\ \text{\AA}$.

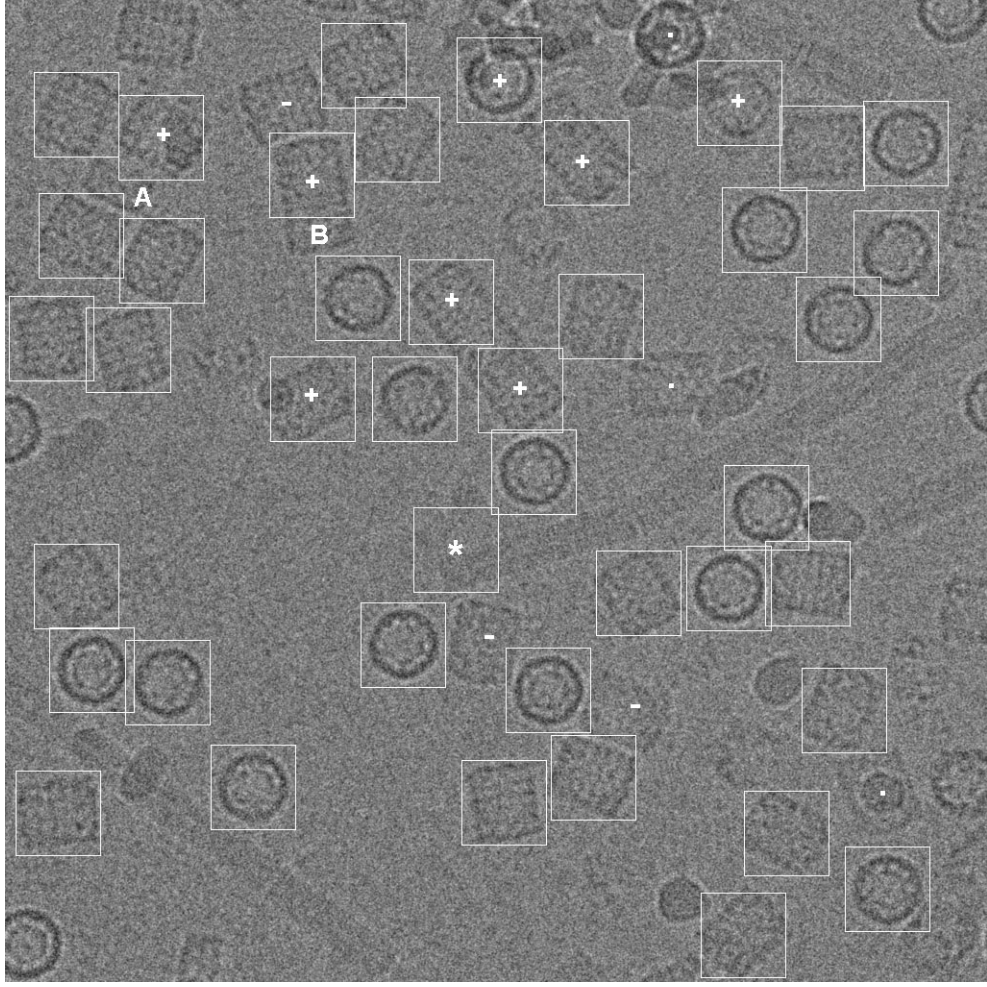


Figure 3: A 3 μm underfocused cryo-EM micrograph of hemocyanin particles. This micrograph shows preferred orientations; the “side views” (rectangles) are the valuable images for reconstruction because they represent various rotations around the axis of the cylinder. There are also particle halves (A) and long variants (B). We score the results on this micrograph as 34 perfect picks, 7 imperfect picks (marked +), 1 bogus pick (marked *), 3 perfect misses (marked -), and 3 imperfect misses (marked ·). (Since bogus picks are rare, we added this one by hand for illustration.) Particles partly off the micrograph cannot be used and hence were not counted at all.

Figure 3 shows a far-from-focus hemocyanin micrograph. Along with the usual hemocyanin (KLH1) didecamer particles, a cylindrical “cage” about 400 Å long, the sample included half particles (decamers) and long particles (multidecamers).

2.1 Probabilistic Model

We consider a cryo-EM image I to be a 2D array of independent random variables I_{ij} . The distribution of the random variable I_{ij} depends upon whether or not a particle is present.

If no particle is present, then each I_{ij} is an *ice random variable*, independent and identically distributed (i.i.d.) with distribution set empirically by a gray-level histogram of a relatively uncontaminated micrograph in the same series. In Figure 5 the curve marked “ice.hist” shows the histogram

