Scanning Electron Microscopy

Volume 3
Number 1 *3rd Pfefferkorn Conference*

Article 23

1984

Interactive Image Processing for Electron Microscopy: Matching Hardware with Software

W. O. Saxton
University of Cambridge

Electron Optical Systems (pp. 221-230)
SEM Inc., AMF O'Hare (Chicago), IL 60666-0507, U.S.A.

0-931288-34-7/84$1.00+.05

INTERACTIVE IMAGE PROCESSING FOR ELECTRON MICROSCOPY: MATCHING HARDWARE WITH SOFTWARE

W. O. Saxton[+]

Department of Metallurgy and Materials Science
University of Cambridge, U.K.

Abstract

The image processing techniques used 'a posteriori' to extract information from electron micrographs are surveyed, including particularly image averaging, selective averaging, 3-D reconstruction, and high resolution focal series restoration; recent developments in on-line image pick up and control have led to fully automatic focussing, stigmating and alignment by a frame store system equipped with a real time correlator board. The diversity of the techniques encountered calls for large integrated program systems with flexible command languages; however, a dilemma exists between providing the user with convenient control of special hardware facilities such as frame stores and array processors, and preventing the programs from becoming so specific that they are extremely short lived. Some of the compromises made in the Semper system are noted.

Key Words: Image processing, image analysis, image averaging, 3-D reconstruction, focussing, stigmating, alignment, high resolution, frame store, array processor

[+]Address for correspondence:
High Resolution Electron Microscope, Free School Lane, Cambridge, CB2 3RQ, England
Phone No.: (0223) 358381 Ext. 328

Introduction

Anyone concerned with image processing (enhancement or analysis) will be only too well aware of the exceptionally rapid rate of hardware development bearing on the field over the last five to ten years: quite apart from the general move towards 32 bit rather than 16 bit computers, and the widespread introduction of virtual memory systems, which have had a universal impact, the advent of 'frame stores'[*] and the growing capabilities of 'array' or 'vector' processors[†] have necessitated a constant re-evaluation of the answers to the common question, "What should I buy as a basic image processing system?".

Frame stores are most important primarily as a means of presenting digitised images in a visual form, being an order of magnitude better for the purpose than any preceding devices, but their former importance as a means of securing reasonably large amounts of inexpensive fast memory in a world dominated by 16 bit computers has already been superseded; they can now be seen as merely one member of the wide range of raster graphics devices which a generation nurtured on home computers now assumes to be present, in a memory-mapped form, in any computer; and it is instead such special capabilities as video-rate filtering and correlation that are now of interest to microscopists. Array processors, formerly hampered seriously for the purposes of image processing by their own limited memory capacities, are now overcoming the problem either through integration into large memory hosts or through massive increases in their own memory sizes.

The sentiment common some years ago that much of what we were doing should be postponed for a few years in anticipation of hardware developments was obviously well founded, and in many respects software is lagging well behind; yet development

[*] Devices characterised by the ability to store one or several image 'frames', roughly of TV quality, in a form that allows rapid input, output and manipulation of the stored image (from a few microseconds per pixel to real time TV rates) simultaneously with the continuous generation of a video signal from the stored image.

[†] Devices characterised by the ability to perform various arithmetical or logical operations on an array of operands at an abnormally high rate, through the use of various possible levels of parallelism.

is not slowing down, and the only sensible course for those of us concerned to use such systems is to continue to pursue the hardware with software systems flexible enough to be capable of substantial adaptation. This theme underlies the present review of image processing within electron microscopy: few workers in the field would pretend to be anywhere near the real front of image processing techniques in either hardware or software terms, and the most useful concern is with making the right compromises with relatively conventional approaches.

Survey of applications and problems

After these introductory remarks, an outline follows of the main reasons for the use of image processing in electron microscopy, and the remaining problems associated with some, with a few mathematical details being given in appendices; reviews in book form can be found in [19, 23].

Applications to scanning electron microscopy have largely been of a kind familiar to workers in several other disciplines, including particularly edge enhancement, particle sizing and counting, the related problem of automatic threshold selection, and image texture classification; earlier volumes of the present series contain numerous relevant contributions, to which we might add the UK Electron Microscopy and Analysis series, a recent thorough account of an earlier directional analysis technique due to B M Unitt [41], and an exciting new algorithm for automatic stereometry [3].

A problem common to all forms of microscopy is the reliable achievement of correct instrumental adjustment; the problem is of course most acute in the context of very high resolution (around 0.2nm) fixed beam (conventional) transmission electron microscopy (CTEM), where the image presented on the fluorescent screen is usually of insufficient intensity and/or magnification, and where the best-adjusted condition is in any case difficult to judge from the image appearance. The use of the new technology has recently made dramatic improvements in what is now possible, as described below under the heading of 'on-line processing'.

The sheer bulk of data involved when images are held in digital form has resulted in much effort being devoted in other fields to their compression by various schemes, which commonly reduces bulk for long-term storage by a factor of 8, at the price of rendering data not immediately amenable to display or manipulation. Within EM, R E Burge's group has used the discrete cosine transform as the basis of a very flexible scheme of this type [44]; more importantly, they have also addressed the question of the best image signal, or rather combination of signals, to use given the variety of signals possible in scanning transmission electron microscopy (STEM) by using principal component analysis (App.1) to identify objectively the most strongly differentiated (mutually uncorrelated) linear combinations [5]; this approach allows them at once to record the most informative images possible and to identify which data may be safely discarded, though it does not necessarily provide a simple recipe for image interpretation. Given the variety of signals possible in normal SEM practice, a similar analysis in this context also could well be useful.

My own enthusiasm and experience is centred on the techniques in use with CTEM - characterised perhaps by a relatively high reliance on transforms and large matrix operations (FFTs, special filters, auto- and cross correlation, clustering and 3-D reconstruction) as described below. (FFT = Fast Fourier Transforms)

Biomolecular applications. Micrographs of biological macromolecules remain extremely noisy because of the limited scattering power and radiation tolerance of the specimen, notwithstanding the effort devoted over the last decade to the alleviation of the damage problem, by very low temperature techniques for example [45]; the growing preference for glucose embedded or frozen hydrated preparations over the more artefact-prone use of negative stain will ensure that this remains the case, and as a result the averaging of large numbers of images of identical molecules is indispensable for the recovery of useful images.

