Notes on image mosaics

B. Vojnovic and P. Barber
Gray Cancer Institute

There are numerous instances when there is a need to acquire a large number of images and to 'mosaic' or tile them to form a larger, composite image. In bright-field or epi-fluorescence microscopy, for example, we may want to characterise a complete tissue section of dimensions of several tens millimetres, at high resolution. A simple calculation quickly reveals that this is not possible to do in a single image since the resolution afforded by a low power objective, necessary to view a large sample, is inadequate, even if we were able to afford cameras having tens of millions of active pixels. The only obvious and practical solution is to assemble the composite image from several images acquired at high magnification.

It sounds simple but this is probably one of the more difficult practical tasks in microscopy to undertake. In fact it is probably fair to say that no system has yet been devised that is able to perform this task easily, quickly and to routinely provide 'perfect' results. It is also probably fair to say that challenges presented are such as to force the developer into many interesting areas of image acquisition, mechanics, image processing and computing, areas that they had no intention in taking an interest in, since the problem is so 'obviously simple and straightforward' to solve: just overlap adjacent images 'a bit', cut them appropriately and 'join'. What could be simpler….and further from the truth!

Here we present some approaches used in our laboratory, which are restricted to the acquisition of images using bright-field (transillumination) and epi-fluorescence methods; while we do not claim that they are unique nor 'perfect every time', they go a long way towards routine operation. Perhaps the most useful aim of these notes is to warn the beginner of most of the traps one can fall into, as we have done (!). It would have been useful to have access to similar notes before we started getting involved in these procedures, but then again, had we been aware of all the difficulties at the start, we may have given up long ago.

Calibration of image imperfections

We bought a very expensive microscope and camera so the results have got to be perfect…..No, they don't have to be. First of all, all imaging systems suffer to a greater or lesser extent from barrel or pincushion distortion, or combinations thereof. The camera (or other e.g. scanning imaging device) produces an image on a rectilinear 'pixel' grid, but the optics will not produce a perfect, grid-like image of a regular array (Figure 1). The greater the camera's pixel resolution, the less likely can image adjacent images be matched down to the last pixel.

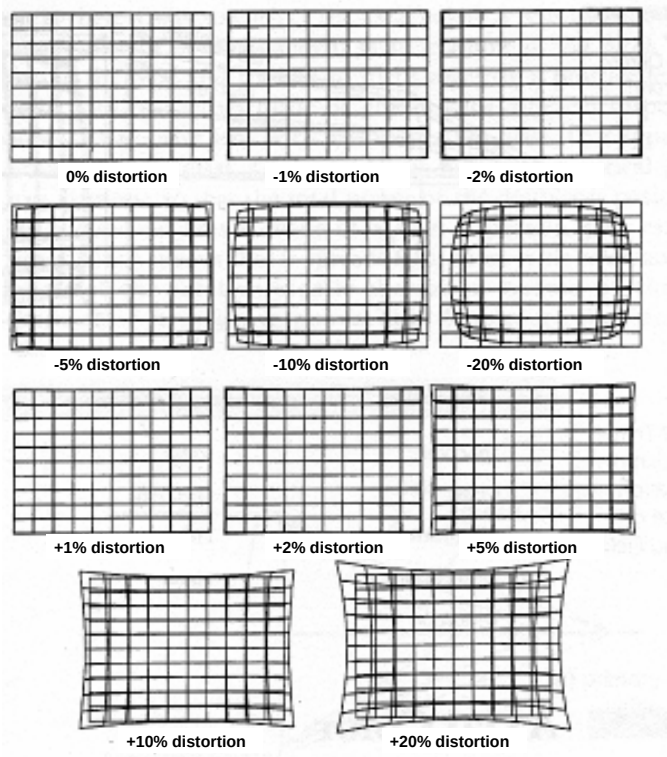

Figure 1: Examples of increasing amounts of barrel (negative) and pincushion (positive) geometric distortions in an imaging system

It is straightforward to correct this type of distortion if the shape of the distortion function is known. For simplicity, and in practice this is more than adequate, a symmetric distortion function can be assumed (i.e. symmetric about the centre of the image). The issue then is how to determine the shape of the function. One method is to 'map' the distortion by moving a point-like light source across (and up and down) the microscope's field of view (e.g. see accessories for fluorescence light microscopy). Another, rather quicker method is to use a regular line pattern and assume that the distortion is centro-symmetric (i.e. the same in X and Y directions away from the centre). A convenient means is to use a Ronchi grating slide of pitch appropriate to the objective in use. Suitable Ronchi gratings can be obtained from Edmund Scientific, as part numbers:

