Scanning Microscopy

Volume 1992
Number 6 *Signal and Image Processing in Microscopy and Microanalysis*

Article 40

1992

# Digital Image Processing of Electron Micrographs: The PIC System II

Benes L. Trus
*National Institutes of Health*

Michael Unser
*National Institutes of Health*

Thierry Pun
*National Institutes of Health*

Alasdair C. Steven
*National Institutes of Health*

**Scanning Microscopy Supplement 6, 1992 (pages 441-451)** 0892-953X/92$5.00+.00
Scanning Microscopy International, Chicago (AMF O'Hare), IL 60666 USA

# DIGITAL IMAGE PROCESSING OF ELECTRON MICROGRAPHS: THE PIC SYSTEM II

Benes L. Trus[1,3*], Michael Unser[2], Thierry Pun[2#] and Alasdair C. Steven[3]

[1] Computational Bioscience and Engineering Lab., Div. of Computer Research and Technology
[2] Biomedical Engineering and Instrumentation Program, National Center for Research Resources
[3] Laboratory for Structural Biology Research, National Institute of Arthritis, Musculoskeletal, and Skin Diseases, National Institutes of Health, Bethesda, MD 20892
[#] Present address: Centre Universitaire Informatique, University of Geneva, 12, rue du Lac, CH-1207 Geneva, Switzerland

## Abstract

The PIC system, an integrated package of Fortran programs and subroutines designed to run on the Digital Equipment Corporation VAX family of computers, has been developed for analysis of electron micrographs with emphasis on the particular requirements for structural analysis of biological macromolecules. The substantially improved VAX version of PIC reported here has been developed from an earlier PDP-11 version which was, in turn, developed from a set of IBM 370 programs called MDPP. PIC now encompasses over 150 commands or processing operations that afford a comprehensive range of image processing operations including image restoration, enhancement, Fourier analysis, correlation averaging, and multivariate statistical analysis including clustering and classification. In particular, we describe our software for correction of imperfect lattices, as well as programs for correlation alignment and averaging of "single particle" images.

Key Words: Image processing, computer software, enhancement, lattice corrections, electron microscopy, correlation averaging, signal processing

* Address for Correspondence:
Benes L. Trus
Building 12A, Room 2033,NIH
Bethesda, Maryland 20892
Phone No.: 301-496-1111
FAX No.: 301-402-2867

## Introduction

While electron microscopes are capable of instrumental resolutions of 2 to 3 A, imaging of biological macromolecules seldom produces usable information on this scale. Reasons for this shortfall include the use of negative stain, which limits resolution to about 20 A, drying artifacts, and radiation damage, which causes progressive degradation of the biological substructure. Quite recently, the emergence of cryo-electron microscopy has made it possible to obtain images of macromolecules in a 'native' state, but unfortunately at low contrast and an adverse signal-to-noise ratio, at least for the untreated images. In order to achieve the highest possible resolution, computer image processing of electron micrographs is routinely used. Combining, averaging, or otherwise analyzing multiple images (either 'single particle' images[1] or unit cell images extracted from a distorted lattice) can produce significant improvement in resolution, signal-to-noise ratio, and then in facilitating interpretation of macromolecular structure.

The PIC software system, originally implemented on a PDP-11/70 (Trus and Steven, 1981) and derived, in part, from the MDPP program written for the IBM 370 (Smith, 1978) has now been adapted to a VAX-8350 and a MicroVAX 3500 computer. This upgrade makes it possible to serve more users and to improve response time. In particular, the large virtual address space and large physical memory of the VAX allow processing of larger images (512 x 512 to 1024 x 1024) with little or no file Input/output (I/O) activity. Our fast Fourier transform (fft) routines do not require power of 2 Fouriers, and the size limitation of 1024 x 1024 can be easily increased by changing one parameter in the program. Furthermore, the new PIC system has been substantially improved by the addition of numerous operations, many of which are discussed below.

The primary philosophy behind PIC has been to provide an easy-to-use, fast, interactive image processing system encompassing state-of-the-art video technology. The scientist interacts with PIC using two-letter commands which switch processing to the appropriate routine. Future plans will incorporate a menu system so that the two-letter commands will become optional. Typically, the operations available include input/output, operations on the real image, Fourier space manipulation, polar coordinate processing for images that contain some angular symmetry, and operations involving use of the video frame buffer memory. New operations make it possible to perform single particle averaging by means of correlation alignment, multivariate statistical analysis, classification, and cluster analysis. Bent or curved filamentous particles may be straightened, and irregular crystalline lattices may be corrected. Because of the compartmentalization of operations, it is relatively easy to add or substitute new routines as necessary. Though PIC was originally designed for image processing of electron micrographs, many of the functions available for image analysis in other disciplines, as exemplified by use at NIH to analyze other biological images such as gel electrophoregrams. An extensive help library has been created, and is available at the command level as needed.

PIC was initially designed as an interactive system, but batch image processing may now be easily performed if desired. Typically, PIC is used interactively to analyze one or a few images, and to work out the most effective strategy for addressing a particular type of problem. Once this strategy has been determined, the ideal approach is to use the batch features to process a (large) group of images with the same commands and parameters. This not only assures that all images will be processed uniformly, but a command file remains for future reference as documentation of exactly how the batch was processed, and which parameters were chosen.

[1] 'Single particle' images refers to multiple images of the same macromolecule or molecular assembly which primarily differ in their two-dimensional orientation. Computer alignment of such ostensibly similar images permits the averaging of the images assuming the signal components are similar, but the noise is random. With N such images a signal-to-noise improvement of $\sqrt{N}$ may be achieved.