Such averaging relied at first on perfect crystallinity in the specimen (usually spatial periodicity, but rotational and even helical periodicities have also been widely exploited), and used Fourier space filters with small 'windows' around the reciprocal lattice sites to suppress image components not sharing the periodicity of the crystal (e.g. [1]). Direct real space superposition was little used, though perfectly viable in most cases (fig.1) and rather simpler, until the more recent interest in averaging molecules in imperfect (i.e. bent) crystals [31,6] and isolated particles [11], the latter in particular due to the stimulus of J Frank. Here, the underlying common technique has been the use of cross correlation with an admittedly noisy reference [10,27] to allow accurate mutual alignment (registration) of individual images for averaging; in automated procedures, spurious correlation peaks due to random structure matching are largely eliminated by a lower threshold imposed on peak heights. The additional problem of unknown orientations in the isolated particle case is dealt with by comparing individual auto-correlation functions (ACFs), which are independent of image translation while preserving orientation and magnification information [11], or by locating individual centres of mass [26], or by simple (subjective) visual alignment [20]. Typically, many hundred images are combined in the case of imperfect crystals, and 20-100 in the case of isolated particles. In the imperfect crystal case at least, such averaging combined with extremely low dose STEM (annular detector mode) of freeze-dried material has allowed direct mass determination of a protein molecule and its major morphological subdivisions [8].

Increasing concern at the possibility that the population of molecules whose images are averaged may in fact be inhomogeneous has stimulated interest in selective forms of averaging. In the context of distorted crystals, restricting the average to the molecules in low

strain sites only has been shown to provide a slight improvement in image quality [30]. More excitingly, a different use of principal component analysis has made possible for stained specimens at least, the objective recognition of several distinct groups within a population (clustering of the individual images) [42]: in this approach, the most highly differentiated linear combinations of a set of images are determined and then used to assign individual images to classes on the basis of their degree of similarity to these (measured by the scalar product, sum squared difference or other measure). The direct handling of large numbers of images requires the inversion of large matrices, but averaging might be usefully extended over large populations on the basis of a classification deduced from a limited "learning" set.

The macromolecules studied by these methods are rarely very large relative to the resolution attainable, and a 3-D structure may be fairly easily obtained from a modest number (10-20) of 2-D (projection) images recorded in different directions through the specimen. Most work here (recently reviewed in [2]) has used the transforms of these projections to provide 2-D sections through Fourier space from which the full 3-D transform is built up, various schemes being employed for establishing the precise tilt angles recorded, the appropriate relative normalisation and registration of the projections and reliable sample values of the continuous 3-D transform, given the relatively noisy character that persists in the data, averaged or not. Most work has also been addressed to perfect 2-D crystals, but the averaging methods used for imperfect crystals have also recently been used on the basis for 3-D reconstruction [7,34] and a significant effort has also been devoted to isolated particles, particularly ribosomes, on both sides of the Atlantic [12,18] using other reconstruction methods reviewed in [15], such as filtered back projection, ART, and non-Fourier decompositions. When all the projections used in a reconstruction derive from one single molecule/particle, the final statistical significance is low; but averaging to provide better defined projections requires 'a priori' recognition of the different projections among a random population of molecules and the unambiguous determination of the corresponding projection directions, which continues to present a substantial challenge.

In all cases the interpretation of the 3-D density distribution is less than straightforward, the molecular boundary being invariably anything but sharp, and the problem is compounded for 2-D crystals by the impossibility of recording edge-on views, which leaves the mean density within each layer of the crystal quite indeterminate. The additional information that can be derived about surface profiles from metal-shadowed preparations is likely to prove increasingly important in the future therefore: a thin coating evaporated at an angle to the specimen plane reveals one component of the surface gradient (App.2), from which the actual surface profile can be reconstructed by a suitable Fourier plane filter [38], particularly if the specimen has sufficient symmetry to make recording the other component of the gradient unnecessary.

High resolution studies In studies of inorganic specimens, resolution is commonly of the order of 0.2nm rather than 2nm, and the demands made on image processing techniques quite different. Averaging many unit cells is not indispensable in this context, and it is not therefore universally practised, but it is still frequently helpful (fig.2) and is likely to be employed with increasing frequency. Variations such as averaging over 1-D lattices only are appropriate to materials with planar defects or planar interfaces between different crystalline forms such as are often found in nonstoichiometric materials. Special 'ad hoc' filters have been used to emphasise features difficult to see clearly in the untreated micrographs - e.g., boosting the super lattice reflections only in images of charge density waves (work in progress at J.W. Steed's group) to locate in an image the regions giving rise to particular diffracted beams, and to remove from an image strong crystal periodicities masking fainter local features.

Examination of the power spectra (diffractograms) of disordered image areas has long been used to evaluate instrumental defocus and astigmatism; the 'band-limit' found in these, i.e. the finest periodicity transferred from specimen to image, can also be used to estimate the two main illumination parameters, namely the effective focus spread and beam divergence [22]. A relatively sensitive band-limit estimate can be made by comparing two images recorded in succession, and tabulating in a radial correlation function [31] the level of mutual agreement found for progressively higher spatial frequency bands (fig.3).

The importance of such estimates for high resolution image interpretation lies in the substantial influence such factors have on image detail. Most practical evaluation of high resolution images, except in the very thin limit, has relied on image simulation techniques (reviewed by M A O'Keefe in this volume), in which the correctness of a structural model is tested by the extent to which the images it predicts match those recorded experimentally; even granted general readiness to discard all imperfectly stigmated images, the calculation of images for a reasonable range of specimen thickness, focus, focus spread and divergence is a formidable task, frequently revealing spurious 'matches' with experimental results, and any reliable observed values for some of these parameters makes the structural evaluation much less ambiguous. In this connection, the emerging realisation that small levels of residual beam tilt (around 1mrad) can affect image detail substantially without being detectable in a diffractogram is alarming [36], and this point is taken up again below.