10 line pairs mm^{-1}	part #38-258
20 line pairs mm^{-1}	part #38-259
40 line pairs mm^{-1}	part #38-260
80 line pairs mm^{-1}	part #38-261
100 line pairs mm^{-1}	part #38-562

Once images of these are obtained (one for each objective/magnification), image processing and curve fitting procedures can be used to determine the appropriate functions. The underlying assumption is that the camera/imaging system has been appropriately aligned to the microscope's output port. It is essential to do this in any case to minimise required image 'overlaps' as will be described later.
Here are the tools we have developed to perform these corrections.

The 'Distortion correction panel' allows us to obtain the image of the Ronchi lines, ('Acquire Image', Figure 2) A new image panel then appears together with the 'Profile panel'.

By clicking and draging the mouse on this new image panel a red line will appear; this can be dragged to one corner of the image. The red line will have blue crosses along it where black lines have been detected. On the Profile panel, the red line shows the image intensity profile along the red line on the image. This profile is thresholded at the value shown by the white horizontal line (which can be dragged up and down), the result is the blue line. A cross is placed at the centre of each region below the threshold (above if bright lines are being detected).

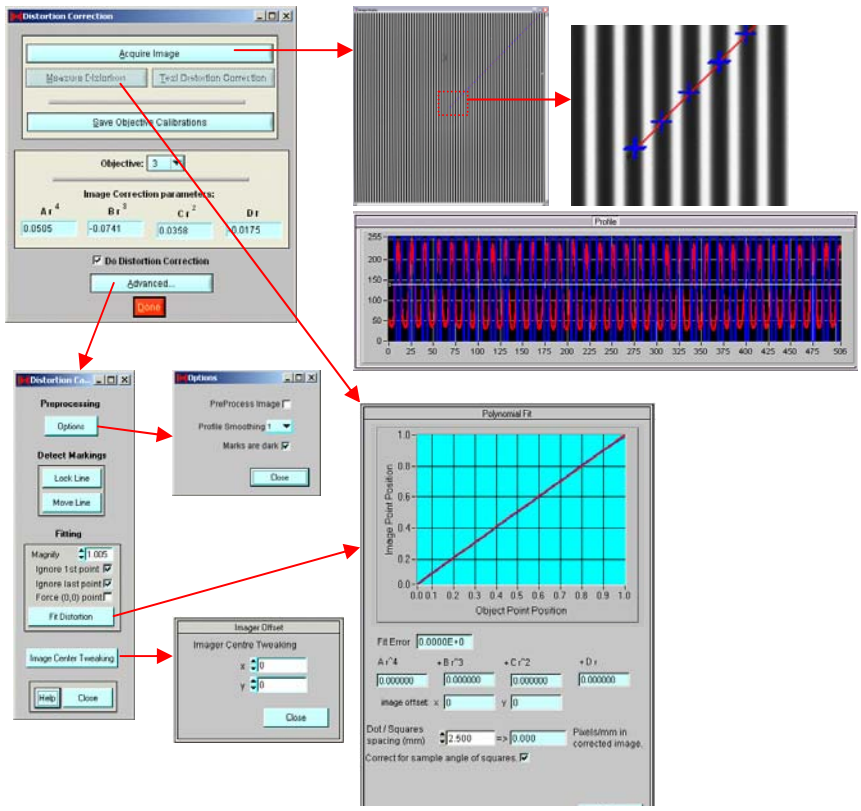

Figure 2: Procedures for using a Ronchi grating to correct for imaging optics geometric distortions.

By clicking the 'Measure Distortion' button a polynomial fit panel appears showing how the crosses are distributed along the red line and hence how far along the line (image point position) each cross is compared to a set of equally spaced crosses (as we know the object point positions are equally spaced on the Ronchi slide). The poynomial fit can then be saved ('Save objective calibrations') or the result of the procedure can be examined ('Test distortion correction') to view an image after the correction has been applied.

