ELSEVIER

Particle picking by segmentation: A comparative study with SPIDER-based manual particle picking

Umesh Adiga [a,*], William T. Baxter [b], Richard J. Hall [a], Beate Rockel [c], Bimal K. Rath [b], Joachim Frank [b,d], Robert Glaeser [a,e]

[a] *Physical Biosciences Division, Lawrence Berkeley National Laboratory, 1, Cyclotron Road, Berkeley, CA 94720, USA*
[b] *Wadsworth Center, Health Research Inc., Empire State Plaza, Albany, NY 12201-0509, USA*
[c] *Department of Molecular Structural Biology, Max-Planck-Institute for Biochemistry, Martinsried, Germany*
[d] *Howard Hughes Medical Institute at Wadsworth Center, Department of Biomedical Sciences, University at Albany, SUNY, Empire State Plaza, Albany, NY 12201-0509, USA*
[e] *Department of Molecular and Cell Biology, University of California at Berkeley, CA 94720, USA*

Abstract

Boxing hundreds of thousands of particles in low-dose electron micrographs is one of the major bottle-necks in advancing toward achieving atomic resolution reconstructions of biological macromolecules. We have shown that a combination of pre-processing operations and segmentation can be used as an effective, automatic tool for identifying and boxing single-particle images. This paper provides a brief description of how this method has been applied to a large data set of micrographs of ice-embedded ribosomes, including a comparative analysis of the efficiency of the method. Some results on processing micrographs of tripeptidyl peptidase II particles are also shown. In both cases, we have achieved our goal of selecting at least 80% of the particles that an expert would select with less than 10% false positives.
© 2005 Published by Elsevier Inc.

1. Introduction

To minimize radiation damage to ice-embedded biological specimens, electron micrographs have to be acquired at low electron dose, which results in a low signal-to-noise ratio (SNR) (Glaeser, 1971). Single-particle reconstruction (Radermacher et al., 1987; see Frank, 1996) is based on the premise that the information necessary to reconstruct a biological particle (a molecule) can be collected from different images showing different "copies" of the molecule in different views. To improve the SNR in the data, a large number of images of particles in many orientations must then be merged to produce a three-dimensional (3-D) den-sity map. When images of currently available quality are used, it is believed that images of as many as one million particles are required to reconstruct a protein-molecule at an "atomic" resolution (Henderson, 1995; Frank, 2002; Sali et al., 2003).

Picking hundreds of thousands of particles is a significant bottle-neck in the process of single-particle reconstruction at high resolution (Glaeser, 2004; Nicholson and Glaeser, 2001). As a result, there are several recent papers describing automatic particle picking techniques (Huang and Penczek, 2004; Rath and Frank, 2004; Roseman, 2004; Sigworth, 2004; Volkmann, 2004). The number of false positives may, however, still exceed 30% when existing methods are applied to micrographs of routinely used quality (Zhu et al., 2004).

In the course of a single particle reconstruction, it is a general practice that the particle images are boxed along

* Corresponding author. Fax: +1 510 486 6488.
 E-mail address: UPAdiga@lbl.gov (U. Adiga).

with a considerable background area around the particle. Information from this background is required to normalize the data (Boisset et al., 1993). In micrographs that contain a high density of particles, this practice results in the appearance of partial signatures of one or more neighboring particles that intrude into the boxed area. Boxes that contain intruding particles are routinely rejected for further processing, as the partial signatures of the neighboring particles are believed to interfere with the alignment process as well as the normalization. This rejection of particles reduces the number of useful particles that can be picked from any one micrograph.

In an earlier paper addressing the particle-picking problem, we have proposed a method that uses a combination of steps that involve pre-processing, segmentation and feature analysis (Adiga et al., 2004). We now demonstrate that the segmentation approach can help in detecting partial signatures of the neighboring particles in a boxed single-particle image. Segmentation thus allows the option of replacing partial signatures with a representative background texture patch rather than eliminating such particle images from the data set. To accomplish this, we have introduced a novel post-processing operation called shrink-wrapping that will center the particles in a box area and facilitate the elimination of intruding particle signatures. The same operation can be further used to create a binary mask of the particle that may prove to be useful for quality indexing, pre-classification of particles, etc. The method is carefully designed to avoid invalidating the statistical signature required for normalization.

This paper also describes improvements made over the initially proposed segmentation method as well as a generalization of the software so that it can be used to pick non-convex particles. We have increased the robustness of the pre-processing and segmentation operations by making them more adaptive to histogram-feature variations in the micrograph. The current paper reports a large-scale comparative analysis of more than one hundred and thirty thousand ribosome particles in 55 micrographs acquired at Wadsworth Center, Albany. Some preliminary results on non-convex shaped, TPP-II particles are also shown.

2. Method

Our approach to particle picking, based on segmentation, follows conventional image-analysis steps of pre-processing followed by segmentation and post-processing. The control flow-diagram shown in Fig. 1 summarizes the segmentation-based software design.