The hardware available to PIC at our installation is shown in Figure 1. To maintain PIC's hardware independence, modular, spawned processes (secondary processes initiated by the main process), which may communicate by means of either shared global memory sections or temporary disk files interact with our specialized image processing hardware. This arrangement is conducive to flexibility and exportability to other VAX VMS computers. Routines which would require alteration or replacement at other installations are segregated into separate processes.

PIC has been given to about a dozen laboratories. However, some components or algorithms developed for of PIC are more widely used in the image processing of electron micrographs. Plans are currently underway to convert PIC to an X-window environment under the VMS operating system. However, moving to UNIX is not planned for the immediate future. This description of the PIC system is not intended as a review of image processing packages available to the electron microscopist. A up-to-date review of such software now available (Hegerl, 1992).

## Special Hardware

Several special-purpose image processing devices are integrated into the VAX system at our installation. We currently have a Vicom IP8500 multiuser image processor system with three medium resolution (512 x 512) and one high resolution (1024 x 1024) displays and a digital video processor, a Matrix 4007 camera station for rapid photographic output, a Hamamatsu video camera, and a Perkin-Elmer 1010MG microdensitometer for high resolution digitization. Each image processing 'workstation' comprises a terminal, Summagraphics digitizing tablet, joystick, as well as a TV monitor.

The Vicom IP8500 contains a Digital Video Processor or DVP which can perform arithmetic operations on a 512 x 512 x 8 bit image in one video frame time (1/30 sec). This specialized hardware has been used to replace some software operations resulting in substantial time savings. For example, the command Compute Mass[2] will total the number of pixels under a mask or compute their integral in one frame time. (A mask is typically a different 512 x 512 x 8 bit image.) Without the DVP, 5 seconds of CPU time and substantial memory would be required. This option has been useful both in mass analyses by electron microscopy (Steven et al., 1983a; Wall and Hainfeld, 1986) and in morphometric studies. Software

[2] The names of programs called by PIC are underlined.

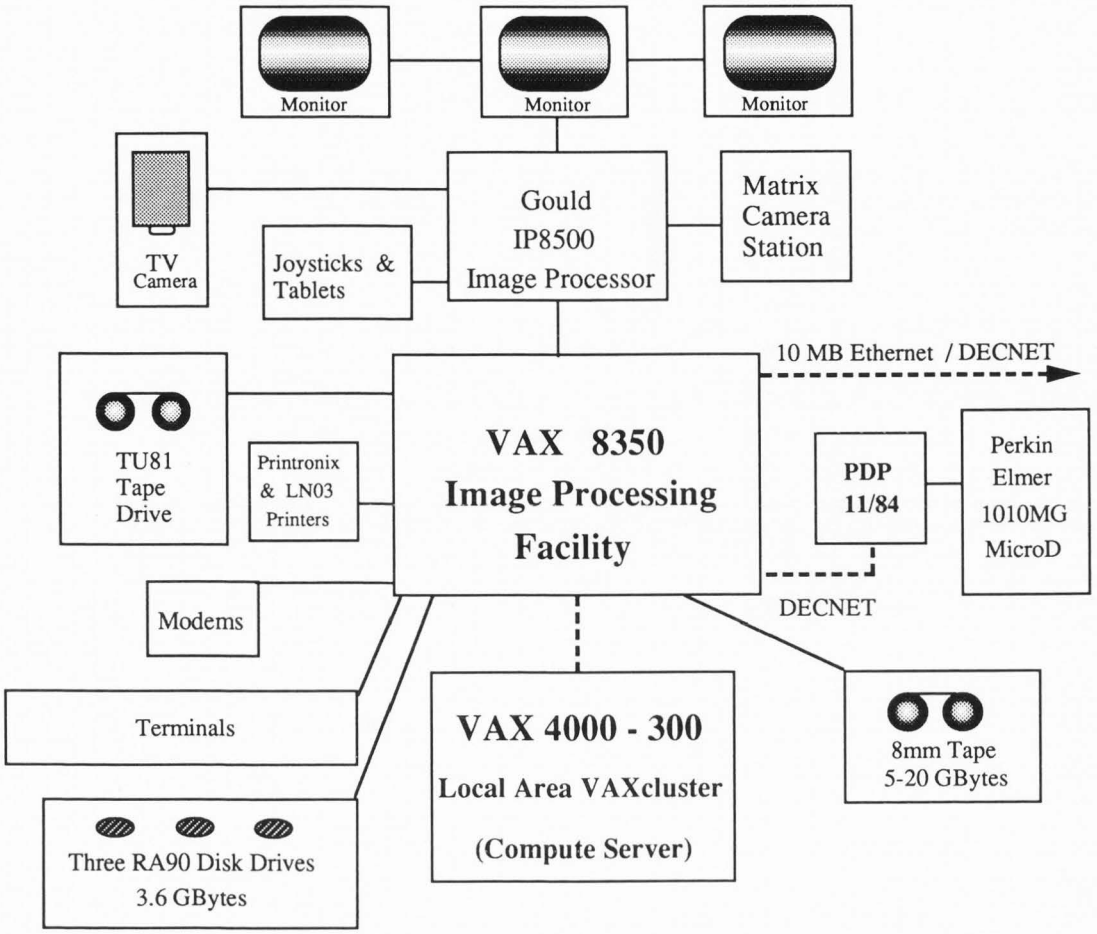


Figure 1 Schematic diagram of the hardware configuration at NIH on which the PIC system has been implemented.

using the DVP can interpolatively zoom images or convolve images using a variable kernel, in near real-time. However, in general, video software is implemented in the VAX rather than in the DVP.

## New Operations

Illustrative examples of new PIC operations are described below. These were either unavailable on the earlier 11/70 version of PIC (Trus and Steven, 1981) or have undergone major revision and improvement as a result of the new hardware architecture.