'Advanced' calls up an additional panel where we can change the pre-processing on the image, the profile, or make the system detect bright lines rather than dark. 'Lock line' just locks the red line on your image so that it is not accidentally changed, and 'Move line' frees it again. We can modify the calibration result by changing 'Magnification', this should be set fairly close to 1.0 for good imaging systems, but if the result of distortion correction has black bands at the edges of the image, we can increase this value to reduce these bands. Sometimes the system cannot properly detect the black line at the centre of the image or lines near the edge, and we can then use the 'Ignore 1st point' and 'Ignore last point' options to disregard these. Sometimes, it is also better to force the fitting to use the (0,0) point, this can be forced on this panel. Ideally, a black line should fall on the image centre, if it does not you can 'tweak' the effective centre of the image by a small amount to obtain better results. This is not critical, and only affects the very centre of the image but sometimes it can give a poor fit.

It is also highly desirable to determine a reasonably accurate pixel scaling factor, i.e. pixels/micron scalings. Although these can be calculated, there are a number of uncertainties associated with image magnifications in most microscopes and it is easiest to measure the scaling factor by using a calibrated scale slides. There are numerous suppliers of these; we use parts manufactured by Pyser-SGI Ltd:

5 mm length, 50 micron steps	02A 00401
reflective 1mm length, 10 micron steps	02B 00421
50 micron squares	02A 00406
10 micron grid	02B 00429
S30 Halton grid finder	02A 00413

Dealing with motorised stage errors

It is assumed that some form of motorised stage will be used to acquire image mosaics, or at least a stage with a readout system, so that the relative coordinates of each of the composite images are known. Although modern stages are very accurate, it is invalid to assume that they are perfect and certainly not to the nearest pixel. By far the most commonly used type of microscope stage is a stepper-motor driven stage. Typically these use a precision leadscrew with a pitch of 1 mm (though many similar variants are possible), driven by a stepper motor which can be moved in steps of around 1/10 000 of revolution. The resolution is thus of the order of 100 nm. The absolute accuracy is however determined by the leadscrew non-linearity, by other mechanical imperfections, by the fact that stage platters are driven from the side (linear bearing twisting), by 'lost' steps etc.; to get to better than 1 micron absolute is very difficult. More typical absolute accuracies are of the order of 3-4 microns, particularly where a stage has been extensively used, with inevitable slightly worn mechanical components. It should thus be obvious that we cannot rely on sample positioning systems to provide perfectly tiled images, unless the lowest magnifications are used, providing fields of view of the order of a few millimetres per image. Any form of XY tiling must thus rely on making use of information from the image itself.

Microscope stages with built-in position encoders are available from a number of manufacturers; these operate in a closed loop or feedback mode and, in principle are more accurate. However, it is unwise to fall into the trap of believing that the presence of an absolute or relative position encoder is the solution to all the above problems. However good the encoder is, it almost never senses the position of the sample, but rather senses the position of an axis parallel to, and some distance away from the sample. This is an inevitable consequence of the fact that biological microscope stages must have a 'hole' in the middle as the ability to perform transillumination imaging is considered essential. This means that we are at the mercy of the quality of the linear bearings: any twist introduced by the fact that the stage is driven from the side means that the encoder does not reflect the true position of the sample. As far as 'lost' motor steps are concerned, the presence of an encoder is definitely beneficial. Similarly, long-term accuracy and repeatability is superior. But the absolute accuracy is almost never better than around +/- 1 μm. There are definite advantages when high slewing speeds are used (i.e. when there is the greatest probability of loosing motor steps, but again, it pays to evaluate the performance carefully. A not uncommon problem is that the feedback loop time-constants may not always be optimised for the particular load or changes in mechanical performance etc. A poorly adjusted feedback loop may result in a performance far worse than that of an open-loop system. In our opinion, a dc-motor-driven stage, with a feedback encoder is a good choice, though this is not readily available from most suppliers: one reason is that the proportional-integrative-derivative terms are not readily adjusted; the more cynical view is that expertise in understanding and setting up feedback loops (i.e. sound, old-fashioned control engineering expertise!) seems to be on the decrease. The trend towards implementing the feedback loop using digital techniques is of course highly desirable, but then you need to be a good machine code programmer as well as a good control engineer – an even rarer breed! With digital, i.e. sampled, control loops, the key specification is the open-loop response speed, i.e. the position sampling rate. This is usually in the kilohertz range and often barely adequate for the fastest moving stages. A few (very few) suppliers offer devices operating at sampling rates above around 10 kHz.