2.1. Pre-processing

The pre-processing steps are designed to increase the particle contrast, smooth the fine texture of the background, and remove most of the low-frequency intensity variations from the micrographs. These goals are achieved by a combination of contrast stretching, partial diffusion of
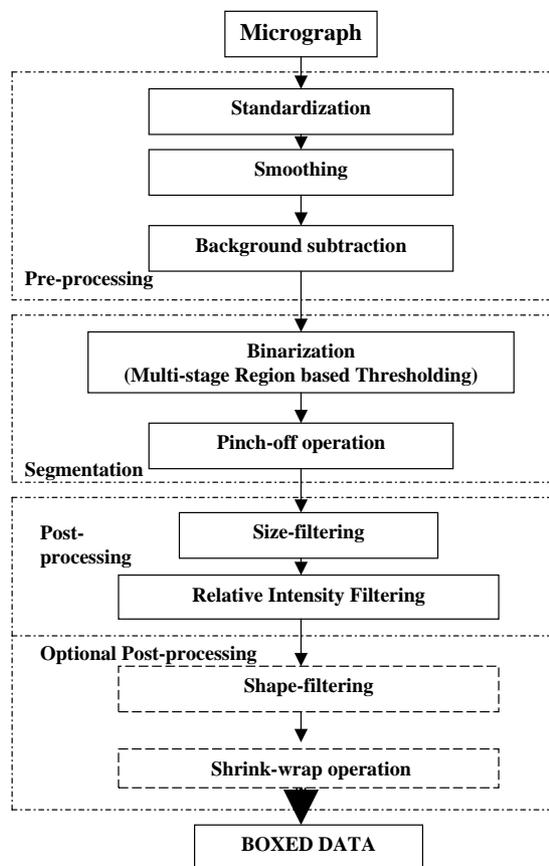

Fig. 1. Control flow diagram of the particle-picking software.

the pixel intensities and a background-subtraction process. All these processing steps are used to identify the locations of putative particles, after which we revert to the original data to extract the particle images that are to be used for particle reconstruction.

A non-linear contrast-stretching operation performed as described previously (Adiga et al., 2004), first redistributes the gray-scale of the pixels that fall within an adaptively selected gray-scale range. This process results in a similar (i.e., standardized) intensity distribution for all the micrographs with an approximately common mean. This initial image standardization step is essential to avoid retuning of control parameters for the steps further down the processing pipeline. Figs. 2A and 3A show a small part of representative micrographs of ribosomes and TPP-II particles, respectively, before contrast-stretching, and Figs. 2B and 3B show the same micrographs after contrast-stretching.