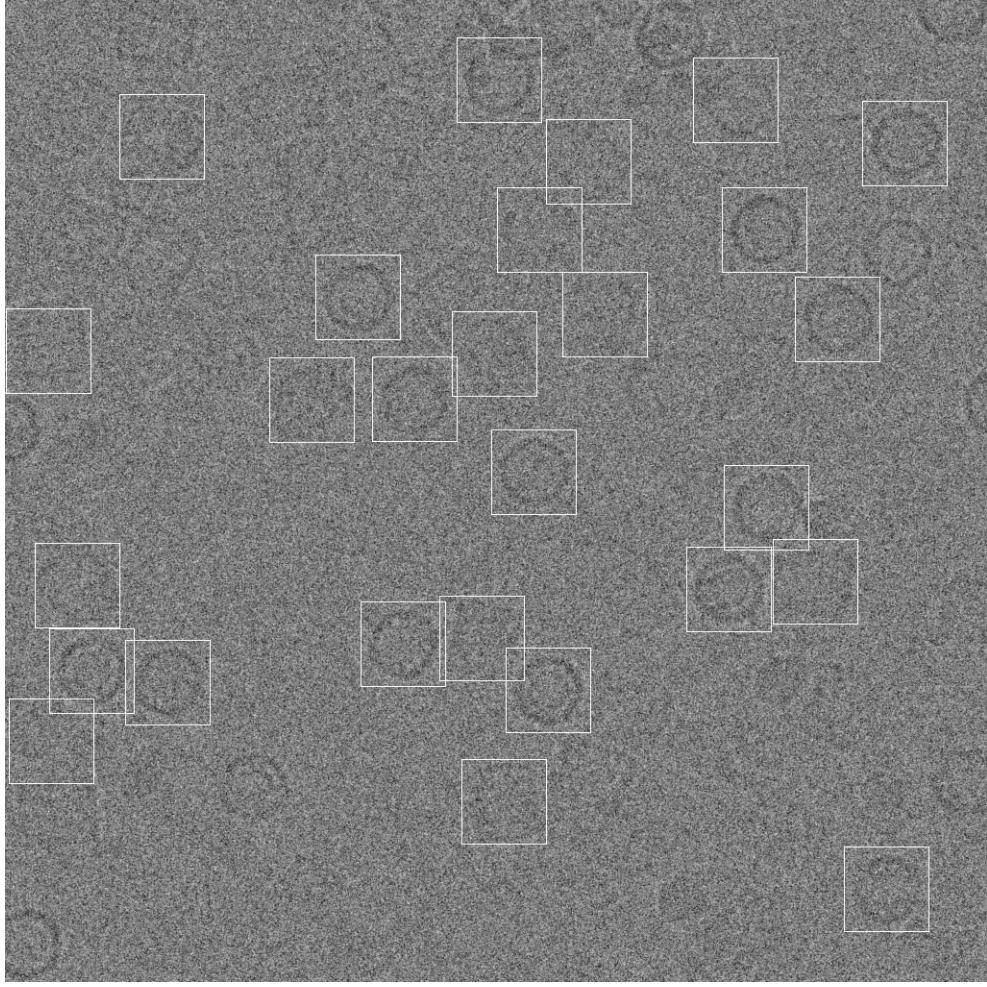


Figure 4: A semi-synthetic micrograph made by averaging Figure 3 with Gaussian noise in the proportion 30% to 70%, reducing the signal-to-noise ratio to about .1 for side views. By adding noise to a very visible particle like hemocyanin, we can simulate faint particles while still knowing the ground truth. Of the 27 picks made by our automatic picker only 1 is bogus.

used for the p97 micrographs. The micrographs have 8-bit or 12-bit gray levels, but we quantized to 0–127 for probabilistic scoring. Similar quantization would enable the use of 16-bit or 24-bit gray micrographs without increasing the program’s memory requirements.

If a particle is present, then I_{ij} is a weighted average of two random variables: an ice random variable with the same “ice.hist” distribution, and a *template random variable* with distribution determined by a 3D template Φ , along with parameters α , β , γ , and τ , giving the orientation of Φ relative to the overall image I . Parameters α , β , and γ specify rotation (Euler) angles around the x -, y -, and z -axes; we always project parallel to the z -axis, so we can think of α and β as specifying a viewing direction in 3D and γ as specifying a planar rotation. Parameter $\tau = (x, y)$ specifies a planar translation of the projection of Φ relative to I .

Given the orientation of Φ relative to I , the template random variable for I_{ij} is a Gaussian random variable with mean equal to the sum of density values in Φ along the line projecting onto I_{ij} , normalized to a 0–127 scale. The standard deviation σ is set equal to the square root of the mean. This σ approximates the Poisson counting statistics of capturing electrons; a better approximation

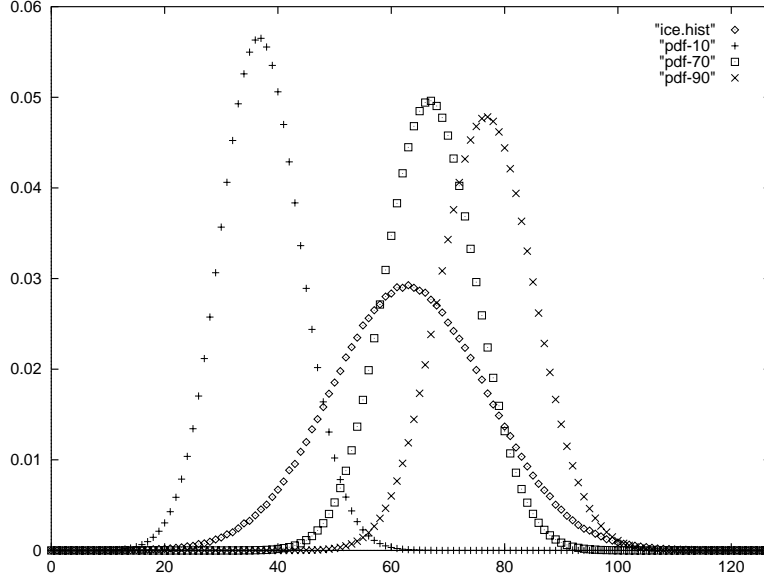


Figure 5: The “ice.hist” curve shows the gray-level histogram of a filtered p97 micrograph with 7-bit gray levels, that is, values from 0 to 127. This curve fits a Gaussian distribution quite well; its right tail is just a bit fat. The curves pdf-10, pdf-70 and pdf-90 show the model probability density functions for micrograph pixels corresponding to template pixels of gray levels 10, 70, and 90, respectively.

could be obtained—at least in principle—by using either a binomial or Poisson distribution with parameters set according to electron dose, contrast transfer function, scintillator efficiency, and pixel size.

Regardless of the number of pixels in image I , there are only 128 different probability distributions for I_{ij} ’s, because each I_{ij} distribution is determined by “ice.hist”—which is fixed over the micrograph series—and by the 0–127 electron density of Φ along the appropriate projection line. We can compute the probability density function (p.d.f.) of I_{ij} for a given gray level simply by convolving the individual p.d.f.’s of the ice and template random variables. (The sum of two random variables has p.d.f. equal to the convolution of their individual p.d.f.’s.) The ratio of template to ice we leave as a tuning parameter, but in the experiments described below, we left *ice-ratio* fixed at 50% template and 50% ice, even though the p97 shows lower contrast than the hemocyanin. In general our algorithm is robust to changes in parameters: any *ice-ratio* from 30% to 60% template seems to work almost equally well. For further robustness, we cut off p.d.f. values at 10^{-6} ; that is, any observed gray level had probability at least 10^{-6} .