### Frame Buffer Operations

One of the most useful and important features of PIC is the ability to readily display images or graphs, thus making possible truly interactive processing. The Raster program can be used to display either the memory-resident PIC image, or various differently formatted picture files from disk. The program Multi is used to display a montage of (similarly dimensioned) images for comparison or analysis. Plot produces a graphics plot of one-dimensional (line) data. Two programs, Font and Label overlay text on the displayed image, which is particularly useful in composing material for slides or photo-montages.

Numerous programs read, examine, process, or modify the contents of the Vicom memory. Manipulate can be used to zoom, scroll, and alter the contrast of the video image. Tablet uses the Summagraphics digitizing tablet, and joystick to examine specific areas of the image. Histogram equalization may be performed in the frame buffer, or the Full range of contrast may be utilized on the tv display monitor. This range is determined by calculating the 2nd and 98th percentile points of the image histogram and mapping the image so that these values define the limits of the 8 bits available (0-255). Images which are rotationally symmetric (both isotropic and anisotropic) may be azimuthally averaged by Radial to produce a one-dimensional representation. Curved

filamentous particles can be Unbent (Steven et al., 1986b; Egelman, 1986) i.e. straightened, to facilitate averaging, Fourier processing, or projection analysis. Similarly, 2-dimensional crystals can be Unwarped for averaging or filtering. Distances or contour lengths can be determined. Chop is used to save all or part of the Vicom memory in an unformatted byte disk file.

**Single Particle Averaging**

Molecules in general may attach to the grid in random three-dimensional orientations. However, some specimens have a preferential attachment site so that all (or most) images present the same projection view of the macromolecule, but differ essentially in rotational setting. Correlation averaging techniques may be used to combine images, yielding a significant signal-to-noise improvement (Frank et al., 1981). We have developed software which performs correlation averaging, and is explained in the next section. Additional programs utilize multivariate statistics (van Heel and Frank, 1981) cluster analysis, classification, and outlier detection methods (Unser et al., 1986) to evaluate the results of averaging. Such software may be used to separate data into different classes or clusters, if appropriate; or, alternatively, to select the best preserved particles for averaging, while excluding those which are noisy, less consistent, or otherwise undesirable for inclusion in a global average.

## Correction of Lattice Distortions

**Straightening of Linear or Filamentous Particles**

Flexible helical particles often attach to the electron micrograph grid in a bent or curved configuration. Of primary importance is the ability to determine helical parameters from their diffraction patterns, and to analyze or reconstruct the three-dimensional structure of such filaments (DeRosier and Moore, 1970). Accurate analysis of these helices is dependent upon their being precisely straight. One solution is computational a posteriori straightening by interpolating the images on to a curvilinear coordinate system defined by the particle's axis (Fraser et al., 1976).

We have developed methods which use natural cubic spline interpolation. Our approach uses one-dimensional cubic splines (Steven et al., 1985a; Steven et al., 1986b, Steven et al., 1991). Bilinear interpolation is used to resample the image. However, to correct tightly curved particles, or those whose axis is not approximately horizontal or vertical, the procedure has been extended to two-dimensions by representing both coordinates of each point along the particle's axis by two separate piecewise cubic functions of the contour length (Kocsis et al., 1991).

Using a graph-pen and tablet one interactively selects N nodes along the filament's axis (on a video display) to describe N-1 sets of cubic spline coefficients. Our coefficients are constrained so that the first and second derivatives are continuous across each node, and zero at the ends (a natural spline). To straighten the filament, we resample the particle at unit pixel distances along this curve and on straight lines running perpendicular to the curve, thus generating a two-dimensional image. The process is iterative, in that the results may be inspected and corrected after each functional fit. Adjustment of node points may be interactive, or automatic. Interactive refinement typically involves looking along the straightened filament and locating deviations from linear, then estimating corrected node coordinates to achieve a more linear filament. While node coordinates are originally integers, as determined by the graph-pen and tablet, we have found that non-integer values are usually required to achieve the straightest possible filaments. Automatic refinement of nodes can be accomplished either by using a cross-correlation function which compares the local average along the filament with the global average or by comparing a line across the straightened filament with its computational two-fold symmetry and adjusting the node to achieve a better center of symmetry. The physical interpretation of this unbending (spline fit) mechanism is that it follows the center of the particle and thereby represents the minimum strain energy along the particle.

**Correction of Two-dimensional arrays**

For periodic two-dimensional arrays, averaging (Carrascosa and Steven, 1979, Smith and Aebi, 1976) is an exceptionally useful procedure for improving their signal-to-noise ratio. This may be accomplished in Fourier space by masking out the nonperiodic components, or in real space by summing unit cells extracted from the periodic lattice.

Unfortunately, not all two-dimensional macromolecular crystals are perfect. In this case the lattice distortions should be computationally corrected before averaging, in order to achieve the highest possible signal-to-noise improvement. In a particularly problematic case, thin cryo-sections of skeletal muscle (Trus et al., 1989), compression of up to 50% is not uncommon. The necessary corrections may be applied in one of two ways. The lattice can be

corrected[3] after determination of lattice displacement vectors, and then the non-periodic Fourier components can be filtered as before (Saxton, 1978). Alternately, single or groups of unit cells may be removed from the distorted lattice by an unwarping function which corrects the individual unit cells (Unser et al., 1987a, Unser et al., 1989a).