Difficulty in reducing the noise and the associated granular texture in cryo-micrographs is a major factor hindering the automation of particle identification. The goal of the noise-reduction process is to obtain as much as possible a smooth background while retaining good contrast between the background and the particles. Use of non-linear anisotropic filters such as the one defined in Eq. (1) has been proved to be effective in achieving the desired reduction in noise (Adiga et al., 2004). The non-linear smoothing filter in its explicit reaction-diffusion form is written as
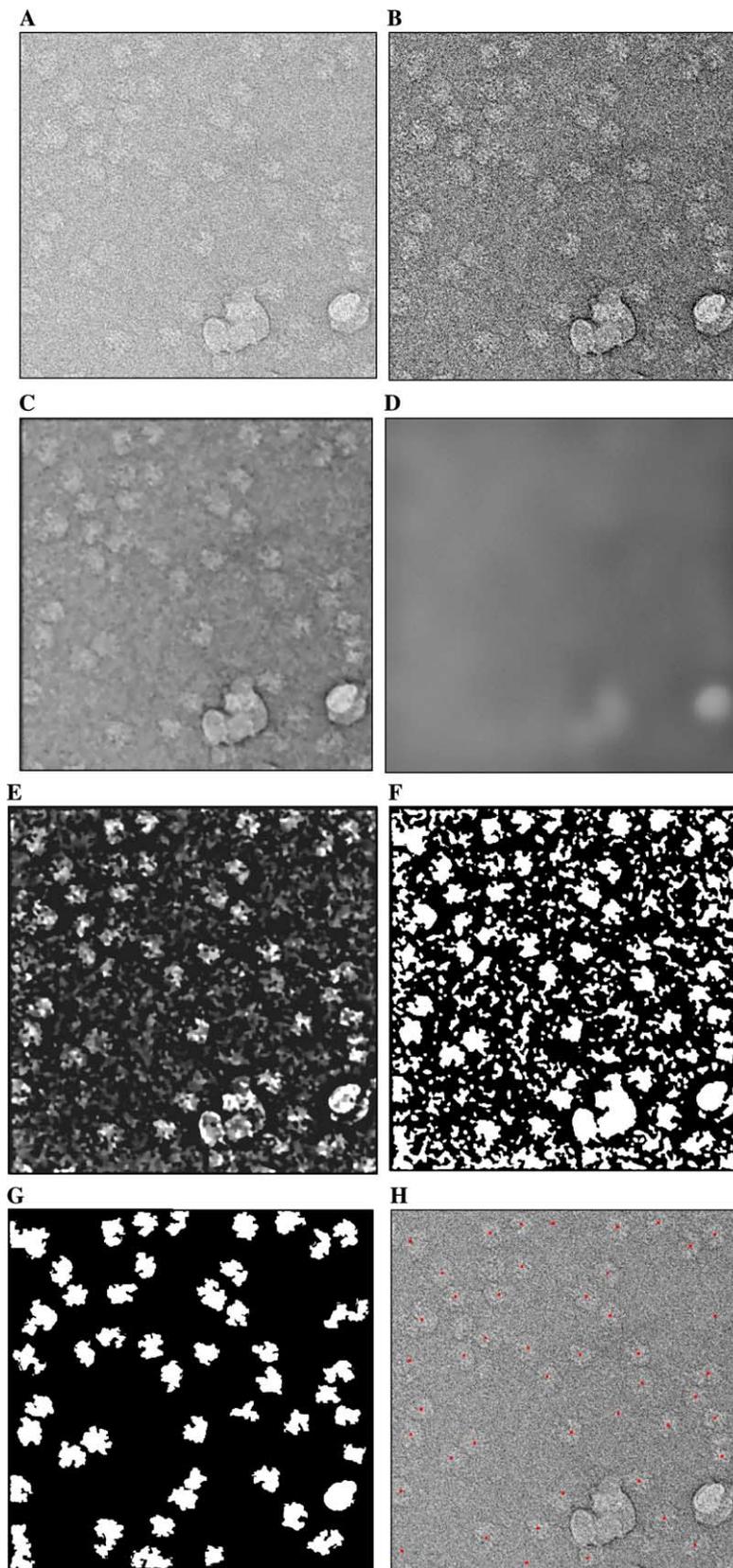
Fig. 2. Results obtained at intermediate steps of processing for picking ribosome particles by segmentation. (A) A small part of an original micrograph. (B) Contrast-enhanced micrograph. (C) After reaction-diffusion filtering. (D) Background image of the micrograph in (C). (E) Background-subtracted and rescaled micrograph. (F) Result of binarization and hole-filling. (G) Result of size filtering. (H) Result of boxing the particles after removing the brightness-saturated particles.
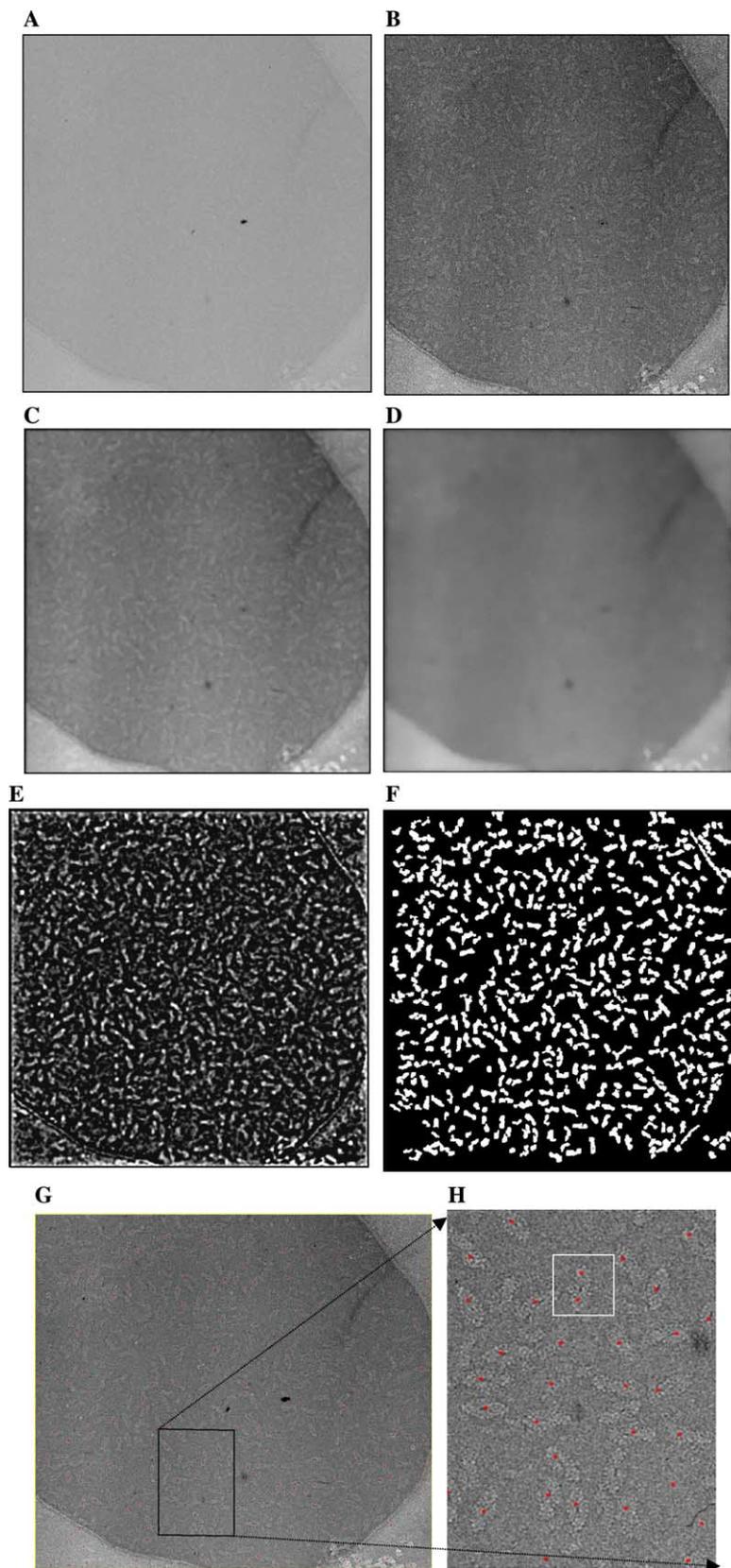
Fig. 3. Results obtained at intermediate steps of processing for picking TPP-II particles by segmentation. (A) Original micrograph. (B) Contrast-enhanced micrograph. (C) After reaction-diffusion filtering. (D) Background image of the micrograph in (C). (E) Background-subtracted and rescaled micrograph. (F) Result of binarization, hole-filling and size-filtering. (G) Result of picking the particles (shown by placing a 'red' dot over the particles). (H) A small area is magnified for a better view. The square box indicates a navette-shape particle that was picked twice due to the fact that its high-intensity ends are similar to two distinct end-views/top-views.(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

$$\frac{\partial I}{\partial t} = (\sin \beta) \cdot h \cdot \nabla^2 I + (\cos \beta) \cdot \nabla h \cdot \nabla I, \qquad (1)$$

where, $h = 1/(1 + (\frac{\partial I}{\partial x})^2 + (\frac{\partial I}{\partial y})^2)$ is an image intensity gradient operator known as an *edge indicator function*. The first term, $\{(\sin \beta) \cdot h \cdot \nabla^2 I\}$ is a diffusion term responsible for smoothing, while $\{(\cos \beta) \cdot \nabla h \cdot \nabla I\}$ is responsible for edge-enhancement. The parameter 'β' determines the relative contribution of diffusion and enhancement terms to the process of non-linear smoothing. Figs. 2C and 3C show the result of smoothing the micrographs for 100 time steps, using the above described filter with $\beta = 65°$. The size of the time step is $\Delta t = \frac{1}{2}(\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2})$, where $\Delta x = \Delta y = 1$, defines the distance between pixels in the 'x' and 'y' direction (Malladi and Ravve, 2002).