Accurate imaging parameters are of course equally vital for direct image 'deconvolution' - the recovery of artefact free images from the aberrated images directly recorded. This is straightforward even in principle only for sufficiently thin and/or light specimens, which scatter only a small proportion of the incident electrons; in this regime the imaging is at least described by a simple specimen-independent spatial frequency response called a contrast transfer function (CTF), which can be 'divided out' of the

image transform. In practice, such a simple procedure is frustrated by substantial bands of low transfer - around the CTF 'zeros' - where division by the CTF results in noise amplification, but a more mature version, drawing on a focal series of images mutually aligned by cross correlation (App.3), and using a 'Wiener' approach to prevent noise amplification [35], provides a highly robust means of recovering faithful images at high resolution [28,16] (fig.4). Computationally, all that is involved is the calculation of a weighted average of the transforms of the individual images (the weighting ranging from point to point), and the only real difficulty of the technique - particularly in the application to images without disordered specimen regions present - is the accurate estimation of the parameters controlling the CTF from the image diffractograms. This technique also figures occasionally in biological applications, and is likely to be crucial to the use of frozen hydrated specimens.

Outside the weak specimen limit, such a variety of schemes was propounded during the '70s to little real avail that it is a matter of some embarrassment to recall much of it! Almost every conceivable combination of images and/or diffraction patterns, bright field, dark field or half-plane apertured has been proposed as the basis of specimen wave recovery schemes, but experimental difficulties, ill conditioning of the data or sheer inability to find the solution to a given set of nonlinear equations variously prevented their providing satisfaction to any but the mathematically inclined! (For full review, see [24].) Few schemes even considered the finite illumination profiles known to affect image quality dramatically, and the most recent [17] relies on an incorrect description of its effects; the most attractive option still open seems to be the subtraction from a bright field image of a corresponding dark-field image recorded with a central beam stop (App.4), for this procedure recovers a difference image exactly described by the (linear) theory applicable to the weak specimen case [25], and deconvolution can accordingly be carried out as described above for that case. At all events, there most certainly remain real problems in interpreting images outside the thin specimen limit; even simulation methods must be suspect beyond 10-20nm thickness, because of absorption effects not yet properly understood.

Real-time processing

The area that has seen the most dramatic impact of the new technology - particularly frame stores - is that of microscope/computer interactions [37,14,39,4]. The more obvious part of this impact has been in image acquisition and presentation: where the direct image is noisy and/or faint, the use of a frame store equipped with TV rate digitisation, the ability to average incoming frames (normally recursively, i.e. taking a weighted average of the incoming frame and that already stored), and present the result continuously on a normal brightness TV monitor overcomes completely the limitations of conventional systems relying on long persistence phosphor, high specimen irradiation levels or scan converter tubes. Microscope interfacing is relatively straightforward for scanning microscopes; but fixed beam instruments currently rely on a TV camera coupled to a fluorescent screen, which is at present the source of considerable loss of resolution, especially in high voltage instruments - a 2:1 or 3:1 tapered fibre optic coupler between the two may provide a simple answer. Given the low contrast of many high resolution images, a real time compensation for the TV camera shading pattern is also normally valuable; this facility depends on the frame store being capable of multiplying one frame by another also at video rates.

A rather less easily anticipated result has been the success of the recently developed procedures for fully automatic instrumental adjustment - particularly in high resolution CTEM applications, where the problems of focussing and stigmating have long been notorious, and those of beam alignment are in fact arguably worse still [36]. In contrast to previous approaches, which for SEM relied mainly on maximising the 1-D signal gradient (being therefore unsatisfactory in high noise situations) and for CTEM on visual assessment of diffractograms or ACF peak profiles, Erasmus and Smith [9] showed that in the (electronic!) hands of a computer the single parameter of image signal variance furnishes an adequate basis for accurate focussing and stigmating in both cases. In the system attached to the Cambridge University 600kV HREM, a minicomputer observes the signal variance at about 40 different values of a supply current in turn, finally selecting the optimum current by locating the required point (usually the centre) of the resulting curve. The variance is determined by using a hardware 'correlator' board to obtain the cross-sum between two successively digitised frames, and several readings are obtained each second. High accuracy is achieved even though the variance is not very sensitive to normal levels of astigmatism because the computer can make use of the higher sensitivity found far from the well stigmated condition.

Given the recent realisation of the importance of perfect beam alignment for high resolution microscopy, and the impossibility mentioned previously of diagnosing residual misalignment unambiguously 'a posteriori', the extension of the variance method to embrace the alignment process too (fig.5), giving a reliable procedure for adjusting all five supply currents automatically within a minute, in spite of their various interactions [33] is potentially equally important.

Any system for digital image acquisition from the microscope provides data in a form immediately suitable for any of the processing techniques described in the previous section, of course; however for CTEM the lower field of view possible with a TV camera (about 1% of the normal recorded field!) means that emulsions and microdensitometry (still surprisingly expensive) will remain with us for some time, even if we assume the widespread introduction of a 1024 square norm for future frame store systems.

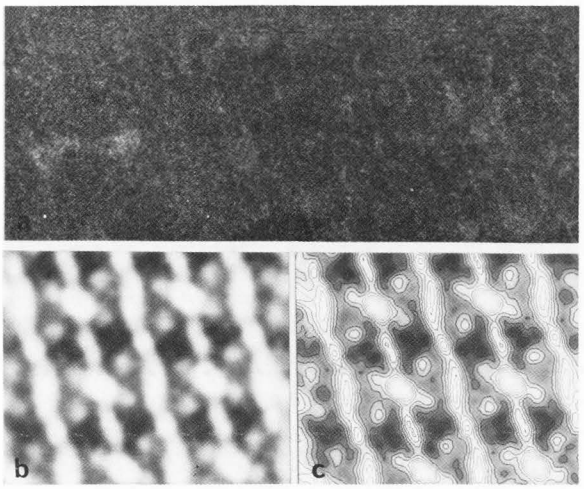

Fig.1 (a) CTEM image of negatively stained Chlamydomonas Reinhardi (due to P J Shaw); (b) average image obtained by superposing corresponding sites in real space, and centro-symmetrising the result by superposing it on a 180-degree rotated copy; (c) the same with density contours marked. (Unit cell dimensions 24 by 13nm.)