We have implemented the latter case as described below. First, we create a cross-correlation function (XCF) between the area of interest and a reference unit cell or group of unit cells. The area of interest typically contains from 30 to 60 distorted unit cells. The reference image might be obtained from a straightforward Fourier filtration of the area of interest, or it may be a model describing ideal unit cells. In either case, the reference may be low-pass filtered as well. A Gaussian window filter is usually applied to the result of a simple Fourier filtration to reduce the effective area of the filtration to a few (e.g. 1.5 to 2.5) unit cells in each direction. This window is typically a radially symmetric Gaussian filter with 1.0 at the center, and a half-width of about 1 unit cell repeat.

Next, a peak search program displays graphically the major peaks over the image of the cross-correlation function, and the better unit cells (those without any obvious defects) are selected either interactively or semi-automatically. These fiduciary points are used as input for the unwarping program, which estimates a spatial distortion function by fitting piece-wise quasi-hermite two-dimensional polynomials across the peaks in the x- and y-directions. The unwarping process is iterative, in that the fiduciary points are unwarped, and may be cross-correlated again with the reference allowing their positions to be refined. The refinement of the fiduciary peaks optimizes the signal-to-noise ratio (SNR) and has been found to improve the SNR by 2 to 30% (Unser et al., 1989a) (for badly distorted cryo-sections of muscle).

This optimization procedure has the following advantages: (a) There is complete flexibility in the choice of fiduciary points. The traditional use of a rectangular or quadrilateral unit cell is not longer required. Fiduciary points need not describe a rectangular or trapezoidal unit cell, and more than four points may be utilized. (b) Because of the iterative refinement, inaccurate fiduciary points may be corrected after the first cycle of unwarping. We have found the method robust, expecting quick convergence (one or two cycles of refinement) with points that were initially displaced by 1 to 2 pixels from their ideal locations. (c) The maximum global SNR is thus achieved by iterative refinement of the fiduciary points.

Individual unit cells are stored in separate disk files, and subsequent analysis includes outlier detection by the OMO (Odd Man Out) algorithm (Unser et al., 1986) to screen out anomalous data, unit cell averaging, and resolution determination by the Spectral Signal-to-Noise Ratio criterion (Unser et al., 1987b).

## Correlation Averaging and Factorial Analysis

Correlation averaging and subsequent statistical analyses as performed with the PIC system actually consists of six sequential steps, each of which will be described in some detail. (a) Preprocessing the input data; (b) Single images are aligned by correlation with a reference template. (c) The results of the alignment are subjected to statistical analysis in order to remove outliers. (d) Only then can the images be averaged to reduce noise. (e) The resulting average image may require further refinement, filtering, or symmetrization. (f) The resolution of the averaged image is calculated.

### Preprocessing of Data

In the beginning, the analyst must designate a reference image to be used in the correlation alignment. An image, usually containing many individual particles, is displayed on the monitor and the subimages of each particle are extracted interactively with the tablet and a video image. The analyst must specify the image size, and then simply point to the center of the particle. Such extracted files are N x N pixels, where typically $N \leq 512$. Initially, one such file must be chosen which does not move (and is called a reference) and a second file (called the subject) is aligned against it. As one or more subject files are aligned, they may be averaged with the reference to produce a less noisy reference. Typically, the analyst uses the PIC option Select to extract a set of particles from one or a few 512x512 images.

Once the subimages have been obtained they require further pre-processing. The gradient of the background density across the field may or may not be uniform. To assure uniformity of background, the gradient is determined and subtracted. A function of

[3] There are a number of types of lattice correction algorithms including use of bilinear-interpolation (Trus et al., 1989, Crowther and Sleytr, 1977) polynomials (Unser et al., 1987a, Eden et al., 1986) or a two-dimensional cubic splines (Henderson et al., 1986) to represent the lattice.

the form ax + by + c is calculated by least squares for the background, and then subtracted to correct its gradient. Either the entire image, or a specified portion may be used to determine the function's parameters (a,b,c) (e.g. for a circular molecule the central circular region of the image is excluded from the least squares fit). Next, images are normalized. We have generally used a constant mean and constant variance normalization (Unser et al., 1989b). An optional preprocessing step is to bandlimit the initial images to remove high frequency noise. This option is also available as part of the correlation alignment procedure (see below).

**Correlation Alignment**

Correlation alignment (CA) may be performed easily using a range of options from semi-automatic through highly interactive (with no default parameters). Our philosophy of correlation alignment has evolved from our practical experience in this area, based upon pooling published reported methods (Frank et al., 1981; van Heel and Frank, 1981; Saxton, 1978; Frank et al., 1982; Frank and van Heel, 1982) with procedures of our own design. The default parameters described below are in fact variables which may be user-specified. Our method was first used to illucidate the structure of the hexavalent capsomers of herpes simplex virus type 2 (Steven, et al., 1986c).

Figure 2 details the procedure for rotational and translational alignment. In the first pass, one can choose to rotationally align the autocorrelation function (ACF) of the subject with the reference (not shown) (Saxton, 1978) since the ACF is insensitive to lateral displacements. However, our experience and that of others (Steinkilberg and Schramm, 1980) indicates that use of the ACF for the first alignment is less satisfactory than using the images directly, especially if circumstances permit a reasonable estimate of the particle's alignment, In fact, we feel that use of the ACF is usually not as good as starting rotational refinement with the image, due in part to the two-fold redundancy in the ACF and to the fact that the ACF contains less information than the original image (the phase information has been lost). We are generally able to pick the centers of test particles to within 1 or 2 pixel units, which is generally satisfactory to begin with real space alignment. The translational alignment is given by the peak of the cross-correlation function (XCF) between the rotated subject and the reference. The usual procedure is then to repeat the alternation of cycles of rotational and translational alignment using the images directly.