To reduce the intensity variation in the background, a background image is first reconstructed by carrying out a higher degree of diffusion of the micrographs, as shown in Figs. 2D and 3D. This "unsharp mask" is then subtracted from the previously smoothed micrograph. The result of background subtraction is shown in Figs. 2E and 3E.

2.2. Segmentation

The process of segmentation that identifies and extracts the particles consists of two stages, region-based binarization and pinch-off filtering. The region-based adaptive-binarization operation is designed to convert the initial gray-tone image into a two-tone one. The algorithm is as follows.

- Threshold the pre-processed micrograph at its mean intensity.
- Fill the holes within the foreground of the binary image. A hole is defined as any isolated background component whose size is less than 25% of the particle size. The particle size is itself an input parameter that defines the approximate size of the particle in pixel units.
- Label the components in the binary image using a connected-component labeling algorithm (Jain et al., 1995).
- Calculate a new threshold value for each labeled region by regional histogram analysis. The new regional threshold value for each component corresponds to the gray value of the point in the respective histograms that is less than the gray value of the mode point, where, in addition, the curvature of the histogram is maximum. The curvature 'κ' of a plane curve in the form of $y = f(x)$ can be calculated as $\kappa = \frac{\frac{d^2 y}{dx^2}}{(1 + (\frac{dy}{dx})^2)^{3/2}}$. The derivatives were calculated as finite differences.
- Threshold each component, based on the regional thresholds calculated as described above.
- Fill the holes, if any, created by this second stage of thresholding. Fig. 2F shows the final result of binarization and hole-filling of a pre-processed micrograph.

There is a possibility that some of the closely located particles, when binarized, are found to touch one another through frail connections. When the objects are labeled, such connected objects are given the same label. This generally results in rejection of the whole cluster of two or more particles. To avoid the rejection of the particles in such clusters, we have segmented them by a method of ultimate erosion and conditional dilation. All the components in the binary image whose sizes are below 25% of the pre-defined particle-size are considered to be artifacts and are first removed from further processing. The individual objects in the binary image are then eroded until the size of the eroded signatures falls below 50% of the predefined particle size. The object signatures are then dilated under the constraint that no two dilated signatures should touch one another, and the pixels of the dilated signatures do not occupy the background area defined by the original binary image, e.g., Fig. 2F. This operation eliminates the frail connections between closely located objects, and hence it is called a pinch-off operation.

In the next step, all the objects are component-labeled again, and those with size greater than the pre-defined particle size are eliminated from further processing. Figs. 2G and 3F show the result of segmentation and size-filtering of a ribosome micrograph and a TPP-II micrograph, respectively.

For each component in the segmented image, the weighted centroid of that component is calculated. The pixel intensity in the contrast enhanced image was used as the weight. Such a weighted centroid provides the coordinates of an approximate center of the particle. Figs. 2H, 3G and H shows the result of extracting the particles by placing a 'RED' dot over the particles.

2.3. Post-processing

Additional post-processing filters are required when images of non-convexly shaped particles are processed. If the particles have a very distinct shape, for example, tobacco mosaic virus (TMV) particles (Zhu et al., 2001), it is possible to include shape-based features to eliminate the false positives. Shape can be quantified by computing a measure of the similarity between a particle and a template (using, for example, the correlation coefficient), or by computing a metric of the geometry of the binary mask of a particle as demonstrated for an application to TPP-II particles in this paper.

It is a general practice to discard boxed particles in a manual-selection (editing) phase, if portions of adjacent particles intrude into the boxed area. Some techniques discard a particle when it is located within a preset distance from an already selected particle (Rath and Frank, 2004). As an alternative option to rejecting particles that are close neighbors, we have implemented a heuristic operation that recognizes the intruding particle within a boxed image and replaces the intruder by a background-texture pattern. Fol-

lowing that, the operation also approximately re-centers the particle within a selected box, since the changes in the shape of a particle due to use of the shrink-wrap operation, described below, also change the location of its center of mass. The algorithm to do the shrink-wrap operation is illustrated in Fig. 4 and it involves the following steps.

- For each particle, define a box that is approximately twice the size of the particle. The particle size is defined as an area occupied by the particle in the image in pixel units.
- Component-label all the objects within the box using the information from the segmented binary micrograph. Figs. 4B1 and B2 show examples of such boxed images from the binary micrograph.
- Using the binary image as an initial model of the particle mask, apply the iterative dilate and shrink-wrap operation described below.
- Remove the intruding particle signatures within the box. An intruding particle signature is any signature inside the box that does not occupy the center of the box.

Each step of shrink-wrapping consists of dilating the object into a predefined neighborhood area within the box followed by erosion of individual boundary pixels which have gray-values below the average intensity of the dilated object. The gray-scale image obtained after pre-processing is used for calculating the average intensity of the object and the intensity of the boundary pixels of the dilated object. The process of eroding the object's boundary pixels stops when no more surface pixels can be eroded.

The process of 'dilate and shrink-wrap' is repeated until the change in size of the object between two consecutive iterations of dilate and shrink-wrap is zero. If the objects are bloated beyond their actual boundary before the start of the shrink-wrap process, such objects are simply shrunk back to their shape and no change in the shape and size takes place in the subsequent iterations. The advantage of this shrink-wrapping process is that the shrinking takes place based on the intensity characteristics of the individual objects.