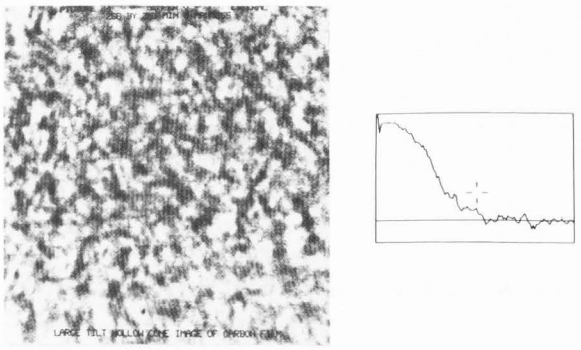

Fig.3 (Left) Photograph of TV monitor display of CTEM image of amorphous carbon film, recorded in the hollow cone illumination mode (due to D J Smith); (right) radial correlation function between the transforms of this image and another recorded immediately subsequently, showing mutual agreement between Fourier components to a period of 0.2nm (marked by the cross).

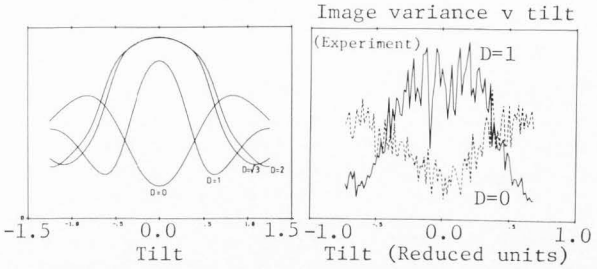

Fig.5 (Left) Theoretical dependence of image variance on tilt over a range of 1.5 reduced units (typically 9mrad) for three focus levels as marked in units of the Scherzer defocus; (right) experimental curves for a slightly lower range of 1 reduced unit, recorded via the 600kV HREM pick up system.

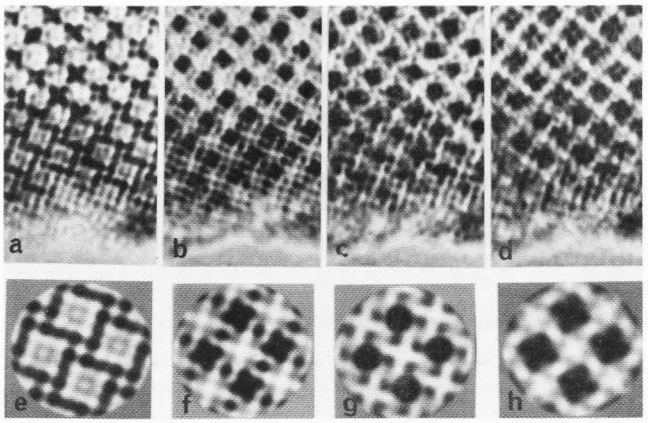

Fig.2 (a-d) CTEM images of a wedge shaped crystal of VNbO, at focus levels of 0.56, 1.67, 2.12 and 2.82 times the Scherzer defocus; (e-h) corresponding clear images obtained by local lattice averaging over the thin crystal areas and P4 symmetrisation.

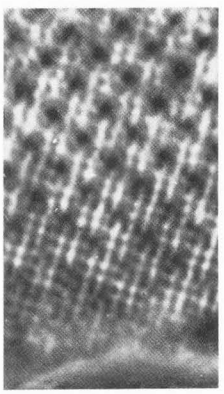

Fig.4 Restoration of focal series in fig.2, using linear method only, so that only the thin edge region is reliable. Tunnels are shown white.

Practical Systems for General Processing

The enormous variety of processing requirements implied by the survey above makes it clear that ultimately the only environment that really ensures adequate flexibility in storage and manipulation of data representing or derived from micrographs is the normal command level of an interactive computer terminal where commands can be entered in turn to apply particular operations to particular pictures, data types and file formats are largely at the user's disposal, and new operations can be defined in the conventional high level languages, such as Fortran*, and

* Is Fortran now in danger of being displaced from its position as the uncontested leader for what the industry calls 'scientific' applications? Perhaps because of the huge burden of Fortran's powerful input/output package more than any real deficiencies of the language, Pascal is increasingly preferred...

compiled to form new programs to be run by subsequent commands.

However, it takes a considerable time to master such an environment for the non enthusiast, and in any case it is in other respects than flexibility rather badly suited to our requirements, involving far too much typing for simple operations, poor inter-program communication, limited looping facilities and error handling, and encouraging the evolution of a plethora of unrelated data formats. It will also be obvious that most of the procedures described above require basic tools at a higher level than Fortran source language statements. For these reasons, almost all groups in the field have developed unified program systems of one kind or another, within which the control facilities approximate more closely to those actually required, and a variety of well established procedures can be applied relatively easily to data in a limited number of formats. The systems are modular, because modularity with well defined interfaces makes subsequent adaptation of the system to changing environments much more feasible. The discipline of writing programs within such a system is repaid directly by a relatively long program lifetime, by a common (and therefore more easily used and documented) user interface, and indirectly - but perhaps most critically - by the convenience of being able to use it in arbitrary (and easily redefinable) combinations with other system commands through the system's command processor.* Interacting with such a system, command by command, preferably with frequent display of intermediate results, is an efficient and satisfactory, if still imperfect, way of working.