To perform the rotational alignment, the image is mapped onto a polar coordinate system. Ring functions are created by sampling the image around each circumference, at distances increasing linearly from the center. The maximum sampling rate is a function of the resolution expected, and may be determined as follows. Each ring function has the same number k*N of samples, where N is the image size and k is a proportionality factor to be chosen. The distances between ring and center is d*N/2, with $0 < d < 1$. The linear sampling step along each circumference is then

$$T = 2 \pi (d*N/2)/kN = \pi d/k \qquad (1)$$

The angular sampling step is 360/(k*N) degrees. Using, for example, five rings at relative distances d = 0.15, 0.30, 0.45, 0.60, and 0.75, and with k = 2, the normalized sampling steps T are 0.24, 0.47, 0.71, 0.94, and 1.18. This provides a reasonable compromise between under- and over-sampling. In fact, the number of rings used is a variable.

Each ring function from the subject particle is correlated with the corresponding one from the reference, and a weighted sum of all the XCFs is calculated. Weights can simply be taken as the radii themselves, to compensate for the increase in sampling step when going further away from the center. A polynomial fit on the peak of this sum yields the rotational angle.

The procedure is repeated until a self-consistent result is obtained, i.e. no more significant changes are observed in rotational (typically, < 0.5 degrees) or translation (typically, < 1 pixel) alignment. Convergence may occur in as few as 3 cycles, but is highly dependent upon the nature of the images and the convergence limits set. Since successive rotations and translations involve interpolations on the subject image, a concomitant (interpolation-associated) blurring is incurred at each step. At convergence, equivalent net rotation and translation operations are then performed on the original image, with overall displacements determined by the sum of the individual displacements from each iteration.

There may remain an ambiguity of mirror symmetry as a result of which side of the specimen is attached to the grid. After completion of the alignment, the cross-correlations of the reference with the result and with the mirror result are computed. Since reference and subject are aligned, these products yield the peak value of the XCF; the higher of the two tells whether to retain the result or its reflected version.

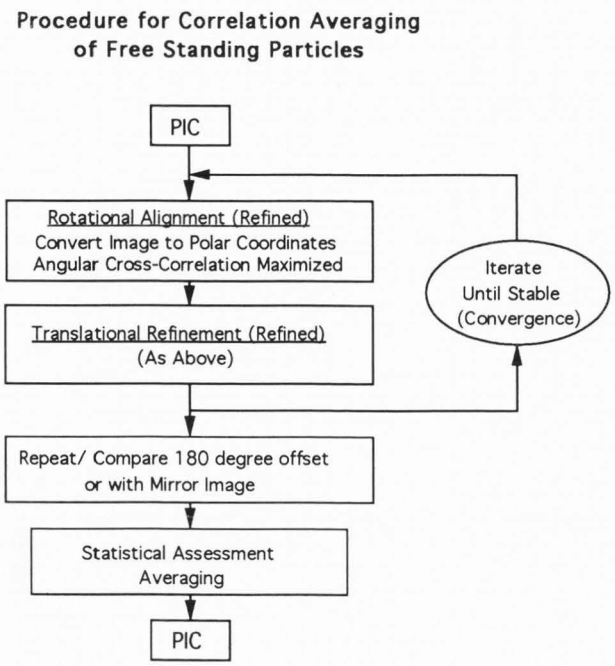


Figure 2 Schematic diagram of the procedure for translational and rotational alignment as components of correlation alignment and averaging.

**Statistical Evaluation of Results**

After images have been aligned, one must determine which images are appropriate to average. We use three types of statistical methods to evaluate the results before combining them; viz (a) factorial analysis to identify clusters of "like" images (Mardia et al., 1979; Jolliffe, 1986); (b) classification to recognize different classes from the noisy images; and (c) our OMO algorithm (Unser et al., 1986) to rank images from "best" to "worst" in each class or cluster.

A variety of approaches to the classification problem have been proposed, and there has been extensive discussion of their respective virtues (Frank et al., 1988, van Heel, 1989a, van Heel, 1989b, Frank et al., 1989), although a consensus has yet to emerge. A detailed discussion of classification is beyond the scope of the present article, and we confine ourselves to a few general remarks. In order to aid classification, it is desirable that the signal-to-noise ratio be improved to the maximum extent possible. Normally this is accomplished by averaging. However, a "hen-egg" situation arises in the sense that the classification must first be performed in order to determine which images should be averaged together, whereas an optimal classification depends on having the best possible signal-to-noise ratio in the data (i.e. prior averaging).

One approach to this problem is to represent the images in terms of the eigenvector space of some multivariate statistical formalism (e.g. correspondence analysis (Benzecri, 1979), principal components (Hotelling, 1933; Engelhardt, et al., 1985; Jolliffe, 1986), etc), and to seek to improve the signal-to-noise by confining attention (in the classification) to a limited number of dimensions in this space. However, this procedure is less than ideal, since the resulting factorial subspace is not noise-free, and the outcome of the classification will, in general, depend on how many factors are selected, which is a somewhat arbitrary decision.

Another issue that must ultimately be confronted, regardless of the approach taken, is how to reach a decision as to how many discreet classes are represented in the data set. We are not aware of a rigorous solution to the problem. Often, this decision is made by applying some criterion in an abstract multidimensional factorial space. This approach has the virtues of being quantitative and objective, but is nevertheless arbitrary at some level. Alternatively, an experienced analyst, after visual review of the data set may identify prototypic members of recurring image types, and these images are then used to seed a "supervised" classification. This approach has the disadvantage that the resulting classification depends on the subjective decisions of the 'supervisor', but has the redeeming feature that the classification decisions relate directly to specific features of the images per se, not to some vaguer properties of the images as represented in an abstract factorial space. The subjectivity problem may be offset to some extent by having independent classifications performed by different supervisors, and seeking to attain consistency.