Figure 5 shows the ice histogram for a p97 micrograph, along with the model’s p.d.f.’s for I_{ij} assuming three different template gray levels. Notice that the middle of the p.d.f. for template gray level 10 occurs at 37, halfway between 10 and 64, the middle of the ice histogram. Notice also that the sum of template and ice has lower variance than the ice alone.

We can use this probabilistic model to compute a likelihood that a given image M is indeed a cryo-EM image of Φ . Let m_{ij} denote the gray value of pixel (i, j) in image M and let $I_{ij}(\alpha, \beta, \gamma, \tau)$ denote the corresponding random variable in our model of a cryo-EM image of Φ oriented by α, β, γ , and τ . Because we do not know the orientation parameters, we define the likelihood that M is an image of Φ as the maximum over all orientation parameters. We also take logarithms, in order to

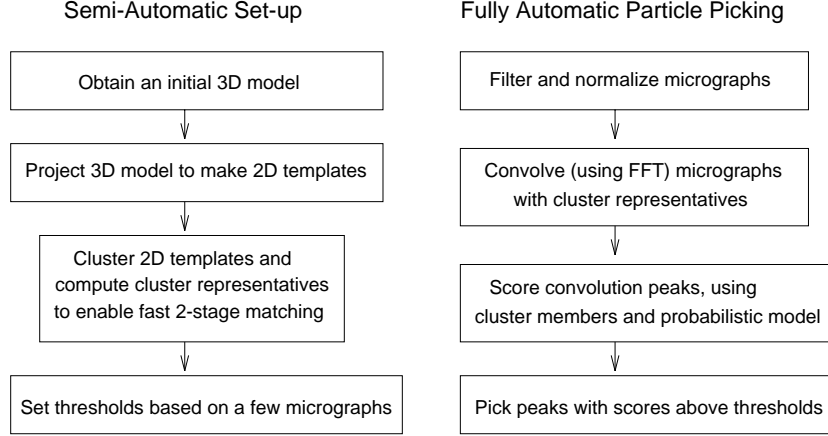


Figure 6: The steps of the algorithm include both preprocessing steps applied to templates (left) and production steps applied to micrographs (right).

turn the product into a sum. Hence the *log likelihood* of image M is

$$L(M) = \max_{\alpha, \beta, \gamma, \tau} \log \prod_{ij} P[I_{ij}(\alpha, \beta, \gamma, \tau) = m_{ij}] = \max_{\alpha, \beta, \gamma, \tau} \sum_{ij} \log P[I_{ij}(\alpha, \beta, \gamma, \tau) = m_{ij}]. \quad (1)$$

Now let I_\emptyset denote an ice random variable, and let $P[M | \emptyset] = \prod_{ij} P[I_\emptyset = m_{ij}]$ be the probability of generating image M without a particle. The *log likelihood ratio* of M is

$$R(M) = L(M) - \log P[M | \emptyset]. \quad (2)$$

Conceptually, our particle picking algorithm evaluates each position (x, y) in a micrograph, forms an image M of the appropriate size centered at (x, y) , and computes $L(M)$ and $R(M)$. The algorithm then picks all M 's with both $L(M)$ and $R(M)$ exceeding manually set, fixed thresholds. The likelihood ratio $R(M)$ is better at distinguishing particles from background noise; however, the simple likelihood $L(M)$ is better at distinguishing particles from high-contrast contaminants. For hemocyanin, we set the $R(M)$ threshold just once and used the $L(M)$ threshold to trade off number versus cleanliness of picks. For p97, the $R(M)$ threshold seems to give lower false positive rate than the $L(M)$ threshold for the same overall number of picks; the danger of picking background noise apparently outweighs the danger of picking dirty particles.

2.2 Algorithm Realization

In practice, searching over all possible choices of image position (x, y) (which implies translation τ) and all possible choices of orientation parameters α, β, γ (quantized to some reasonable amount of precision) would be very slow. To speed up the search for the maximizing orientation parameters, we use a two-stage computation for $L(M)$ and $R(M)$. We precompute many projections of Φ with varying α, β and γ , and group them into a few clusters of similar 2D templates. In the first stage, we generate *candidate picks* by cross-correlation of M with cluster representatives, and in the second stage, we evaluate each candidate pick by computing the actual probabilities given by our model. Figure 6 shows the major steps of the algorithm.

Projecting and Clustering. We start by describing the precomputation, the steps applied to the templates (both 3D and 2D) rather than to the micrographs. Currently these steps require some manual choices, such as the number of projections and the increments for angles α, β , and γ .

For our 3D templates, we used rather high-quality electron density maps computed from 1000 or more particle images. These Φ 's were unnecessarily good, with resolutions of about 16 Å for p97 and 24 Å for hemocyanin. Low-pass filtering of the 2D templates (discussed below) leaves very little power at wavelengths shorter than about 24 Å for p97 and 30 Å for hemocyanin; hence the initial 3D model need not have resolution better than these values, and in practice we found that the 3D model could be substantially worse. We obtained acceptable hemocyanin picking results using a badly formed 3D map, with errors on the order of 60 Å, the result of SPIDER incorrectly assigning the particle's symmetry axis. We imagine that even simple geometric shapes, such as cylinders or ellipsoids of the right size, would be good enough to bootstrap the picking process.