If biological microscope stages did not have holes in the middle, it would all be so much easier! Indeed, feedback stages used by e.g. the semiconductor industry, where a solid stage can be used, are far better. These are driven - and the position sensed – truly in line with the sample. Since the position encoder is then placed directly below the sample, the true position is always known. An interesting possibility thus suggests itself for biological microscopes: using two encoders, placed symmetrically on two opposite sides of a driven axis. Any twist in the stage would result in one of the encoders reading too low, the other too high: take the average and, hey presto, you have the true position in the middle, where your sample is. To the best our knowledge, no one has developed such a stage (perhaps more correctly, no one has dared to develop such a stage, though it was suggested some years ago by B.D. Michael at the Gray Cancer Institute and has remained a challenge to our group!). If only there were more hours in the day, what an interesting project that would be….

A related problem is associated with z-positioning, i.e. focus. When imaging a large sample, it is most unlikely that all of it will be in focus, i.e. in the same z-plane relative to the objective. The only correct solution to this problem is to perform some form of autofocusing on EACH of the images to be tiled. While this is possible, the penalties in terms of acquisition time, and in the case of fluorescence imaging, of sample bleaching, make this approach rarely acceptable. There are other crucial issues associated with the particular autofocusing algorithm used that are discussed separately.

The other 'extreme' approach is to assume that the sample is inherently 'flat' but that it is not necessarily parallel to the stage motion XY plane. Here all we need is to focus on three images, arranged in a triangular pattern near to the sample edges. An arbitrary planar surface can be then defined and the most appropriate 'focus' position can be readily calculated for all intermediate images. This approach actually works rather well for moderate magnifications, down to fields of view of the order of around 250 microns per constituent image. A better, though more time-consuming approach is to define a curved surface with more than three 'focus' points and use e.g. a second or third order

polynomial to define the surface. In practice of course other considerations may dictate that the total number of images in the mosaic be minimised and hence a 'good' starting point is the use of a 20x objective. The depth of focus of most 20x objectives is appropriate for this task.

Ensuring appropriate illumination

In most microscopes it is almost impossible to ensure a completely flat illumination field. Even when the illumination source is perfectly aligned, vignetting errors, the need to ensure that only the imaged area is illuminated etc. all conspire to result in reduced illumination intensities near the edge of the image. When eyepieces are used, this is of little consequence, but the true meaning of 'tiled' images becomes all too readily apparent when a 'raw' images sequence is acquired (Figure 3).

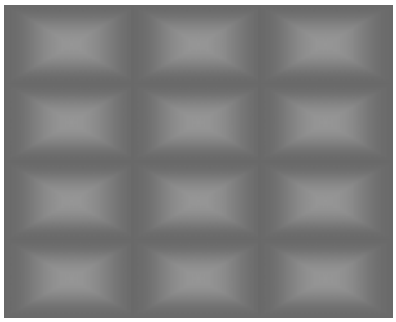

Figure 3a: Example of tiled sequence of images where illumination at images edges has not been corrected.

It is essential to normalise each image in the sequence with respect to the actual illumination profile, or more correctly with respect to the illumination surface. Even an empirical approach with a say a second-order polynomial is better than no compensation, although best results are obtained by acquiring the actual illumination profile. In the case of bright-field illumination, this is straightforward: one simply acquires a 'white' normalisation image from a clear portion of the slide.

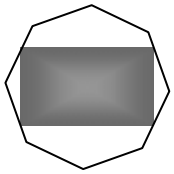

Figure 3b: Optimising the illumination aperture to 'just' cover the imaged area.

In the case of epi-fluorescence microscopy, an evenly fluorescent sample is needed, for example as described in 'Useful accessories for quantitative fluorescence microscopy'. In both cases, it useful to 'remove' minor, localised variations in intensity due to dust, debris etc. from the reference images by applying a median filter of appropriate width to the image. Clearly, correct lamp alignment is essential to minimise variations in the first place.

A related problem in the case of epi-fluorescence microscopy is associated with the illuminated sample area. If this is too great, i.e. where regions of the sample, which are not imaged at a given time, are illuminated, excessive bleaching will result and will become apparent in the 'next' image in the tiling process. Ultimately, there is no solution to this, other than minimising illumination intensities as much as possible by using the most sensitive affordable camera. Minimising exposures is of course always desirable.