J. Frank's account of his group's system SPIDER [13] includes a useful survey of systems suitable for EM image processing, many of which are already ten years old, which indicates reasonable success in keeping pace with technical development. Most frame store manufacturers offer subroutine libraries managing little more than their own peculiar features, though Gould DeAnza supply a somewhat more flexible program system called LIPS, with command processor support, in conjunction with their advanced IP-8500 display, and MicroConsultants offer a system called GPIPS which was developed from our own 'Semper' system. More recently, Logica UK have produced a relatively large (and expensive) image processing system called LUCID (formerly INSIGHT); and a comprehensive subroutine library - also unfortunately called SPIDER! - for memory resident 2-D arrays is available from the Joint System Development Corp., Tokyo (or from Mitsui in the UK at least).

* The usefulness of a program system is only sometimes measurable by how well it does the task it was designed for; in many other cases the relevant question is how easily it can be coerced into doing something different! The considerable progress that has been made was emphasised recently by my finding it possible to implement and test within a few minutes a proposed spatial frequency extrapolation algorithm that would have taken days to explore in the early '70s.

Semper. Our own system, Semper (also available commercially under a British Technology Group license), has been described previously in its Version IV form [29], and illustrates the theme of adapting to hardware changes in having moved easily from the 20KW PDP 8/E computer on which it was first developed to a 32 bit machine (a Systems Engineering Labs 32/27) without appearing grossly inefficient in the new environment [32].

Semper's basic control structure is a cycle of command decoding by an interpreter which relies (apart from a few 'intrinsic' commands) on a command syntax definition array that is updated as fresh commands are added to the system, and calls Fortran subroutines ('extension' routines) appropriately to effect the operations required by the commands encountered. The present extension routines comprise code for almost all the types of processing mentioned above, with the main exceptions of particle counting/sizing and principal component analysis. Within the current interpreter (version V-3), there is provision for command level branches, conditionals and 'FOR' loops, and for dynamically created and edited 'command procedures', as well as for the indirect execution of command sequences prepared outside Semper in ordinary text files. The interpreter manages a table of named numerical variables, which are largely at the user's disposal but which also serve as a means of program intercommunication; it also provides a unified image filing system embracing disc storage, tape and a display (not necessarily of the frame store type), so that commands and extension routines function in a largely device independent manner (e.g., 'Copy 3 to 4' makes a copy of disc picture 3 as a new disc picture numbered 4, while 'Copy 102 to display' causes the file 2 on tape drive 1 to be displayed, without the COPY routine itself being aware of the difference).

Image storage is organised in rows, the extension routines issuing (Semper) system read/write requests as necessary and operating with no more than a few rows available simultaneously; this allows the system to process large pictures on small address space machines - especially 16 bit computers - at some cost in terms of its speed in handling small pictures; we have taken the view that small pictures are handled sufficiently rapidly in any case and that it is the treatment of large ones that matters. Although on small machines the read/write requests result in actual physical transfers, efficiency is achieved in large machines through a 'cache'-like arrangement, Semper's disc input/output routine maintaining a collection of recently used rows in a large memory buffer with physical transfers made less frequently and in larger units. Even on large virtual memory systems such as the DEC VAX 11/780 this arrangement has proved quite successful: an address space of, say, 4MB such as might typically be available to a task under VAX/VMS is still inadequate for large images in floating point representation, and an intermediate buffering or mapping arrangement involving a disc file remains essential. A single large disc file is used with all pictures stored at convenient places inside it, a rudimentary directory being maintained by Semper; this choice - like the

decision to link the whole system as a single task rather than as a set of independent tasks run in response to requests from a central command processor task - was made in the interests of portability, a consideration which we have placed above all others, and is by no means essential; however, a very high level of portability has indeed been achieved, with installations on machines made by DEC, IBM, SEL, ICL, Prime, Data General, GEC, Nord and Apollo amongst others, and we are content with the compromise.

Display access presents a particularly difficult problem for a program system that is not to be too closely dependent on any particular display system. Our compromise at least has been to rely within Semper on no more than display erasure, image row output, text and line generation, a user-driven cursor and optional row input too for use with display devices of the frame store type; these functions are provided by 'primitive' routines written in Fortran-callable form to suit any particular display. Currently, all data are scaled suitably on output to the display device, and re-scaled to the original range on recovery, so that the display device can be used as a storage medium (albeit with limited precision), like some kind of visible disc. All other aspects of display manipulation* have been independently provided in our own installation (fig.6), via a keypad and trackerball continuously serviced by a small minicomputer dedicated to frame store management, or via a small host machine program that transmits instructions to the mini in response to single keystroke requests from a terminal† [32]. Different keys request increased or decreased contrast, brightness, upper or lower threshold; grey-scale wrap-round; contrast reversal; increased or decreased zoom; cursor movement or display scroll under trackerball control; alternative display modes for text and line graphics (stored in a separate memory plane in our case); switching between the two frames available; various levels of erasure; and the transfer of an image frame to or from the minicomputer system dedicated to the 600kV microscope. In this way, the display can be manipulated quickly and conveniently irrespective of the host program to which it is currently allocated. Anyone who has used a system where the frame store look-up-table (LUT), in principle always capable of virtually instant alteration, is in practice only alterable through the running of a program which loads a new LUT from a file will see what is meant in this context at least by matching hardware with software!

Likely Future Developments

Several improvements are likely within Semper itself, independent of any special hardware. Work is already in progress replacing the filing system with one supporting a much larger number of

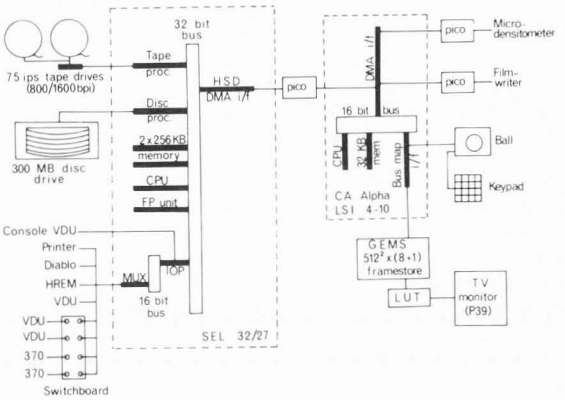

Fig.6 Block diagram of the HREM processing system: SEL computer, and GEMS frame store, with CA minicomputer between the two.

pictures within each physical device, each with a label including a comment or title string and other information (such as a 'protection' flag preventing accidental deletion/overwriting), with provision for multiplanar images in any of four different pixel representations, automatic conversion being performed as necessary by the system row access routines so that (broadly speaking) all routines will operate successfully on all representations. Provision is also being made for access to/generation of tape files in some interchangeable format independent of the Semper filing system, such as the FITS standard [43] - though even this does not propose any truly adequate mechanism for the transport of floating point values, and we are having to propose our own.