Fig. 4 shows an example of this post-processing step in centering the particles and removing the intruding noisy signature. It also shows the binary image of the particle obtained as the result of shrink-wrap filtering. It is possible that this binary image can be used in feature-vector calculation and pre-classification of the particles.
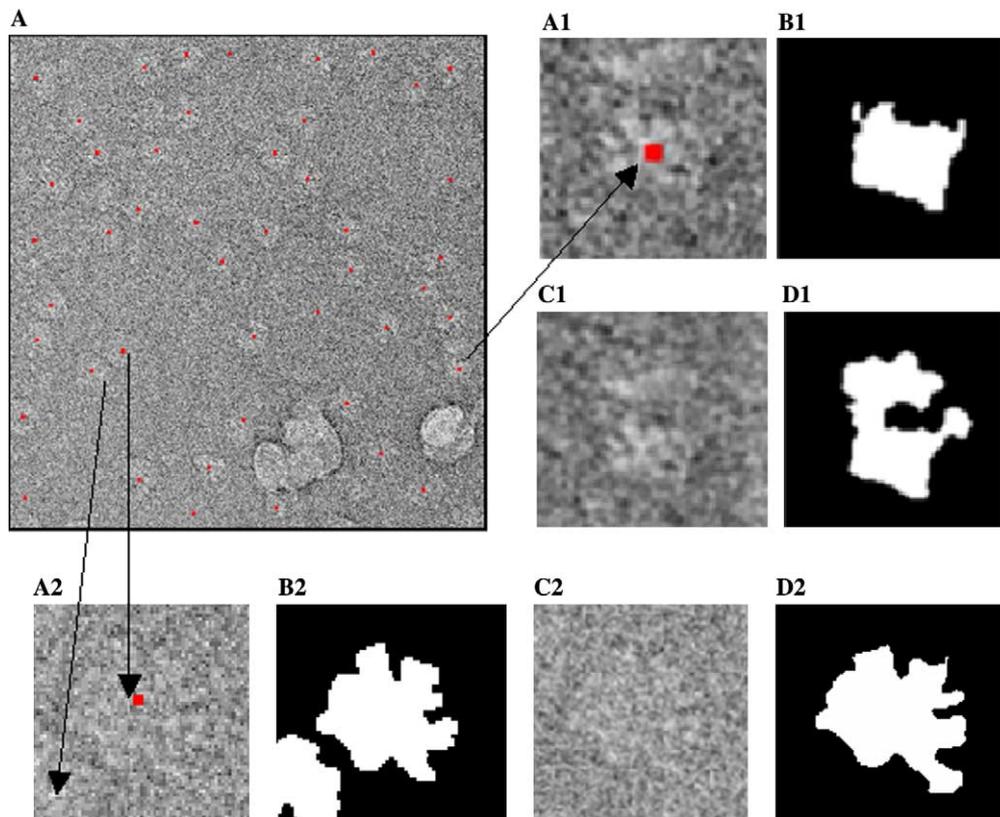

Fig. 4. Illustration of advantages of shrink-wrap operation. (A) An example of a particle that would be eliminated in normal circumstances because of the particle truncation or the intrusion of a neighboring particle. (A1) A boxed particle image with a partial representation of the particle, i.e., particle is cut-off at the top. The particle is also off-centered. (B1) Binary mask of the particle in (A1) as extracted from the binarized micrograph shown in Fig. 2G. (C1) After shrink-wrap operation this particle is completely represented and centered. (D1) Binary mask of the particle in (C1). (A2) Another example with an intruding particle within the box area. (B2) Binary mask of particles within the box shows an intruding particle. (C2) After replacing the intruding particle by a background texture, followed by re-centering the particle after the shrink-wrap operation. (D2) Binary mask of the particle in (C2).

2.4. Implementation

The software was developed in IDL language with required Graphical User Interface (GUI). The C language version of the software was also implemented but without the GUI facility. The derivatives in the reaction-diffusion filter and the curvature calculation were implemented using the difference equation of the form $\frac{\partial I}{\partial x} = \frac{I_{x+1} - I_x}{\Delta x}$.

The particle-picking program can be set in a batch mode and run overnight on a set of micrographs of similar quality. The software takes about 10 min on a single microprocessor machine to complete the particle picking over a micrograph of an approximate size $6K \times 6K$.

We have tested the program on Linux 32bit and 64bit operating systems. All the testing and experiments were conducted on images where data were stored in little-endian order. The data in big-endian order should be converted to little-endian order before using this particle-picking program. The endian order can be changed by 'em2em.e' program freely available at http://www.imagescience.de/em2em/. At present, the segmentation-based software does not allow the user to interact once it has started running, but we plan to provide interactive controls for the user as an option in the next version of the program.

Both IDL and C language versions of the program together with the user-manual for tuning a small set of parameters can be downloaded from the website http://macro-em.org/particle-selection. We will periodically update the program with improved versions at this site.

3. Results and discussion

Several improvements have been made over the earlier version of this software. The pre-processing technique is now structured into independent functions and most tuning parameters adapt to the image statistics. Segmentation is restricted to the binarization and pinch-off filtering steps only. The post-processing operations read their parameters from the particle-size and the image statistics. The approximate particle-size, in terms of number of pixels, is provided as an input parameter. Several additional, post-processing filters are designed as optional features to reduce false positives, to center the particles, and to obtain an approximate binary mask of the particles. In general, only a smoothing-level parameter and an approximate particle-size is required to be tuned by the user for a representative micrograph in a data set, and the rest of the parameters are either independent of the variations in the data or they are adaptively calculated based on image statistics.