The above comments relate to acquisitions with a modern monochrome camera. When colour cameras are used, the problems are significantly greater. Indeed, since most fluorescence imaging requires the sensitivity of a monochrome camera, we consider ourselves fortunate in not being able to afford a top-of-the-range high-sensitivity 3-CCD camera! Fluorescence illumination systems on most microscopes are not achromatic and minor variations in illumination profiles in different spectral bands are inevitable: strictly speaking we would need to acquire 'red', 'green' and 'blue' illumination profiles and use these to process the camera's three channels separately. For transillumination work (e.g. histology), there is no problem in acquiring the reference images. Nevertheless, it is advisable to appropriately set the camera's shading controls, particularly where dichroic-split 3-CCD cameras are used.

Nevertheless, there are instances where a 'colour' fluorescence image is required and this is usually accomplished by merging 2 or more monochrome images, each acquired with a different filter set. In those instances, it is definitely advisable to use separate illumination profile images for the different wavelength bands. Any chromatic variations in the illumination will thus tend to cancel out.

A final issue in fluorescence work is lamp stability. By their very nature, arc lamps are fundamentally unstable sources. In the short term, particularly following switch-on, enough time must be allowed for the lamp/housing to reach thermal equilibrium before reference images are acquired. In the medium-to-long term, the lamp electrodes wear out, most often with a consequent increase in arc length and definitely with a change in position. For this reason, quite apart from the fact that it is probable that someone may have been fiddling with the lamp alignment controls (it is always someone else!) it is recommended that a set of reference images is always acquired for a particular experimental session.

For fluorescence work, a fast-acting excitation shutter is almost always essential. An example of the arrangement we use in conjunction with Nikon microscopes is shown in figure 4. Note the coupling component between the shutter itself and the lamp bayonet fitting: it is essential to ensure appropriate ventilation in that area to dissipate the heat from the lamp and from the optical beam. The shutter must be synchronised with the image acquisition to minimise over-exposure of the sample. Shutter opening/closing times of a few milliseconds are readily achieved (see notes on 'Making your own automated microscope'); we use the Uniblitz range of shutters made by Vincent Associates.

Figure 4: Fast optical shutter inserted between mercury arc lamp and microscope epi-illumination optics

An improved arrangement would be to separate the lamp from the microscope altogether and dispense with free-space optical coupling in favour of fibre-optic coupling. Although we have not often used this method, largely because of cost and light loss issues, it should have significant advantages and would ensure stable and consistent alignment between fibre and microscope. The way light is launched into the fibre (typically 1 mm diameter) is of secondary importance since the light is 'scrambled' in the fibre. Any lamp-fibre alignment issues will tend to cause a change in intensity (rather than changes in the observed intensity profile). Of course a second set of UV-transmitting optics is required between lamp and fibre and light loss is inevitable. At some point in time (i.e. when we can afford it!) we intend to install such a system. Assemblies are available from Technical Video Ltd.

Selecting the imaged area

Modern microscopes are capable of producing excellent images over increasingly larger fields of view. Nevertheless, it is not advisable to use the 'full' available field of view: the greater that is, the greater will be the geometric and intensity corrections needed. In practice, we have never been practically successful in routinely utilising the full field of view. A more appropriate solution is to dispense with camera coupling optics altogether and use a large format imaging chip (e.g. 2/3", 6.6 x 8.8 mm; 11 mm diagonal). At the very least, try to minimise the de-magnification in the camera coupling optics: most image distortions tend to be introduced by these systems.

Selecting the image overlap area

This clearly depends on the accuracy of the stage and the degree of distortion in the imaging system. On the one hand, as the image overlap is increased, it will be easier (but more computationally intensive and time-consuming) to perform accurate edge registration and correlation. The increase in acquisition time is due to the fact that more images are needed to cover the sample area; computation time is also increased since more pixels are utilised in the correlation calculations. On the other hand, the smaller the overlap, less image edge information is 'used' and the success of the correlation is more dependent on heterogeneity of image features at the edges. A reasonable compromise is to use an

overlap of typically less than 10% of the image linear dimensions, though larger overlaps are preferred, particularly where distortions associated with coupling optics are to be avoided.