Ideas so far less well developed include providing an environment for some extension routines to operate on in-memory 2-D arrays containing the whole of a picture simultaneously (subject to a lower maximum picture size, of course), with automatic assembly/splitting of picture rows by the system as necessary - this seems a sensible move towards growing VM capabilities in future systems; also some of the simpler routines might be reorganised under an intermediate controlling routine that implements something similar to a UNIX 'pipe' (applying several commands in turn to each row of a picture, rather than applying each command to all rows before the next command is begun) - this would improve efficiency again [40], and allow immediate viewing of the cumulative result if the final destination was the display.

But what of more hardware-oriented developments? In spite of the flexibility of the display control mechanism described above, we are still not using all of the capabilities of our display system, nor have we made any provision for the use of an array processor (which we do not have!). The problem is of course that the better any software uses particular hardware, the less easily it is likely to be to transfer it to another environment.

As far as frame store exploitation is concerned, it seems simple and sensible to make use of a column access mode as well - a widely

---

* We have a GEMS Mk I, with a dual 512 by 512 by (8+1) bit configuration.

† It is remarkable how difficult it often is to achieve individual keystroke servicing in large computer environments, an added RETURN being all too frequently necessary.

available hardware feature; this could greatly simplify image rotation, transposition and warp correction, for example. By a somewhat less clear route, a mechanism is needed for controlling the use of multiple image frames - e.g., selecting individual planes, or sets of three for three-plane full colour images; the problem here is more one of how the user is to express his intentions in a convenient and memorable form than of actually achieving any particular desired goal, but this frequently occurring problem should not be overlooked!

As far as array processors are concerned, we can perhaps do no more within the context of a standardised portable package than to identify and isolate in subroutine calls vector operations that may reasonably be expected of an array processor; beyond this, adjustments of a very ad hoc character are to be expected, as the present range of architectures and interfaces is already rather wide. In our present context, by far the most useful single operation likely to be encoded in hardware (or at least moderately firmware!) is the 2-D FFT, used both by itself and as part of the process of auto- or cross correlation. Such a capability is already found in some array processors, and at least one frame store, and others will doubtless follow. Depending on precisely how the operation is defined by the hardware, however, we may have to fix or float an image before transformation, move it to a particular address, alter floating point formats, define an address range to be transformed, shift transform origins between corner and centre, pack or unpack real data in complex pairs, insert normalisation factors, and so on. An interesting account of the practical problems of using an existing array processor for FFT calculation appears in [37] (pp85-91). It is all ultimately worthwhile, as the anticipated speeding up is very considerable; but this seems to be the point at which we settle for short-lived - and necessarily home-grown! - software as we explore the new tools.

Acknowledgement

I acknowledge gratefully the support of a Royal Society research fellowship.

Appendix 1: Principal Component Analysis.

Given a set of corresponding pixels $x_1, x_2, x_3 \ldots$ from different images, the interdependence of the images might be quantified via the covariance matrix

$$v_{ij} = \mathrm{cov}\,(x_i, x_j)$$

(calculated by averaging over all pixels). The matrix v_{ij} is real and symmetric: its eigenvalues λ_i are therefore real, and we may find an orthonormal eigenvector set e_{ij}. If we consider now a set of new images, formed as linear combinations of the original images, with each pixel derived from the original pixels via the equation

$$x'_i = e_{ij}\, x_j$$

(i.e., if we decompose the original pixel vectors in terms of the eigenvectors), we find that the new covariance matrix

$$\begin{aligned} \mathrm{cov}(x'_i, x'_j) &= e_{ip}\, e_{jq}\, \mathrm{cov}(x_p, x_q) \\ &= e_{ip}\, e_{jq}\, v_{pq} \\ &= \lambda_i\, \delta_{ij} \end{aligned}$$

so that the resulting images $x'_1, x'_2, x'_3 \ldots$ are uncorrelated, and in this sense at least contain maximally different information. It also follows that λ_i gives the variance of image x'_i, so that the most important 'principal component' images are simply those which correspond to the largest eigenvalues. (The transformation from x to x' is commonly referred to as the 'Karhunen-Loeve' transformation [e.g., 21].)

Appendix 2: Interpreting thin metal coatings.

If a metal coating of thickness d is deposited uniformly at an angle θ to the specimen normal, on a specimen with a height profile h(x), then the coating thickness at x (projected normally) is

$$\begin{aligned} t(x) &= h(x + d\sin\theta) + d\cos\theta - h(x) \\ &= h(x + x_o) - h(x) \end{aligned}$$

if we drop the constant term and write x_o for $d\sin\theta$, which obviously corresponds to differentiation of h(x) if x is small. On Fourier transformation (denoted by tildes), we have

$$\begin{aligned} \tilde{t}(k) &= \tilde{h}(k).\exp(2\pi i k x) - \tilde{h}(k) \\ &= \tilde{h}(k).[\exp(2\pi i k x) - 1] \\ &= \tilde{h}(k).2\pi i k x \qquad \text{for small } k. \end{aligned}$$

Appendix 3: Linear image deconvolution.