For our 2D p97 templates, we used 60 projections of the p97 3D template, as shown in Figure 7. A set of 60 projections gives rather thin coverage of the space of views, for example we used only $\alpha = 5^\circ, 35^\circ, \text{ and } 95^\circ$; $\beta = 0^\circ$; and $\gamma = 0^\circ, 10^\circ, \dots, 170^\circ$, except for $\alpha = 5^\circ$, for which we used only $\gamma = 0^\circ, 10^\circ, \dots, 50^\circ$. Here we are taking advantage of the C_6 and nearly D_6 symmetry of the p97 particle. For hemocyanin we used 35 projections of the hemocyanin 3D template, 5 top views and 30 side views (every 6°), taking advantage of our knowledge that the particle does not show tilted views. We made our choices of projections intuitively, for example, we picked 6° as the increment for the slopes of hemocyanin rectangle views, judging that 3° would be unnecessarily fine and 10° would be too coarse. We expect that up to 1000 2D templates will be needed for asymmetric particles without preferred orientations; we should not need any more than 1000 (and probably not any more than 200 or 300), because the far-from-focus micrographs have little signal at high spatial frequencies.

We note that EMAN already includes automatic selection of 2D templates; the user specifies only the number of projections, with 60 as a recommended number. We found that 60 EMAN projections of the p97 3D template looked very much like our own intuitive manual choices; 60 projections of hemocyanin looked like our choices along with about 30 tilted views. EMAN chooses projections in a principled way, by computing a large number of projections, uniformly sampled over the space of views, and then ordering them by dissimilarity to previous projections.

After forming the 2D templates, we filtered them to remove high spatial frequencies. Filtering the 2D templates softens the discretization of views to allow each template to represent a patch of angles. Instead of low-pass filtering, it may be better to “super-sample” the patch of angles by averaging a number of templates from the patch.

We grouped filtered 2D templates into clusters, choosing 8 clusters for the 60 p97 templates as shown in Figure 7, and 5 clusters for the 35 hemocyanin templates (not shown). Each cluster is represented by its center, that is, the pixelwise average of its members. Clustering was accomplished using a slightly modified k -means algorithm, with the distance between two images being the L_2 distance (square root of the pixelwise sum of squared differences). We chose L_2 distance so that the cluster center would minimize the sum of distances to the images in the cluster. The modifications to the standard k -means algorithm [Duda, 2001] were as follows. The first cluster seed was chosen by picking an image at random and then picking the farthest-away image as the seed. Then the i -th cluster seed, for $i = 2, 3, \dots, k$ was chosen to be the image with the largest minimum distance to a previously chosen seed. At the end of the clustering, each cluster with radius more than twice the median cluster radius is split in two, and the k -means algorithm is run for a few more iterations.

We note that Fred Sigworth (personal communication) uses an eigenimage decomposition rather than clustering in his work on particle picking with 3D templates. Each of the 2D templates is approximated by a linear combination of a small number of eigenimages; the cross-correlation of an individual template with a micrograph is approximated as a linear combination of the cross-correlations of the eigenimages. This method should give as great a speed-up as clustering, but offer the advantages of precisely quantified approximation and more accurate correlation peak positions.

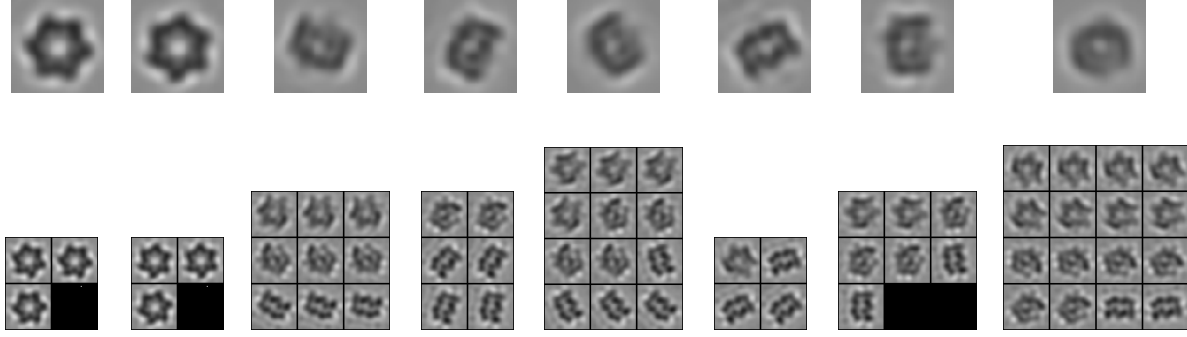


Figure 7: The top row shows the eight cluster centers for p97 particles. Below each cluster center are the individual templates in the cluster.

Filtering Micrographs. The first stage of image comparison will convolve each micrograph with each cluster representative, using Fourier transforms for efficiency. Before carrying out the convolutions, however, we perform some image processing steps on the micrographs. We remove very low spatial frequencies (period longer than about 100 pixels), such as smooth shading across the image, so that different parts of the micrograph have about the same mean gray level, and so that the Fourier transform does not see a discontinuity at the image boundaries. We also remove very high spatial frequencies (period shorter than about 3 pixels).

We stretch the gray scale histogram of each micrograph in order to standardize the micrographs and better use the range of gray values. It turns out to be quite important to do the histogram stretching in a robust way. Stretching that sets the darkest pixel to 0 and the lightest pixel to 255 gives poor results, because some micrographs include very dark contaminants and some do not. What works is to break the micrograph into 8×8 blocks of pixels, and compute the lowest, highest, and average gray level within each block. These values are voted over all blocks to determine the most popular lowest, highest, and average gray levels, L , H , and A . Because the majority of 8×8 blocks in even the most particle- and contaminant-rich micrographs are pure ice, L , H , and A are good fiducial gray values. The gray scale histogram is standardized by rescaling the gray values below A using the linear function that takes the range $[L - 20, A]$ to $[0, 128]$, and rescaling gray values above A using the linear function that takes $[A, H + 20]$ to $[128, 255]$. Gray values that land below 0 or above 255 are truncated. Figure 2 shows the effect of micrograph filtering and gray-scale stretching.