How many images and where to store them?

This clearly depends on the total tissue area that needs to be imaged and on the pixel resolution of the camera, as well as the required optical resolution. Megapixel cameras are recommended and reasonably commonly available, so a 1024 x 1024 pixel image would be considered typical. A tissue section may be around 20 x 20 mm and if we need to resolve submicron features, where say every pixel represents some 500 nm, each image in the mosaic would represent 500 x 500 microns; some 40 x 40 images are thus needed, i.e. we need to cater for storing and tiling 1600 images. This is clearly at the upper end of the scale of what is possible, since 1600 images would require around 2 GigaBytes of uncompressed storage if acquired at 8 bit intensity resolution, or twice that if we were interested in intensity details requiring 12-16 bit resolution.

It thus pays to use common sense rather than the 'sledgehammer' approach. The resolution of the objective needs to be considered so as to minimise 'empty' magnification. A simple examination of the available numerical aperture of the objective tells us what resolution is achievable: it is not uncommon to then find that far more pixels than needed are used and that all we will get is a less pixelated but still blurry image. If you want a pretty picture, that's fine; if you are after information, you will be wasting resources. In the above example, simply relaxing the resolution to 1 micron reduces the memory requirement 4 fold, i.e. to 400 images.

Speed of acquisition

Acquisition speed is primarily determined by the imager sensitivity and the intensity resolution required, as well as by the stage motion time when moving from frame to frame. These features are determined by hardware and are essentially 'fixed'. The other main factor determining speed is the extent of image processing performed 'on the fly', i.e. as the images are acquired. In practice, this can be made arbitrarily fast; indeed the other extreme is to eliminate all processing and accept that image corrections and normalisations are performed in a separate, post-acquisition phase. Conversely, in some cases, it is essential to process during acquisition if one needs to 'capture' as fast as possible to minimise diffusion of fluorophores, for example. Nevertheless, it is still useful to check that the appropriate image correction routines have been successfully applied by building up an approximately 'correct' mosaic during the acquisition process.

So let's see how the numbers pan out in practice. When we have plenty of light, i.e. when dealing with bright-field images (or intense fluorescence images), a typical camera exposure time (integration time) of a few milliseconds would be used and the acquisition time would largely depend on the image transfer time (time required to transfer a given image from camera to PC). This in turn depends on the type of camera interface and the operating mode of the camera. A commonly used interface is the IEEE 1394 ('FireWire') standard which is limited to some 10 frames sec^{-1} for megapixel images. The maximum sustained throughput with such an interface is 100 Mbits sec^{-1}; a common type of imager chip (e.g. as used in the popular Hamamatsu ORCA series of cameras) provides some 1.4×10^6 pixels. This is 11.2 Mbits per image, and will take around 100 ms to transfer.

If we need to integrate for longer (integration times of 100 – 1000 ms are not uncommon when dealing with low-light-level fluorescence) this must be added to the image transfer times. When using integration times longer than a few milliseconds, it is far better to use a camera that can be triggered externally, preferably through a hardware connection. The usual approach is to software-trigger the shutter, which, when fully opened, in turn triggers the start of camera integration. When that is complete, the camera notifies the software, this closes the shutter and the image transfer process is

started. During that time, the shutter power supplies are recharged, ready for the 'next' image and, if at all possible, the stage is advanced to the 'next' position.

The fact that several independent, but ultimately synchronised, processes should take place leaves us with a slight dilemma: a simple, sequential approach to the programming is straightforward to implement but is rather inefficient in terms of speed: while, say, the stage is moving, the software is 'waiting'. A multi-threaded approach to the programming is far better, though considerably more complex to implement. With that approach, several operations can be started simultaneously and the slowest one limits the overall speed.

For example, if we wish to acquire low pixel number images, image transfer times will be short and stage moves (or shutter recharging times) will dominate. Larger images will take longer to transfer and the stage will be 'ready' and positioned to the next sample area within the transfer time. In general it is a good idea to time the processes carefully, or at least analyse/predict the times before final software design is started. A continuously evolving 'problem', or rather issue, is caused by the fact that computers and associated peripherals are becoming faster all the time: for example fast graphics cards, the use of PCI bus architectures and similar improvements have made the speed of image display and storage a relatively trivial problem, while in the past, it was a major issue. It thus pays to keep a good timing record of the different processes to see if and where changes are required to make use of the latest technology. In particular, multi-threading programming approaches can be useful here, but it is always necessary to evaluate the advantage introduced by such rather more complex programming techniques.