It should also be noted that classification and correlation alignment are interrelated, i.e. in order to achieve the highest possible resolution, the alignment procedure should be repeated with reference to the in-class average, after a first cycle of the analysis (Schatz and van Heel, 1990). It is also the case that the initial data must have a reasonably good signal-to-noise ratio, otherwise any classification scheme, no matter how sophisticated, is condemned to ambiguity.

We have systematically compared two factorial representations, Correspondence Analysis and Principal Components to determine whether either facilitated a more efficient identification of "like" images (classes) (Trus et al., 1988; Unser et al., 1989b). The results suggested that a slightly better

separation of classes for this set of negatively stained data was obtained by using Principal Components with appropriate pre-normalization (constant mean and constant variance).

If there is reason to believe that only one class exists, or that the images have been correctly assigned to several different classes, then the OMO (Odd Men Out) algorithm is useful for identifying outliers in each class. By classifying these images from "best" to "worst", the "worst" images can be eliminated, up to a predetermined threshold of acceptability and the remaining "best" images used for averaging.

We are of the opinion that the classification problem has not yet been conclusively solved, and is still an area of active research. Moreover, the more tools available for classification, the more likely a judicious solution can be found for any given set of images. One method, which we have used in some of our analyses (Steven, et al., 1988; Trus, et al., 1988; Trus, et al., 1989; Unser et al., 1989b), is a supervised classification system. Software has been developed which computes the distance of each image from a set of reference images. Each reference image is obtained by averaging a sub-set of images that have been designated as representative of a given class; for N classes, there will be N such reference images. Automated classification is then accomplished by using a standard minimum distance classifier (Duda and Hart, 1973). In this approach, the pixels of each image define the components of a feature vector. Similarity is expressed in terms of the Euclidean distance between the feature vectors.

**Averaging of Images**

When images are identified as appropriate for averaging, then these images can be combined in several ways. The easiest implementation is direct, i.e. unweighted averaging of the "best" images, so characterized by the OMO algorithm. Alternatively, the images can be averaged according to a likelihood-based weighting scheme. The weights can, for example, be assigned so as to maximize a global signal-to-noise ratio criterion (Unser and Eden, 1990). Interestingly enough, we found that this approach produces weights that were highly consistent with the outcome of the OMO algorithm (Unser et al., 1986). Another technique is to consider the distribution of grey values assigned to each pixel, and to take the median pixel grey level at each spatial location. However, in practice, we have usually found that direct averaging of the "best" images is quite satisfactory.

A novel approach to correlation alignment involves correlation averaging of separate components of a macromolecular assembly. A study, using T7 tail fibers (Steven et al., 1988), was probably the first use of piece-wise correlation alignment - where several distinct parts of the particle under analysis were aligned and averaged separately, and then re-synthesized into a final composite image. A similar re-synthesis after averaging was performed on clathrin triskelion legs (Kocsis et al., 1991).

**Refinement of Average Images**

Once averaged images have been obtained, there are two optional steps which can be used for further noise reduction. First, if the average image has symmetry (e.g. n-fold rotational or mirror symmetry), then strict imposition of this symmetry can be used to further reduce noise. Second, a final low-pass band-limiting filtration may be appropriate if the image still contains an appreciable amount of high frequency noise. For instance, one might use a cosine cutoff with a value of 1.0 at $(20\ A)^{-1}$ decaying smoothly to 0.0 at $(15\ A)^{-1}$.

**Quality Assessment**

There are two measures of quality control that are applicable to the final averaged image. It is advisable that both should be performed. First, the entire correlation alignment and averaging process should be repeated on at least two independent sets of data to establish reproducibility. Second, the resolution of the averaged images should be determined. Initially, resolution was determined by two methods (Frank et al., 1981, Saxton and Baumeister 1982), both of which divided the data into two arbitrary subsets. A newer method has now been introduced (Unser et al., 1987b) which is based on the Spectral Signal-To-Noise Ratio (SSNR) of the set of individual images. The SSNR method has the advantages that a) it uses all of the images rather than an arbitrary division of the data into two averages; b) the estimate of the resolution has a lower statistical uncertainty than the previous two methods (the uncertainty of the estimate is less); c) it can be used to predict how many images are necessary to achieve a desired improvement in signal-to-noise; and d) the method is directly related to resolution criteria for crystallographic two-dimensional arrays.

## Mass Calculation Programs

We have developed many programs for the purpose of analyzing micrographs from the Brookhaven Scanning Transmission Electron Microscope (STEM) (Hainfeld et al., 1982). These programs typically use the video hardware to achieve optimal

(real time) speed and are highly interactive. Backgrounds may be either estimated by fitting a smooth function across observed backgrounds or measured locally with the routine Fixit. Box examines specified rectangles (which need not be orthogonal to the coordinate system of the array) and is used to determine integrated masses or mass per unit length (e.g. for TMV as an internal standard, intermediate filaments (Steven et al., 1983a; Steven et al., 1983b); or bacterial fimbriae (Steven et al., 1986a). Circle is comparable to Box, but operates on specified circular areas (e.g. for coated vesicles (Steven et al., 1983c)). Programs have also been developed to determine the radial mass distribution in filamentous or spherical specimens using a real space algorithm which accepts as input the projected STEM image (Steven et al., 1984; Steven et al., 1985b). This program accepts unflattened filamentous images as well as data slightly influenced by flattening on to the grid. The Sample program produces Vernier sampling of Brookhaven STEM data perpendicular to a filamentous axis (Steven et al., 1984). The Shape routine may be used to evaluate images of arbitrary shaped particles. Many of these routines may also be applied in morphometric analyses.