Identifying and Scoring Candidate Picks. Now we convolve each micrograph with each cluster representative. Up to 600 peaks in each convolution image are then identified, but satisfying a constraint that no peak has distance less than say 30 pixels to a previously picked peak in the same image. If we have 8 clusters, we then have up to 4800 identified peaks, each given by its location (x, y) and the index c , $1 \leq c \leq 8$, of the cluster it matched. (A single location (x, y) can match more than one cluster.) Assume cluster c contains individual templates $P^{c1}, P^{c2}, \dots, P^{cn(c)}$, where each P^{ck} is a (filtered) projection of Φ with some setting of angles α , β , and γ .

The second stage then evaluates each cross-correlation peak $((x, y), c)$ by forming an image M of the appropriate size centered on (x, y) and computing the log likelihood and log likelihood ratio of M . Let $I_{ij}(P^{ck})$ denote the random variable for pixel ij implied by the projection P^{ck} , that is, $I_{ij}(P^{ck})$ has p.d.f. pdf- P_{ij}^{ck} (notation as in Figure 5) where P_{ij}^{ck} is the gray level of the ij -th pixel of

P^{ck} . Then we compute the following approximations to the log likelihood and log likelihood ratio:

$$L_c(M) = \max_{1 \leq k \leq n(c)} \log P[I_{ij}(P^{ck}) = m_{ij}], \quad R_c(M) = L_c(M) - \log P[M | \emptyset]. \quad (3)$$

These expressions are only approximations because we are limiting attention to a predefined set of α , β , and γ 's, and fixing the translation τ to be the one that gave the cross-correlation peak. We also tried searching for a translation within a small radius, such as 3 pixels, around the cross-correlation peak, but we found that it gave nearly identical results to the fixed translation.

Thus the inner loop of the computation for equation (3) contains just a table lookup and an addition: for each pixel in template P^{ck} , we look up the log probability of observing the corresponding pixel in M in a table containing the logarithms of p.d.f. values like those shown in Figure 5. We have 128 log p.d.f. tables, each containing 128 entries. Finally, we accept peaks by decreasing order of $L_c(M)$ (on a master list containing $L_c(M)$'s for all the possible cluster numbers c), picking each peak for which both $L_c(M)$ and $R_c(M)$ exceed the thresholds so long as it is not too close to a previous pick, where "too close" is set to about one particle diameter.

Thresholds. We set the thresholds by examining the picked particles for about 10 micrographs from each series, thereby simulating the case of thousands of micrographs, too many for a person to examine. Thresholds turn out to be somewhat a matter of taste. For p97 the picked particles with lowest likelihoods were vague, and we were unsure if they were real or not. For hemocyanin, which is an unusually visible particle, the last picked particles were real particles, but damaged or dirty. One of our criteria in setting the thresholds was to pick approximately as many particles as had been hand-picked by the experts (Mouche for hemocyanin, and Rouiller for p97), so that automatic-pick and manual-pick 3D reconstructions could be compared fairly.

2.3 Relationship with Cross-Correlation

Because the ice histogram is so nearly Gaussian, our model-based method turns out to have an interesting relationship with cross-correlation. The cross-correlation of an observed image $M = \{m_{ij}\}$ with a template $T = \{t_{ij}\}$, each with mean μ , is

$$\sum_{ij} m_{ij} t_{ij} = \sum_{ij} (m_{ij} - \mu)(t_{ij} - \mu) + \sum_{ij} \mu^2. \quad (4)$$

Our interest is in whether or not the cross-correlation exceeds some threshold, so we can ignore the constant term $\sum_{ij} \mu^2$. Rewriting the other term, we see that

$$2 \cdot \sum_{ij} (m_{ij} - \mu)(t_{ij} - \mu) = - \sum_{ij} (m_{ij} - t_{ij})^2 + \sum_{ij} (m_{ij} - \mu)^2 + \sum_{ij} (t_{ij} - \mu)^2. \quad (5)$$

Now say we have a hypothetical image I , that is, an array of random variables I_{ij} from our probabilistic model. Let $T = \{t_{ij}\}$ denote a projection of the 3D map Φ , that is, a 2D template. Assume, however, that we have "grayed down" T so that its gray levels match the centers of the p.d.f.'s in Figure 5. And further assume that the p.d.f.'s in Figure 5 are all Gaussians with standard deviation σ . This assumption is not a bad approximation, since the differences in the variances of the curves pdf-10, pdf-70, and pdf-90 are not terribly large. Under these assumptions, the probability that I_{ij} takes the value m_{ij} is

$$P[I_{ij} = m_{ij}] = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-(m_{ij}-t_{ij})^2 / (2\sigma^2)} \quad (6)$$

and $\log P[I_{ij} = m_{ij}]$ is just $-(m_{ij} - t_{ij})^2 / (2\sigma^2)$, plus a small constant, the same for all ij . Thus we see that the first term on the right-hand side of equation (5) is essentially proportional to the log likelihood in our probabilistic model.

What about the terms $\sum_{ij}(m_{ij} - \mu)^2$ and $\sum_{ij}(t_{ij} - \mu)^2$? The term $\sum_{ij}(t_{ij} - \mu)^2$ is just a fixed per-template offset, so we can ignore it when analyzing the effect of cross-correlating a micrograph with a single template. The term $\sum_{ij}(m_{ij} - \mu)^2$ is the interesting one. Assume the micrograph normalization and high-pass filtering together set the mean gray level of I to $\mu = 64$, the middle of the “ice.hist” p.d.f., and further assume that the “ice.hist” curve is Gaussian with standard deviation σ_\emptyset . Again these are not unrealistic assumptions. Then under our probabilistic model, the probability that I_{ij} takes the value m_{ij} when no particle is present is proportional to

$$P[I_{ij} = m_{ij}] = \frac{1}{\sqrt{2\pi}\sigma_\emptyset} \cdot e^{-(m_{ij}-\mu)^2 / (2\sigma_\emptyset^2)} \quad (7)$$

and $\log P[I_{ij} = m_{ij}]$ is just $-(m_{ij} - \mu)^2 / (2\sigma_\emptyset^2)$, plus a constant. Thus we see that the second term of equation (5) is essentially proportional to the log likelihood of the “null hypothesis” in our probabilistic model, that is, the log likelihood of generating M from pure ice. The constant of proportionality, however, is different than for the first term, with the difference stemming from the different variances of “ice.hist” and the curves such as pdf-70.