The number of parameters to be tuned in the software is one of the issues in defining the utility of the software. A few user-set parameters can be tolerated if they are required to be set on only one or two training images, and the rest of the images can be processed automatically with the same set of parameter values. In the work reported here, parameters are set based on experiments with one micrograph. In the segmentation-based technique, the set of parameters include one that defines the number of iterations of contrast stretching; a smoothing-filter parameter β; the number of time-steps for smoothing; the amount of diffusion used for construction of the background-image; and the particle-size. The number of iterations for contrast stretching is set based on the visual improvement in the contrast of the particles in the micrograph. This was set to a default value '1.'

In our experiments, the algorithms were found to be robust against small changes in the parameter values. For example, in the case of ribosome-particle picking, a 10% increase in the number of time steps for diffusion resulted in less than a 1% change in the particle selection efficiency. Excessive diffusion causes an obvious loss of particles, while insufficient smoothing results in an increase in the number of false positives. Similarly, variation of β in the range of 65–85° did not show any significant changes in the particle picking efficiency. A further decrease in β results in enhancement of noisy structures along with true particles, which results in increased false positives. The particle size is not intended to be a tuning parameter, but the user is free to treat it as such.

We have observed that when micrographs of a different type of particle are processed, we had to tune only the time-steps for smoothing and reset the particle-size on one micrograph while the rest of the parameters were left unchanged from values used previously for ribosomes. The parameter tweaking can be done in less than half an hour by following the control-flow of the operations as described in this paper.

Our goal is to consistently achieve 80% or more efficiency in boxing the same particles that were selected manually and to do so with less than 10% false positives, without human intervention. If $\Theta_{man} = \{p_1, p_2, \ldots, p_n\}$ is the set of manually boxed particles, where the size of the set $\#\{\Theta_{man}\} = n$, and $\Theta_{aut} = \{p_1, p_2, \ldots, p_m\}$ is the set of particles boxed automatically, where $\#\{\Theta_{aut}\} = m$, then $\{\Theta_{man} \cap \Theta_{aut}\}$ is the set of particles that are picked by both automatic and manual methods. The symbol '∩' is the set intersection that brings out those elements that are common to both sets. The set $\{\Theta_{man}\Theta_{man} \cap \Theta_{aut}\}$ is the set of particles that are manually picked but not picked up by the automated software. The symbol '\' denotes the set difference. The set, $\{\Theta_{aut} \setminus \{\Theta_{man} \cap \Theta_{aut}\}\}$ is the set of false positives, i.e., the set of particles boxed by the automatic method while rejected as non-particles by the manual method. The percentage efficiency of the automatic method is then calculated by % efficiency $= \frac{\#\{\Theta_{man} \cap \Theta_{aut}\}}{n} \times 100$.

3.1. Comparison of ribosome picks

Experiments were conducted on 55 micrographs containing images of more than one hundred and thirty thousand ribosome particles obtained by electron microscopy. These micrographs were collected in Albany, New York,
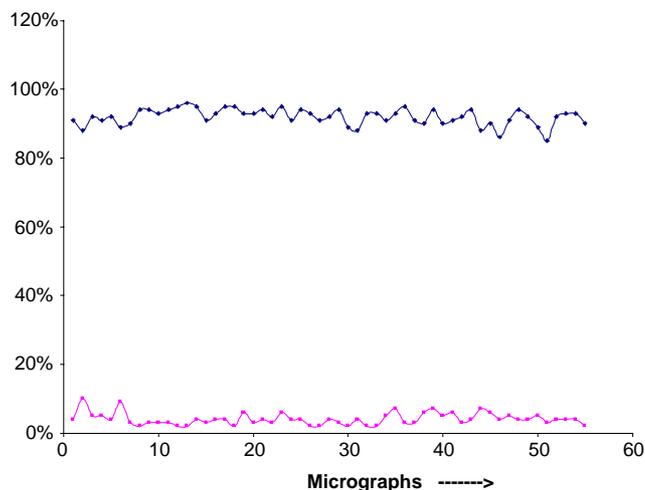
Fig. 5. Graphical representation of the efficiency of the segmentation-based software for particle-picking, tested on a set of 55 micrographs. The blue curve at the top of the chart shows the percentage efficiency in correctly picking the particles and the pink curve at the bottom of the chart shows the percentage rate of false-positives. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

obvious false-positives, for example 'red' spots over a background region, were eliminated. Fig. 5 shows a graph indicating the efficiency of the segmentation process for particle picking based upon manual editing by the first author of the paper. It can be observed that the automated segmentation program consistently satisfied the condition that it should show more than 80% efficiency in picking the same particles that are picked on the basis of human judgment, with less than 10% false positives. There is, however, a reasonable chance for introducing some sort of bias in quantifying the efficiency of the software when a person who has designed the software for automatic particle picking also does the manual picking and compares the results.

To reduce the likelihood of such a bias in the analysis, the second author of this paper has picked particles from a randomly selected subset of 14 micrographs. The micrographs were first processed by SPIDER software batch file *lfc_pick.spi* (Rath and Frank, 2004), to produce a set of candidate particles. During manual editing of this set of candidate particles, obvious false positives were first eliminated. In the next step of manual inspection, some of the true-positives were also eliminated on the basis of several objective and subjective criteria such as intensity, contrast, shape, visual estimation of high-frequency information, etc.

Table 1 shows a more detailed comparative evaluation that was made for the fourteen micrographs whose particles were picked by the second author of this paper. This list of particle coordinates is considered to be a gold standard for comparison. The statistics clearly indicate that the segmentation-based method performs well as regards to efficiency but appears to perform poorly with respect to the number of false positives. Fig. 6 shows a small part of a high-density micrograph in which the particles are marked. White boxes indicate ribosome particles marked by the segmentation-based software while red dots indicate

on a Philips Tecnai F30 at 300 kV using low-dose conditions at a magnification of 39 000×. The pixel size in the digitized micrographs corresponds to 3.72 Å/pixel (Valle et al., 2003).