## Conclusion

We announce and briefly describe the VAX version of the PIC image processing system. New software is described which is applicable to straightening filamentous particles, unwarping distorted lattices, single particle correlation alignment, averaging, and multivariate statistical analyses. The software, written entirely in Fortran for a VAX is modular in design allowing modification or substitution for other laboratories, and is available without charge to other institutions.

## Acknowledgement

We are especially grateful to William Risso and Arthur J. Pashayan for contributions to the initial development of hardware and library software routines for the PIC system. We thank Eva Kocsis and James Conway for recent contributions, and Murry Eden for support of algorithm development.

## References

Benzecri, J-P (1979) L'Analyses des Donnees, Vol. 1: La Taxinomie (Dunod, Paris) 1-615.

Carrascosa, J L and Steven, A C (1979) "A procedure for evaluation of significant structural differences between related arrays of protein molecules", Micron 9, 199-206.

Crowther, R A, and Sleytr, U B (1977) "An Analysis of the Fine Structure of the Surface Layers from Two Strains of Clostridia, Including Correction for Distorted Images", J. Ultrastruct. Res. 58, 41-49.

DeRosier, D J and Moore, P B (1970) "Reconstruction of Three-dimensional Images from Electron Micrographs of Structures with Helical Symmetry", J. Mol. Biol. 52, 355-369.

Duda, R and Hart, P (1973) Pattern Classification and Analysis New York, John Wiley & Sons 10-43, 189-260.

Eden, M, Unser, M, and Leonardi, R (1986) "Polynomial representation of pictures", Signal Proc 10, 385-393.

Egelman, E (1986) "An Algorithm for Straightening Images of Curved Filamentous Structures", Ultramicroscopy 19, 367-373.

Engelhardt, H, Guckenberger, R, Hegerl, R, and Baumeister, W (1985) "High resolution shadowing of freeze dried bacterial photosynthetic membranes: Multivariate statistical analysis and surface relief reconstruction. Ultramicroscopy 16, 395-410.

Frank, J, Verschoor, A, Boublik, M (1981) "Computer averaging of electron micrographs of 40S ribosomal subunits", Science, 214, 1353-1355.

Frank, J, and van Heel, M (1982) "Correspondence Analysis of Aligned Images of Biological Particles", J. Mol. Biol., 161, 134-137.

Frank, J, Verschoor, A, Boublik, M (1982) "Multivariate Statistical Analysis of Ribosome Electron Micrographs; L and R Lateral Views of the 40S Subunit from HeLa Cells", J. Mol. Biol., 161, 107-133.

Frank, J, Bretaudiere, J, Carazo, J M, Verschoor, A, and Wagenknecht, T (1988) "Classification of images of biomolecular assemblies: a study of ribosomes and ribosomal subunits of Escherichia coli" J Microscopy 150 99-115.

Frank, J, Bretaudiere, J-P, Carazo, J-M, and Verschoor, A (1989) "Purity, Purposes, and Possibilities of Classification Techniques in Electron Microscopy: a Reply to Comments by Marin van Heel", J Microscopy 159, 117-122.

Fraser, R D B, MacRae, T P, Suzuki, E, Davey, C L (1976) "Image processing of electron micrographs of deformed filaments", J. Microscopy (Oxford) 108 343-348.

Hainfeld, J F, Wall, J S, and Desmond, E J (1982) "A Small Computer System for Micrograph Analysis", Ultramicroscopy, 8, 263-270.

Hegerl, R (1992) "A brief survey of software packages for image processing in biological electron microscopy" Ultramicroscopy, 47, 417-423.

Henderson, R, Baldwin, J M, Downing, K H, Lepault, J, and Zemlin, F (1986)

"Structure of Purple Membrane from Halobacterium Halobium: Recording, Measurement and Evaluation of Electron Micrographs at 3.5 A Resolution", Ultramicroscopy 19, 147-178.

Hotelling, H (1933) "Analysis of complex statistical variables into principal components", J. Educ. Psychol. 24, 417-441, 498-520.

Jolliffe, I T (1986) Principal Component Analysis, Springer, Berlin 1-271.

Kocsis, E, Trus, B L, Steer, C J, Bisher, M E, Steven, A C (1991) "Image Averaging of Flexible Fibrous Macromolecules: The Clathrin Triskelion Has an Elastic Proximal Segment", J of Structural Biol. 107, 6-14.

Mardia, K V, Kent, J T, Bibby, J M (1979) Multivariate Analysis, Academic Press, London 1-521.

Saxton, W O (1978) Computer Techniques for Image Processing in Electron Microscopy, Academic Press, New York 1-288.

Saxton, W O and Baumeister W (1982) "The correlation averaging of a regularly arranged bacterial cell envelope protein", J of Microscopy 127, 127-138.

Schatz, M, and van Heel, M (1990) "Invariant Classification of Molecular Views in Electron Micrographs", Ultramicroscopy 32, 255-264.

Smith, P R (1978) "An Integrated Set of Computer Programs for Processing Electron Micrographs of Biological Structures", Ultramicroscopy 3, 153-160.

Smith P R and Aebi, U (1976), The computer filtration of hexagonal lattices," J. Supramolec. Structure 5, 493-495.

Steinkilberg, M, and Schramm, H J (1980) "Eine verbesserte Drehkorrelationsmethode fur die Struktubestimmung biologischer Makromolelcule durch Mittelung electronenmikroscopischer Bilder", Hoppe-Seyler's Z. Physiol. Chem., 361, 1363-1369.

Steven, A C, Hainfeld, J F, Trus, B L, Wall, J S, and Steinert, P M (1983a) "Epidermal Keratin Filaments Assembled In Vitro Have Masses-per-Unit-Length That Scale According to Average Subunit Mass: Structural Basis for Homologous Packing of Subunits in Intermediate Filaments", J.Cell Biol., 97, 1939-1944.