Restating the above, cross-correlation is essentially $\sum_{ij}(m_{ij} - \mu)^2 - \sum_{ij}(m_{ij} - t_{ij})^2$; our log likelihood is approximately $-\sum_{ij}(m_{ij} - t_{ij})^2$; and our log likelihood ratio is approximately $\alpha \cdot \sum_{ij}(m_{ij} - \mu)^2 - \sum_{ij}(m_{ij} - t_{ij})^2$, for some constant α which depends upon the parameter *ice-ratio*. With *ice-ratio* set to 50% ice, α is about .6. With its greater emphasis on the power term $\sum_{ij}(m_{ij} - \mu)^2$, cross-correlation picks high-contrast contaminants as well as true particles. Doubling the contrast of a candidate particle doubles its cross-correlation, but would typically decrease its log likelihood ratio under our model. Log likelihood leans too far in the other direction, sometimes picking gray background as well as true particles. (Correlation coefficient—dividing by a power term—would have the same flaw.) Log likelihood ratio seems to be the happy medium.

Finally, we should mention that $\alpha \cdot \sum_{ij}(m_{ij} - \mu)^2 - \sum_{ij}(m_{ij} - t_{ij})^2$ can be computed using Fast Fourier Transforms and an efficient local variance calculation [van Heel, 1982]. Hence if this approximation were substituted for our log likelihood ratio, we would not have to limit the probabilistic scoring to translations given by cross-correlation peaks in order to achieve fast running time.

3 Results

Figures 1, 2 and 3 show results. Exactly the same programs were used for p97 and hemocyanin, with the only choices being the set of projections, the number of clusters, and the two likelihood thresholds.

On Figure 1 our automatic picker picked many more genuine p97 particles than an expert (Rouiller) using SPIDER’s automatic picker followed by manual screening. Over all 17 micrographs, however, the disparity was—as intended—much smaller, our program made 617 picks and Rouiller made 561. The number of semi-automatic picks was limited by SPIDER’s accuracy; a threshold that found all the real particles would find so many false particles that the manual screening would be very tedious.

EMAN discarded 50 of our 617 automatic picks and 88 of the 561 semi-automatic picks, because they did not align to a consistent center; these numbers give an indication of the false positive rate. Both data sets gave 3D reconstructions of resolution 22.8 Å, as defined by the .5 threshold of the Fourier shell correlation. As shown in Figure 8, both reconstructions were in good agreement with a

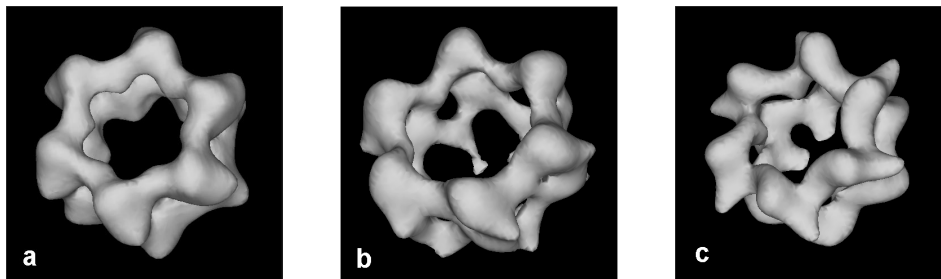


Figure 8: Isosurfaces of p97 density maps computed from (a) 617 automatic picks, (b) 561 semi-automatic picks, and (c) more than 4000 semi-automatic and manual picks. The resolutions of (a) and (b) are both 22.8 Å as determined by the .5 threshold of the Fourier shell correlation.

high-quality reconstruction, with the semi-automatic reconstruction (b) showing a little better visual similarity than the automatic reconstruction (a). (The electron density maps are actually in closer agreement than the isosurfaces; features such as the “spurs” in the back center of (c) are present in (a) but fall below the isosurface value.) Why did 473 semi-automatic picks give a reconstruction just as good as—or even slightly better than—the reconstruction from 567 automatic picks? The answer is that the semi-automatic picks had more uniform coverage of the sphere of views, with relatively fewer of the “saw blade” top views (which can be easily confused with bottom views) and more of the harder-to-see tilted views.

We also visually compared our p97 particle picking results to those from EMAN’s particle picker BOXER, which we trained and ran on the same micrograph (thereby helping EMAN). Using 2D templates computed from the micrograph shown in Figure 1, BOXER gave poor results, well under 50% accuracy. Using the initial 3D map to make 60 2D templates, BOXER gave good results, comparable to our own results but with a few more picks in the contaminant region. The contrast threshold settings that worked for this micrograph, however, did not work well with other micrographs; for example, in the next micrograph (alphabetically by name of the file) BOXER’s performance slipped to about 13 correct picks and 5 false positives, where our picker had about 40 correct picks and 5 false positives.

Figure 3 shows a hemocyanin example. For hemocyanin, a person’s picks can be considered ground truth, so we can give precise quantitative results from the 2D images alone. We counted statistics for five categories: perfect picks (clean particles of the right kind), imperfect picks (damaged, dirty, or variant particles), bogus picks (background noise or contaminant), perfect misses (clean particles not picked), and imperfect misses (damaged, dirty, or variant particles not picked). Particles partially off the micrograph cannot be used and hence were not counted at all. Using a relaxed likelihood threshold (not much discrimination by quality), our picker had 33 perfect picks, 8 imperfect picks, 0 bogus picks, 3 perfect misses, and 3 imperfect misses on Figure 3. Two of the 3 perfect misses seem to result from abutting particles reducing likelihood scores. We compared our results to results from BOXER, which we trained and ran on the same micrograph. Using 2D templates computed from the micrograph, BOXER gave 31 perfect picks, 9 imperfect picks, 9 bogus picks, 4 perfect misses, and 2 imperfect misses. Using the initial 3D map to make 60 2D templates, BOXER improved to 27 perfect picks, 8 imperfect picks, 1 bogus pick, 8 perfect misses, and 3 imperfect misses. One difficulty when using BOXER was to find contrast thresholds that simultaneously worked for both rectangle and circle views.