In the first step of comparative analysis, the first author of this paper has visually inspected 55 micrographs of ribosome particles and counted all the false-positives marked by the segmentation software. To perform this check, the micrographs were displayed on a computer monitor and thick 'red' dots were used to indicate the coordinates of particles selected by the software. Every particle was then inspected in the context of its surroundings, and the

Table 1

Comparative analysis of particle picking software performance using 14 micrographs of ribosome particles: No. of picks in manual edit is the number of particles in the initial gold-standard data set that was obtained after manually editing the set of particles picked by SPIDER software batch file lfc_pick.spi

File names	No. of picks in manual edit	No. of auto picks	Apparent efficiency (%)	Apparent false positives (%)	Actual false positives (%)
msk003.spi	1561	2474	89	43	11
msk004.spi	1470	2353	89	44	11
msk005.spi	1338	2511	89	52	14
msk006.spi	1279	2398	90	51	13
msk008.spi	1621	2335	89	37	9
msk010.spi	1707	2541	91	38	9
msk013.spi	1424	2623	91	50	13
msk015.spi	1165	2456	89	57	15
msk016.spi	1155	2370	90	55	15
msk018.spi	1605	2601	93	42	11
msk019.spi	1493	2587	91	47	12
msk020.spi	1387	2561	90	50	13
msk021.spi	1423	2513	90	48	13
msk022.spi	1632	2493	92	39	10
msk027.spi	1504	2386	89	43	11

No. of auto picks is the number of particles picked by the segmentation-based method. Apparent efficiency and apparent false positives are the efficiency and false-positives, respectively, calculated by comparison with the initial gold-standard. Actual false positives shows a corrected estimate of the percentage of false positives in which it is assumed that 2/3rd of the apparent false positives as true positives (see the text for details).
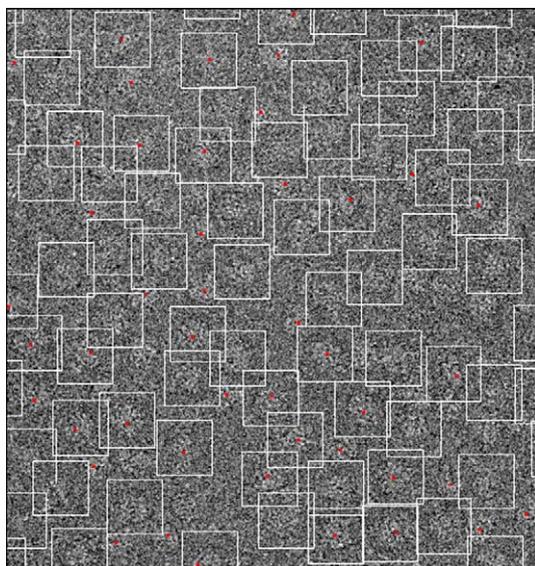
Fig. 6. A small part of a micrograph that contained a high density of particles. Particles marked by the automated segmentation-based software are shown in white boxes, and the particles included in the original gold-standard set are marked by red dots. Many of the white boxes that do not have red dots in them nevertheless appear to have boxed ribosome particles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

the particles manually selected. It can be observed that most of the white boxes have ribosome particles in them. Several of them were not selected as 'good' ribosome particles in the gold standard, however, based on subjective criteria. Thus, the efficiency and the percentage of false positives calculated by the direct comparison with the particles picked by the second author are referred to as the 'apparent efficiency' and the 'apparent false positives,' respectively.

To re-evaluate whether all particles that were boxed by the software but were not in the initial manually selected data set should be classified as false positives, the second author was provided with the same set of 14 micrographs in which only the 'apparent false positives' were marked by a red dot. The second author found that at least 2/3 of these 'apparent false positive' particles were, in fact, true particles that should be included in the gold standard set. Accordingly the percentage of false positives for automatic

particle picking was recalculated. The recalculated values are shown in column 6 of Table 1 as 'actual false positives.'

The automatic particle-picking software described in this paper is not designed to eliminate the particles based on quality criteria. It will pick every object in the micrograph that is deemed to be a particle based on the size and intensity characteristics of the segmented object. This has resulted in picking several particles that were not considered by a human expert to be good enough particles to be in the gold standard. At present, evaluation of the quality of a particle is subjective and depends on the expertise of the researcher. An important next objective will be to design algorithms to quantify the subjective criteria that are used in order to develop a quality index for the boxed particles. We believe that it will be possible to develop such an operation that would further reduce the percentage of false positives when compared to human analysis.

3.2. Comparison of TPP-II picks

Our long-term goal is to design a particle-picking approach that is more general in nature, i.e., one that does not need completely new developments each time that a new type of particle is to be picked. To test the generality of the current software and to identify aspects for further improvement, we have run the software on five micrographs of TPP-II complex, which is a giant protease isolated from *Drosophila* (Rockel et al., 2002). TPP-II complex has a twisted shape and exhibits distinct 2-D projections that cannot be classified as a simple blob-like structure, as is the case for ribosomes. The TPP-II complex shows several distinct shapes when projected onto a 2-D plane. Various shapes of side views are referred to as "navette" or boat-like, "bow-tie," and "fish" views, as well as more spherical end-views. These shapes are pictorially described in Rockel et al. (2002). Fig. 3 shows different stages of processing of the TPP-II micrographs using the above described operations and filters.