Steven, A C, Hainfeld, J F, Trus, B L, Wall, J S and Steinert, P M (1983b) "The Distribution of Mass in Heteropolymer Intermediate Filaments Assembled in Vitro", J.Biol.Chem., 258, 8323-8329.

Steven, A C, Hainfeld, J F, Wall, J S, and Steer, C J (1983c) "Mass Distributions of Coated Vesicles Isolated from Liver and Brain: Analysis by Scanning Transmission Electron Microscopy,", J.Cell Biol., 97, 1714-1723.

Steven, A C, Hainfeld, J F, Trus, B L, Steinert, P M, and Wall, J S (1984) "Radial distributions of density within macromolecular complexes determined from dark-field electron micrographs," PNAS (USA), 81, 6363-6367.

Steven, A C, Stall, R, Steinert, P M, and Trus, B L (1985a) "Computational Straightening of Images of Curved Macromolecular Helices by Cubic Spline Interpolation Facilitates Structural Analysis by Fourier Methods", in Proc. of the 43rd Annual Meeting of the Electron Microscopy Society of America, Bailey, G W (Editor) San Francisco Press, 738-739.

Steven, A C, Trus, B L, Hainfeld, J F, Wall, J S, and Steinert, P M (1985b) "Conformity and Diversity in the Structures of Intermediate Filaments", in Intermediate Filaments (Vol 455 Annals of the New York Academy of Sciences, Eds. Wang, E, Fischman, D, Liem R K H, and Sun T T), 371-380.

Steven, A C, Bisher, M E, Trus, B L, nhomas, D, Zhang, J M, and Cowell, J L (1986a) "Helical Structure of Bordetella pertussis Fimbriae", J. of Bacteriology, 167 (3), 968-974.

Steven, A C, Roberts, C R, Hay, J, Bisher, M E, Pun, T, and Trus, B L (1986b) "Hexavalent Capsomers of Herpes Simplex Virus Type 2: Symmetry, Shape, Dimensions, and Oligomeric Status", J of Virology, 57, 578-584.

Steven, A C, Stall, R, Steinert, P M, Trus, B L (1986c) "The Structure of Intermediate Filaments from Electron Microscopy: Potentialities of Computational Straightening of Curved Filaments to Admit Fourier Analysis of Helical Organization", in Electron Microscopy and Alzheimer's Disease, J. Metuzals, Ed., pp 31-33.

Steven, A C, Trus, B L, Maizel, J V, Unser, M, Parry, D A D, Wall J S, Hainfeld, J F, and Studier, F W (1988) "Molecular Substructure of a Viral Receptor-recognition Protein", J. Mol. Biol. 200, 351-365.

Steven, A C, Kocsis, E, Unser, M, Trus, B L (1991) "Spatial disorders and computational cures", Int J. Biol. Macromol. 13, 174-180.

Trus, B L and Steven, A C (1981) "Digital Image Processing of Electron Micrographs - The PIC System", Ultramicroscopy 6, 383-386.

Trus, B L, Unser, M, and Steven, A C, (1988) "Effects on image classification by multivariate statistical methods of different normalization procedures and factorial representations", in Inst. Phys. Conf. Ser. No. 93, 3, Chapter 14, Proc. of EUREM 88, York, England, 367-368

Trus, B L, Steven, A C, McDowall, A W, Unser M, Dubochet, J, and Podolsky, R J (1989) "Interactions between actin and myosin filaments in skeletal muscle visualized in frozen-hydrated thin

sections", Biophys. J. 55, 713-724.

Unser, M, Steven, A C, and Trus, B L (1986) "Odd Men Out: A Quantitative Objective Procedure for Identifying Anomalous Members of a Set of Noisy Images Of Ostensibly Identical Specimens", Ultramicroscopy 19, 337-348.

Unser, M, Eden, M, and Trus, B L (1987a) "Unwarping of Slightly Distorted Periodic Structures Using Bidimensional Polynomial Representation," Signal Processing 12, 83-91.

Unser, M, Trus, B L, and Steven A C (1987b) "A New Resolution Criterion Based on Signal-to-Noise Ratios," Ultramicroscopy 23, 39-52.

Unser, M, Trus, B L, and Eden, M (1989a) "Iterative Restoration of Noisy Elastically Distorted Images", Signal Processing 17, 191-200.

Unser, M, Trus, B L, and Steven, A C (1989b) "Normalization Procedures and Factorial Representations for Classification of Correlation-Aligned Images: A Comparative Study", Ultramicroscopy 30, 299-310.

Unser, M and Eden M (1990) "Weighted averaging of a set of noisy images for maximum signal-to-noise ratio", IEEE Trans. Acoust., Speech, Signal Process., ASSP-38 (5), 890-895.

van Heel, M, and Frank, J (1981) "Use of Multivariate Statistics in Analyzing the Images of Biological Macromolelcules", Ultramicroscopy, 6, 187-194.

van Heel, M (1984) "Multivariate statistical classification of noisy images (randomly oriented biological macromolecules)", Ultramicroscopy, 13 165-184.

van Heel, M (1989a) "Classification of very large electron microscopical image data sets", Optik 82 114-126.

van Heel, M (1989b) "Comments on: 'Classification of Images of Biomolecular Assemblies: A Study of Ribosomes and Ribosomal Subunits of Escherichia coli'", J. Microscopy 159 113-117.

Wall, J S and Hainfeld, J F (1986) "Mass-mapping with the scanning transmission electron microscope", Ann Rev Biophys Chem, 15 355-376.