Over all 82 hemocyanin micrographs our particle picker gave 1979 perfect picks, 285 imperfect

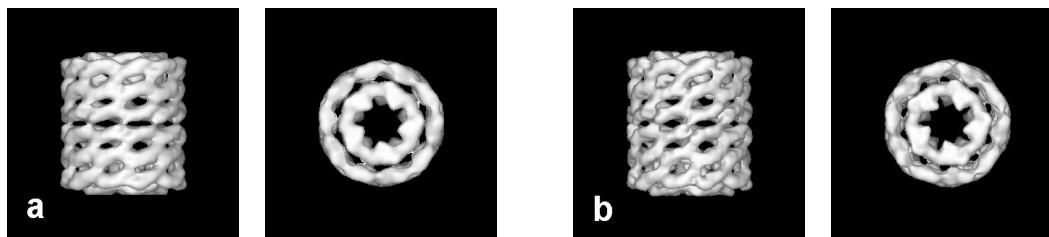


Figure 9: Hemocyanin density maps (side and top views of isosurfaces) computed from (a) 1157 automatic picks and (b) 1042 manual picks. The resolution of (a) is 23.3 Å and that of (b) is 23.4 Å.

picks, 7 bogus picks, 76 perfect misses, and 227 imperfect misses. Restating these results another way, our automatic picker had 4% false negative rate (perfect misses) and 0.3% false positives (bogus picks). It picked 56% of the imperfect particles, but only 13% of all picks were imperfect.

For 3D reconstruction we raised the likelihood threshold in order to reduce the number of imperfect picks, and raised it still higher for the circle views (recognized by which cluster they matched), because these views represent only a single point on the sphere of viewing directions. We then had 1024 rectangle views and 133 circle views, less than half the original number of picks. Figure 9 compares a 3D reconstruction made from these automatic picks with a 3D reconstruction made from 1042 manual picks, of which only 10–20 were circle views. These reconstructions, made by SPIDER, are virtually identical: the .5 threshold of the Fourier shell correlation gives a resolution of 23.4 Å for the manual-pick reconstruction and 23.3 Å for the automatic-pick reconstruction.

Running time is dominated by the Fourier transforms for the cross-correlation. The number of transforms is $n + m + nm$, where n is the number of micrographs and m is the number of template clusters. Using 1024×1024 transforms, we obtained times of about 2 minutes per micrograph; with 512×512 transforms times should be well under a minute. We estimate that running times would grow to 5 minutes per micrograph with 1000 templates in 50 clusters.

4 Discussion

We believe that our particle picker is usable already, and should give near human performance level for any fairly round particle. We do, however, have some plans for further improvement. One idea is to use irregularly shaped templates, rather than only square templates; this should improve the performance on abutting particles and also enable the system to pick longer, skinnier particles, although our approach probably would still not be the method of choice for filamentous particles. We should be able to determine a particle boundary automatically in the 3D map, and use this boundary to define the irregularly shaped templates. Another possibility is to stay with square templates, but include variances for the pixel p.d.f.’s in order to weight some pixels more heavily than others. We would also like to explore quality measures for images in order to screen out dirty, distorted, and variant particles; we think that after the likelihood or likelihood ratio indicates that a particle is real, a different measure—perhaps computed from the close-to-focus micrograph—would give a more sensitive measure of quality.

Aside from leading to practical tools, model-based methods bring another benefit. By making the assumptions underlying an algorithm explicit, these methods enable more informed analysis of algorithms. As we saw in Section 2.3, cross-correlation is appropriate for a model with i.i.d. additive Gaussian noise, an unrealistic assumption for micrographs with contaminants. Our model also as-

sumes i.i.d. additive noise, but with lower noise overtop of desirable particles than in other sections of the micrograph; this slightly more realistic model gives much greater contaminant rejection. Local correlation as in [Roseman, 2003] corrects for local mean and variance and hence fits a model with multiplicative noise, a common situation in macroscopic scenes with locally varying illumination, but not in EM micrographs.

This analysis shows that the least developed part of our method is the noise model. In fact, compared to other computer vision problems such as face detection and recognition, we already have an excellent signal model. One interesting direction for future work would be to train a noise model (for example, one in which the random variables are wavelet or spectral coefficients) automatically by feeding the results of 3D reconstruction back into the particle picker. Picked particles whose close-to-focus partners give poor correlations with the 3D model could serve as negative training examples.

Acknowledgements

We thank Fred Sigworth for sharing ideas. We gratefully acknowledge the help of Bridget Carragher, Clint Potter, Yuanxin Zhu, and Ron Milligan at the Scripps Research Institute, and David Goldberg, David Fleet, and Richard Bruce at the Palo Alto Research Center. Some of the work presented here 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 grant RR17573).