Ultimately of course, it is the mechanical components that will determine the maximum throughput. Up to now, we have discussed stages and shutters briefly, but there is a further section to consider when acquiring fluorescence images of multiple dyes, requiring the use of more than one cube/filter set. It is clearly preferable to acquire images of all the spectral bands of interest at a given stage position: we can thus be sure that they will be co-registered. However, the consequence is that the acquisition process will be painfully slow, since in most motorised microscopes, it may take a second or two to change cube positions. If we have to acquire a hundred images or so per spectral band, the cost in coffee and biscuits while we are waiting can become excessive. At co-localisation resolutions comparable to stage accuracy, a better approach is to scan the complete sample with one cube, followed by subsequent scans with the other filter cube(s). However, this does raise post-acquisition issues, particularly with registration of composite images that have been assembled from different features in the constituent images.

Now we come to the stage. The time required for the stage to move from frame to frame is defined by its acceleration/deceleration, its speed and by any anti-backlash motions. XY stages are reasonably fast and typical parameters are: acceleration: 50 000 mm sec^{-2}; speed: 50 mm sec^{-1}. Now we have to go back to school and remember Newton's laws of motion, or if we can't find our old textbooks, the following table may be useful:

Newton summarized, for people where apples have caused damage to heads

Steady state motion (constant velocity)		
To find	*As a function of*	*Use:*
x (position)	v (velocity) and t (time)	x = v.t
v (velocity)	x (position) and t (time)	v = x/t
t (time)	x (position) and v (velocity)	t = x/v

Accelerated motion (constant acceleration)		
To find	*As a function of*	*Use:*
x (position)	a (acceleration) and t (time)	$x = a.t^2 / 2$
x (position)	v (velocity) and a (acceleration)	$x = v^2 / 2a$
v (velocity)	a (acceleration) and t (time)	$v = a.t$
v (velocity)	x (position) and a (acceleration)	$v = (2.a.x)^{1/2}$
a (acceleration)	v (velocity) and t (time)	$a = v / t$
a (acceleration)	x (position) and t (time)	$a = 2.x / t^2$
a (acceleration)	v (velocity) and x (position)	$a = v^2 / 2.x$
t (time)	v (velocity) and a (acceleration)	$t = v / a$
t (time)	x (position) and a (acceleration)	$t = (2.x / a)^{1/2}$

In most cases, acceleration and deceleration will be equal and a trapezoidal velocity profile is used (Figure 5); thus the total travel time will be: $T_{CV} + T_{DEC} + T_{ACC}$.

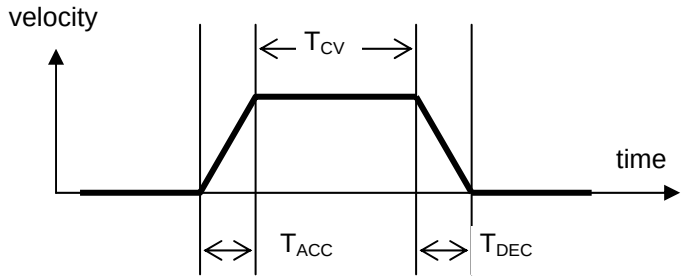

Figure 5: Example of velocity profile associated with stage XY move, from image frame to next image frame.

Focusing

The crucial question related to focusing is how to do it – manually or automatically? Manually is clearly acceptable but perhaps not so elegant and not always either perfect or practical. The automatic approach seems superior but is not always straightforward or indeed fast: the processor we have between our ears is very good at optimisation procedures! Any automatic system must start with a reliable focus 'measure'. We have examined numerous approaches but there is no question that the most reliable measure is based on exactly on what a 'focused' image represents: maximum high spatial frequency content. Performing a Fourier transform on the image and examining the ratio of high to low frequencies is thus the most reliable approach. Indeed it is useful to use this routinely to provide a 'focus indicator' even for manual operations.