Linear space-invariant imaging is described by a spatial frequency response - the ratio of an image Fourier component to the corresponding object component. In high resolution electron imaging of 'weak-phase' objects [e.g., 23], the 'object' is the projected specimen potential distribution $\phi(x)$, and the 'image' the image plane contrast, or fractional intensity variation c(x); and the spatial frequency response or 'transfer function' p(k) embraces the effects of illumination divergence and focus spread (spatial and temporal coherence), astigmatism, beam tilt, drift and vibration (all largely constants of the series) as well as defocus (deliberately varied from one image to the next). For any one such image, $\tilde{c}(k)/p(k)$ provides an estimate of $\tilde{\phi}(k)$, but the estimate will be poor where p(k) is low or zero; a better estimate is obviously provided by averaging the individual estimates, weighted relatively in some way that reflects their reliability; weighting in proportion to $|p(k)|^2$ gives the result

$$\begin{aligned} \tilde{\phi}'(k) &= \sum (\tilde{c}(k)/p(k)).(|p(k)|^2 / \sum |p(k)|^2) \\ &= \sum p^*(k)\tilde{c}(k) / \sum |p(k)|^2 \end{aligned}$$

which is also in fact the estimate minimising the summed squared image residuals; minimising the

summed squared object residuals, in expectation, gives the slightly more reliable Wiener solution instead, differing only in the addition to the denominator of an estimate of the spectral power ratio n(k) between image noise and object:

$$\tilde{\phi}'(k) = \sum p^*(k)\tilde{c}(k)/(\sum |p(k)|^2 + n(k)) \ .$$

Either of these prescriptions yields restorations with essentially flat spatial frequency responses except at very low and very high frequencies where none of the recorded images contains any reliable information. Linear imaging of a weak-phase weak-amplitude specimen producing a Fourier plane scattered wave $\tilde{f}(k)$, yields an image contrast

$$\tilde{c}(k) = \tilde{f}(k)t(k) + \tilde{f}^*(-k)t^*(-k)$$

in terms of a slightly different transfer function t(k); at least two images are now required for any estimate at all to be made of $\tilde{f}(k)$, but similar optimum restoration formulae can be found for exploiting a through-focal series of recorded images.

Appendix 4: Nonlinear image deconvolution.

When nonlinear image intensity components cannot be neglected, the transfer of object information to the image must be described by the equation

$$I(k) = \sum \tilde{f}(q)\tilde{f}^*(q-k)t(q,q-k)$$

in which $\tilde{f}(k)$ is the object transform and $t(k_1,k_2)$ a mutual transfer function, or transmission cross coefficient embracing all the effects noted in App.3 [e.g., 25]. This may be conveniently decomposed into three sets of terms, namely the background term

$$I_0 = |f(k_0)|^2 ,$$

the terms linear in f(k)

$$\tilde{I}_1(k) = \tilde{f}(k_0+k).[\tilde{f}^*(k_0)t(k_0+k,k_0)] + \tilde{f}^*(k_0-k).[\tilde{f}(k_0)t(k_0,k_0-k)] \ ,$$

and the remaining nonlinear terms

$$I_2(k) = \sum_{q \neq k, k_0+k} \tilde{f}(q)\tilde{f}^*(q-k)t(q,q-k) \ .$$

The troublesome terms are $\tilde{I}_2$; but they may be observed independently as the image intensity under central stop dark-field imaging conditions, and subtracted from the full bright-field intensity to leave purely linear terms which can be deconvoluted as in App.3.

References

1 Aebi U, Smith PR, Dubochet J, Henry C, Kellenberger E. (1973). A study of the T-layer of Bacillus brevis. J. Supramol. Struct. 1, 498-522.

2 Amos L, Henderson R, Unwin PNT. (1982). 3-D structure determination by electron microscopy of 2-D crystals. Prog. Biophys. Mol. Biol. 39, 183ff.

3 Atkin P, Smith KCA. (1983). Automatic stereometry and special problems of the SEM, in: Electron Microscopy and Analysis 1983, P.J.Goodhew (ed), Institute of Physics, Bristol. 219-220.

4 Boyes ED, Muggridge BJ, Goringe MJ. (1982). On-line image processing in high resolution electron microscopy. J. Microsc. 127, 321-335.

5 Burge RE, Clark AF. (1981). STEM multiple images: the Karhunen-Loeve transform and data compression, in: Electron Microscopy and Analysis 1981, M.J.Goringe (ed), Institute of Physics, Bristol. 315-320.

6 Crepeau RH, Fram EK. (1981). Reconstruction of imperfectly ordered zinc-induced tubulin sheets using cross-correlation and real space averaging. Ultramicrosc. 6, 7-18.

7 Engel A. (1984) 3-D structure of matrix porin. Ultramicrosc. 13, in press.

8 Engel A, Baumeister W, Saxton WO. (1982). Mass mapping of a protein complex with the scanning transmission electron microscope. Proc. Natl. Acad. Sci. USA 79, 4050-4054.

9 Erasmus SJ, Smith KCA. (1982). An automatic focusing and astigmatism correction system for the SEM and CTEM. J. Microsc. 127, 185-199.

10 Frank J. (1980). The role of correlation techniques in computer image processing, in: Computer Processing of Electron Microscope Images, P.W.Hawkes (ed), Springer-Verlag, Berlin. 187-222.

11 Frank J, Goldfarb W, Eisenberg D, Baker TS. (1978). Reconstruction of glutamine synthetase using computer averaging. Ultramicrosc. 3, 283-290.

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

13 Frank J, Shimkin B, Dowse H. (1981). SPIDER - a modular software system for electron image processing. Ultramicrosc. 6, 343-357.

14 Herrmann K-H, Krahl D, Rust H-P, Ulrichs O. (1976). Aufbau und Anwendung eines Digitalen Fernsehbild-Halbleiterspeichers fur die Elektronenmikroskopie. (Construction and application of a digital TV-halftone store for the electron microscope.) Optik 44, 393-412.

15 Hoppe W, Hegerl R. (1980). 3-D structure determination by electron microscopy (nonperiodic specimens), in: Computer Processing of Electron Microscope Images, P.W.Hawkes (ed), Springer-Verlag, Berlin. 127-185.

16 Kirkland EJ, Siegel BM, Uyeda N, Fujiyoshi Y. (1980). Digital reconstruction of bright field phase contrast images from high resolution electron micrographs. Ultramicrosc. 5, 479-503.