References

- [Abbott, 2002] A. Abbott. The society of proteins. *Nature* 417, 27 June 2002, 894–896.
- [Baker, 1996] T.S. Baker and R.H. Cheng. A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J. Struct. Biol.* 116 (1996), 120–130.
- [Carragher, 2000] B. Carragher, N. Kisseberth, D. Kriegman, R.A. Milligan, C.S. Potter, J. Pulokas, and A. Reilein. Leginon: an automated system for acquisition of images from vitreous ice specimens. *J. Struct. Biol.* 132 (2000), 33–45.
- [Dubochet, 1988] J. Dubochet, M. Adrian, J.J. Chang, J.C. Homo, J. Lepault, A.W. McDowell, and P. Schultz. Cryo-electron microscopy of vitrified specimens. *Quarterly Review Biophysics* 21 (1988), 129–228.
- [Duda, 2001] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley & Sons, 2nd edition, 2001.
- [Frank, 1996] J. Frank. *Three-dimensional electron microscopy of macromolecular assemblies*. Academic Press, 1996.
- [Frank, 2000] J. Frank, P. Penczek, R.K. Agrawal, R.A. Grassucci, and A.B. Heagle. Three-dimensional cryo-electron microscopy of ribosomes. *Methods Enzymol.* 317 (2000), 276–291.
- [Frank, 1984] J. Frank and T. Wagenknecht. Automatic selection of molecular images from electron micrographs. *Ultramicroscopy* 12 (1984), 169–176.

- [Frank, 1995] J. Frank, J. Zhu, P. Penczek, Y. Li, S. Srivastava, A. Verschoor, M. Radermacher, R. Grassucci, R.K. Lata, and R.K. Agrawal. A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature* 376 (1995), 441–444.
- [Frank, 1996] J. Frank, M. Radermacher, P. Penczek, J. Zhu, Y. Li, M. Ladjadj, and A. Leith. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* 116 (1996), 190–199. http://www.wadsworth.org/spider_doc/spider/docs/
- [Gabashvili, 2000] I.S. Gabashvili, R.K. Agrawal, C.M.T. Spahn, R.A. Grassucci, D.I. Svergun, J. Frank, and P. Penczek. Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* 100 (2000), 537–549.
- [Glaeser, 1999] R.M. Glaeser. Review: Electron crystallography: Present excitement, a nod to the past, anticipating the future. *J. Struct. Biol.* 128 (1999), 3–14.
- [Henderson, 1995] R. Henderson. The potential and limitations of neutrons, electrons and x-rays for atomic resolution microscopy of unstained biological macromolecules. *Quart. Rev. Biophys.* 28 (1995), 171–193.
- [Kivioja, 2000] T. Kivioja, J. Ravantti, A. Berkovsky, E. Ukkonen, and D. Bamford. Local average intensity-based method for identifying spherical particles in electron micrographs. *J. Struct. Biol.* 131 (2000), 126–134.
- [Lata, 1995] K.R. Lata, P. Penczek, and J. Frank. Automatic particle picking from electron micrographs. *Ultramicroscopy* 58 (1995), 381–391.
- [Ludtke, 1999] S.J. Ludtke, P.R. Baldwin, and W. Chiu. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* 128 (1999), 82–97.
- [Nicholson, 2001] W.V. Nicholson and R.M. Glaeser. Review: Automatic particle detection in electron microscopy. *J. Struct. Biol.* 133 (2001) 90–101. <http://www.lbl.gov/Publications/LDRD/2000/LSD.html>
- [Ogura, 2001] T. Ogura and C. Sato. An automatic particle pickup method using a neural network applicable to low-contrast electron micrographs. *J. Struct. Biol.* 136 (2001) 227–238.
- [Orlova, 1997] E.V. Orlova, P. Dube, J.R. Harris, E. Beckman, F. Zemlin, and J. Markl. Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *J. Molecular Biology* 271 (1997) 417–437.
- [Roseman, 2003] A.M. Roseman. Particle finding in electron micrographs using a fast local correlation algorithm. *Ultramicroscopy* 94 (2003), 225–236.
- [Rouiller, 2000] I. Rouiller, V.M. Butel, M. Latterich, R.A. Milligan, E.M. Wilson-Kubalek. A major conformational change in p97 AAA ATPase upon ATP binding. *Molecular Cell*, 6 (2000), 1485–1490.

- [Rouiller, 2001] I. Rouiller, J. Pulokas, V.M. Butel, R.A. Milligan, E.M. Wilson-Kubalek, C.S. Potter, and B. Carragher. Automated image acquisition for single particle reconstruction using p97 as the biological sample. *J. Struct. Biol.* 133 (2001), 102–107.
- [Rouiller, 2002] I. Rouiller, B. DeLaBarre, A.P. May, W.I. Weis, A.T. Brunger, R. Milligan, E.M. Wilson-Kubalek. Conformational changes of the multifunction p97 AAA ATPase during its ATPase cycle. *Nature Struct. Biol.* 9 (2002), 950–957.
- [Saad, 1998] A. Saad, W. Chiu, and P. Thuman-Commike. Multiresolution approach to automatic detection of spherical particles from electron cryomicroscopy images. *Int. Conf. Image Processing*, 1998. <http://www.qedlabs.com/pam/Research/Icosahedral/ICIP98.htm>
- [Sigworth, 2002] F. Sigworth. Personal communication.
- [Stoschek, 1997] A. Stoschek and R. Hegerl. Automated detection of macromolecules from electron micrographs using advanced filter techniques. *J. Microscopy* 185 (1997), 76–94.
- [Thuman, 1995] P. Thuman-Commike and W. Chi. Automatic detection of spherical particles in spot-scan electron cryomicroscopy images. *J. Microscop. Soc. Am.* 1 (1995), 181–201.
- [van Heel, 2000] M. van Heel, B. Gowen, R. Matadeen, E.V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, and A. Patwardhan. Single-particle electron microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics* 33:4 (2000), 307–369.
- [van Heel, 1982] M. van Heel. Detection of objects in quantum noise limited images. *Ultramicroscopy* 8 (1982), 331–342.
- [Zhu, 2001] Y. Zhu, B. Carragher, D. Kriegman, and C.S. Potter. Automated identification of filaments in cryo-electron microscopy images. *J. Struct. Biol.* 135 (2001), 302–312.
- [Zhu, 2002] Y. Zhu, B. Carragher, F. Mouche, and C.S. Potter. Automatic particle detection through efficient Hough transforms. Preliminary version in *IEEE International Symposium on Biomedical Imaging*, July 2002.