The use of size and intensity thresholds that match the side-views resulted in elimination of many particles with small area such as top-views or end-views. Relaxing the lower size-threshold value to pick end-views as well as side-views increases the number of false positives. Visual inspection of the segmented, false positives made it appar-

Table 2
Comparative analysis of particle picking software performance using five TPP-II micrographs

Micrograph	No. of manual pick	No. of auto pick	Efficiency (%)	False positives (%)	False positives after shape filtering (%)
TPP000.tif	098	116	93	19	8
TPP001.tif	438	457	90	13	6
TPP002.tif	288	293	87	14	6
TPP003.tif	392	436	94	15	7
TPP004.tif	411	424	89	16	7

Manual selection of the particles was done by the first author. In the first step, the parameters were tuned to select end-views as well as side-views. This resulted in a relatively high number of false positives, as evident from column 5. Column 6 reflects the reduction in percentage of false positives due to shape based filtering.

ent that many of them did not have the near-spherical shape expected of end-views. This made it possible to eliminate such false positives with criteria based on a combination of (1) object eccentricity, i.e., the ratio of maximum radius to minimum radius of the binary mask of the particle and (2) the size of the binary mask.

Table 2 gives the performance analysis of the segmentation-based software for automatically picking the TPP-II particles. When the objects in the binary image that are below a pre-defined size threshold and above a given object eccentricity limit were eliminated, the number of false positives decreased to fewer than 10% as evident from column 6 of Table 2. The results given in Table 2 indicate that the software is now suitable for testing on a much larger data set of TPP-II images.

We have observed, however, navette shaped particles are occasionally recognized as two end-views rather than as one single particle. An example is shown in Fig. 3H within a marked box. These cases do not affect the efficiency of the particle picking since we are not picking a false positive nor do we leave out a true-positive during the particle picking process. The negative effects of such cases are, (1) a same particle gets to be counted twice and (2) the extracted particle image will not be properly centered. We therefore believe that use of an improved shape-descriptor, such as templates, should still be added during the post-processing stages. In the next version of the particle-picking program we propose to implement templates as shape descriptors, to make the software more general in its applications.

Acknowledgments

References

Adiga, U.P.S., Malladi, R., Baxter, W., Glaeser, R.M., 2004. A binary segmentation approach for boxing ribosome particles in cryo EM micrographs. J. Struct. Biol. 145, 142–151.

Boisset, N., Penczek, P., Pochon, F., Frank, J., Lamy, J., 1993. Three-dimensional architecture of human α2—Macroglobulin transformed with methylamine. J. Mol. Biol. 232, 522–529.

Frank, J., 1996. Three-dimensional Electron Microscopy of Macromolecular Assemblies. Academic Press, San Diego.

Frank, J., 2002. Single-particle imaging of macromolecules by cryo-electron microscopy. Ann. Rev. Biophys. Biomol. Struct. 31, 303–319.

Glaeser, R.M., 1971. Limitations to significant information in biological electron microscopy as a result of radiation damage. J. Ultrastruc. Res. 36, 466–482.

Glaeser, R.M., 2004. Historical background: why is it important to improve automated particle selection methods? J. Struct. Biol. 145, 15–18.

Henderson, R., 1995. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q. Rev. Biophys. 28, 171–193.

Huang, Z., Penczek, P.A., 2004. Application of template matching technique to particle detection in electron micrographs. J. Struct. Biol. 145, 29–40.

Jain, R., Kasturi, R., Schunck, B.G., 1995. Machine Vision. McGraw-Hill, NY, pp. 43–47.

Malladi R., Ravve I., 2002. Fast difference schemes for edge enhancing Beltrami flow. In: Heyden et al. (Ed.), Computer Vision ECCV 2002, Lecture Notes in Computer Series, vol. 2350, pp. 343–357.

Nicholson, W.V., Glaeser, R.M., 2001. Review: automatic particle detection in electron microscopy. J. Struct. Biol. 133, 90–101.

Radermacher, M., Wagenknecht, T., Verschoor, A., Frank, J., 1987. Three-dimensional reconstruction from a single-exposure random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. J. Microsc. 146, 113–136.

Rath, B.K., Frank, J., 2004. Fast automatic particle picking from cryo-electron micrographs using a locally normalized cross-correlation function: a case study. J. Struct. Biol. 145, 84–90.

Rockel, B., Peters, J., Kühlmorgen, B., Glaeser, R.M., Baumeister, W., 2002. A giant protease with a twist: the TPP II complex from Drosophila studied by electron microscopy, EMBO J., vol. 21, No. 22, pp. 5979–5984.

Roseman, A.M., 2004. FindEM–a fast, efficient program for automatic selection of particles from electron micrographs. J. Struct. Biol., 91–99.

Sali, A., Glaeser, R.M., Earnest, T., Baumeister, W., 2003. From words to literature in structural proteomics. Nature 422, 216–225.

Sigworth, F.J., 2004. Classical detection theory and the cryo EM particle selection problem. J. Struct. Biol. 145, 111–122.

Valle, M., Gillet, R., Kaur, S., Henne, A., Ramakrishnan, V., Frank, J., 2003. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo EM. Nat. Struct. Biol. 10, 899–906.

Volkmann, N., 2004. An approach to automated particle picking from electron micrographs based on reduced representation templates. J. Struct. Biol. 145, 152–156.

Zhu, Y., Carragher, B., Glaeser, R.M., Fellmann, D., Bajaj, C., Bern, M., Mouche, F., Haas, F., Hall, R.J., Kriegman, D.J., Ludtke, S.J., Mallik, S.P., Penczek, P., Roseman, A.M., Sigworth, F.J., Volkmann, N., Potter, C.S., 2004. Automatic particle selection: results of a comparative study. J. Struct. Biol., 3–14.

Zhu, Y., Carragher, B., Kreigman, D.J., Milligan, R.A., Potter, C.S., 2001. Automated identification of filaments in cryoelectron microscopy images. J. Struct. Biol., 302–312.