17 Kirkland EJ. (1982). Nonlinear high resolution image processing of conventional transmission electron micrographs. Ultramicrosc. 9, 45-64.

18 Knauer V, Hegerl R, Hoppe W. (1983). 3-D reconstruction and averaging of 30S ribosomal subunits of Escherichia coli from electron

micrographs. J. Mol. Biol. 163, 409-430; Oettl H, Hegerl R, Hoppe W. (1983). J. Mol. Biol. 163, 431-450.
19 Misell DL. (1978). Image Analysis, Enhancement and Interpretation. North Holland, Amsterdam.
20 Ottensmeyer FP, Bazett-Jones DP, Hewitt J, Price GB. (1978). Structure analysis of small proteins by electron microscopy. Ultramicrosc. 3, 303-313.
21 Rosenfeld A, Kak AC. (1976). Digital Picture Processing. Academic Press, New York. 109-123.
22 Saxton WO. (1977). Coherence in bright field microscopy of weak objects, in: Developments in Electron Microscopy and Analysis 1977, D.L.Misell (ed), Institute of Physics, Bristol. 111-114.
23 Saxton WO. (1978). Computer Techniques for Image Processing in Electron Microscopy. Academic Press, New York.
24 Saxton WO. (1980). Recovery of specimen information for strongly scattering objects, in: Computer Processing of Electron Microscope Images, P.W.Hawkes (ed), Springer-Verlag, Berlin. 35-87.
25 Saxton WO. (1980). Correction of artefacts in linear and nonlinear high resolution electron micrographs. J. Microsc. Spectrosc. Electron. 5, 665-674.
26 Saxton WO. (1980). Digital processing of electron micrographs: a survey of motivations and methods, in: Proc. 7th Eur. Cong. EM, Vol 1, P Brederoo and G Boom (eds) 486-493. 7th Eur. Cong. EM Foundation, Leiden, The Netherlands.
27 Saxton WO, Frank J. (1977). Motif detection in quantum noise-limited electron micrographs by cross correlation. Ultramicrosc. 2, 219-227.
28 Saxton WO, Howie A, Mistry A, Pitt A. (1977). Fact and artefact in high resolution electron microscopy, in: Developments in Electron Microscopy and Analysis 1977, D.L.Misell (ed), Institute of Physics, Bristol. 119-122.
29 Saxton WO, Pitt TJ, Horner M. (1979). Digital image processing: the Semper system. Ultramicrosc. 4, 343-354.
30 Saxton WO, Baumeister W. (1981). Image averaging for biological specimens: the limits imposed by imperfect crystallinity, in: Electron Microscopy and Analysis 1981, M.J.Goringe (ed), Institute of Physics, Bristol. 333-334.
31 Saxton WO, Baumeister W. (1982). The correlation averaging of a regularly arranged bacterial cell envelope protein. J. Microsc. 127, 127-138.
32 Saxton WO, Koch TL. (1982). Interactive image processing with an off-line minicomputer: organisation, performance and applications. J. Microsc. 127, 69-84.
33 Saxton WO, Smith DJ, Erasmus SJ. (1983). Procedures for focusing, stigmating and alignment in high resolution electron microscopy. J. Microsc. 130, 187-201.
34 Saxton WO, Baumeister W. (1984). 3-D reconstruction of imperfect 2-D crystals. Ultramicrosc. in press.
35 Schiske P. (1973). Image processing using additional statistical information about the object, in: Image Processing and Computer-Aided Design in Electron Optics, P.W.Hawkes (ed), Academic Press, London. 82-90.
36 Smith DJ, Saxton WO, O'Keefe MA, Wood GJ, Stobbs WM. (1983). The importance of beam alignment and crystal tilt in high resolution electron microscopy. Ultramicrosc. 11, 263-282.
37 Smith KCA (ed). (1982). Computer Techniques in Electron Microscopy and Analysis. (1982). J. Microsc. 127, 1-126. (Special issue.)
38 Smith PR, Kistler J. (1977). Surface reliefs computed from micrographs of heavy metal-shadowed specimens. J. Ultrastruct. Res. 61, 124-133.
39 Spence JCH, Disko M, Higgs A, Wheatley J, Hashimoto H. (1982). A digital on-line diffractometer and image processor for HREM, in: Proc. 10th Int. Cong. EM (Hamburg), Vol 1, 519-520. Deutsche Gesellschaft fur Elektronenmikroskopie e.V., Frankfurt.
40 Stevens WR, Hunt BR. (1982). Software pipelines in image processing. Comp. Graph. Im. Proc. 20, 90-95.
41 Tovey NK. (1980). A digital computer technique for orientation analysis of micrographs of soil fabric. J. Microsc. 120, 303-315.
42 Van Heel M, Frank J. (1981). Use of multivariate statistics in analysing the images of biological macromolecules. Ultramicrosc. 6, 187-194.
43 Wells DC, Greisen EW, Harten RH. (1981). FITS - a flexible image transport system. Astron. Astrophys. Suppl Ser. 44, 363-370.
44 Wu JK, Burge RE. (1982). Adaptive bit allocation for image compression. Comp. Graph. Im. Proc. 19, 392-400.
45 Zeitler E (ed). (1982). Cryomicroscopy and radiation damage. Ultramicrosc. 10, 1-177. (Special issue.)

Discussion with Reviewers

F. Lenz: In connection with automatized averaging procedures one might ask what the chances are that a given structure of interest is found even in a completely random intensity distribution.

There are cases where structural details are hardly visible in the original micrographs but come out clearly in processed averages. Would this also occur if one searches for such details in an intensity distribution known to be random?

I should like to suggest to use some "figure of credibility" defined as follows: Let us assume that a given structural detail is found N times in a micrograph. If the same structural detail is searched in a random intensity field of the same area, using the same criteria, and it is found n times, then we may define a figure of credibility e.g. by $\log \frac{N}{n+1}$.

Author: Thank you for your comment.