With analogue output cameras, an alternative to Fourier analysis is possible: remove synchronisation signals from the video signal and filter into high and low frequency bands, determine the relative powers, then ratio the results so as to make the measure independent of intensity. All this is easily achieved in hardware with a handful of chips. It is indeed unfortunate that there is no access to the analogue CCD signal from modern digital cameras, forcing us to perform the spatial frequency analysis in the digital domain; while there is nothing significantly hard in this, it does mean tying up the host processor/image acquisition system for this purpose.

Whatever method is used, it is advisable to ensure that the mechanical coupling between the focus drive and the microscope is as rugged as possible, even if this means dismantling and modifying the microscope. Some manufacturers of motorised focus accessories provide a friction coupling between the drive and the microscope's focus knob, often using 'O' rings. Although this seems at first sight to be a straightforward approach which ought to work well on 'scopes with very low friction (i.e. low torque) focus knobs, in practice it does not. There are few issues that have caused more problems in terms of time wasted or unfair criticism of focus algorithms: backlash, 'lost' motor steps and cumulative errors all contribute to make this approach unreliable. A better, and thoroughly reliable way is to couple the drive directly, as shown in Figure 6.

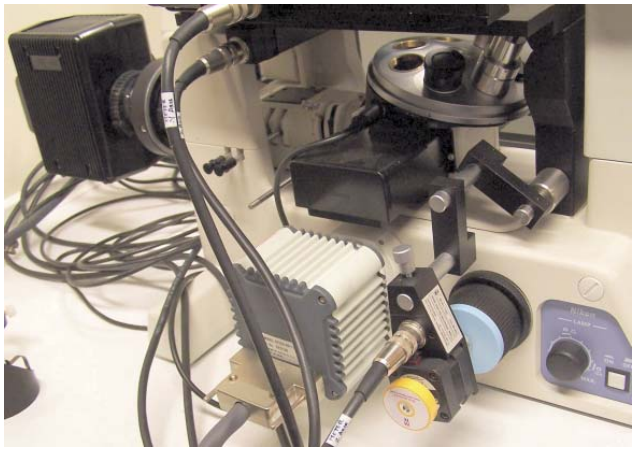

Figure 6: Example of coupling of stepper motor to the 'fine' focus shaft of a Nikon TE300 microscope, using a 'bellows' coupling. Appropriate degrees of freedom are provided on the motor support to ensure that 'scope and motor shafts are co-linear and concentric.

Let's get back to determining the focus measure. While computing speeds will inevitably become faster, it is useful to minimise the computation times as much as possible. We have found that using the central portion of the image is generally more than adequate, and we thus 'cut' a 2^n a 2^n region from the image (256^2, 512^2 or 1024^2) to allow efficient Fourier transform calculations using the Fast Fourier Transform approach. The image information is processed to derive relative high spatial frequency content: this is indicative of image sharpness and hence a focus indicator can be obtained from this measure. The focus measure is derived by performing a 'power' Fast Fourier Transform on the cropped image, but sub-sampled down to 256 x 256 pixels when the image resolution is larger than this. Once the FFT is performed, we can derive the total power at all spatial frequencies and frequencies above 1/16 of the highest frequencies present in the image. This is done by applying suitable 'masks' on the FFT image, as shown in Figure 7. Calculating the ratio from the resulting mask integrals normalises the result and makes it independent of the absolute image intensity. We have tried alternative measures such as gradients, contrast etc., but find that for most microscopy applications, intensity-based focus measures are not adequate.

The next issue is how to optimise the focus based on the focus measure. This is definitely not as trivial as it sounds. The aim is to find a 'peak' in the focus measure, but which way to go and how do we search for the peak? If we were allowed to take a large number of images and trace out the focus curve, everything would be easy. But by the time we found the peak, our sample is likely to have been thoroughly bleached! And the acquisition of each image might require a long integration time: hardly a fast approach. In practice, it is highly desirable to start the search in a defined direction and minimise the number of required images by using a kind of successive search algorithm based on the Fibonacci series, as explained below. The most serious problem is how to deal with situations where there is nothing to focus on, e.g. no fluorescence signal in the particular region imaged. If the search boundaries are not fixed, most approaches will eventually 'find' something to focus on: probably a speck of dirt on the distal surface of the slide! The only redeeming feature is that even the sophisticated systems used in consumer digital cameras also fail under similar conditions, even though the manufacturers never admit to this! At least microscopes allow manual focusing…..

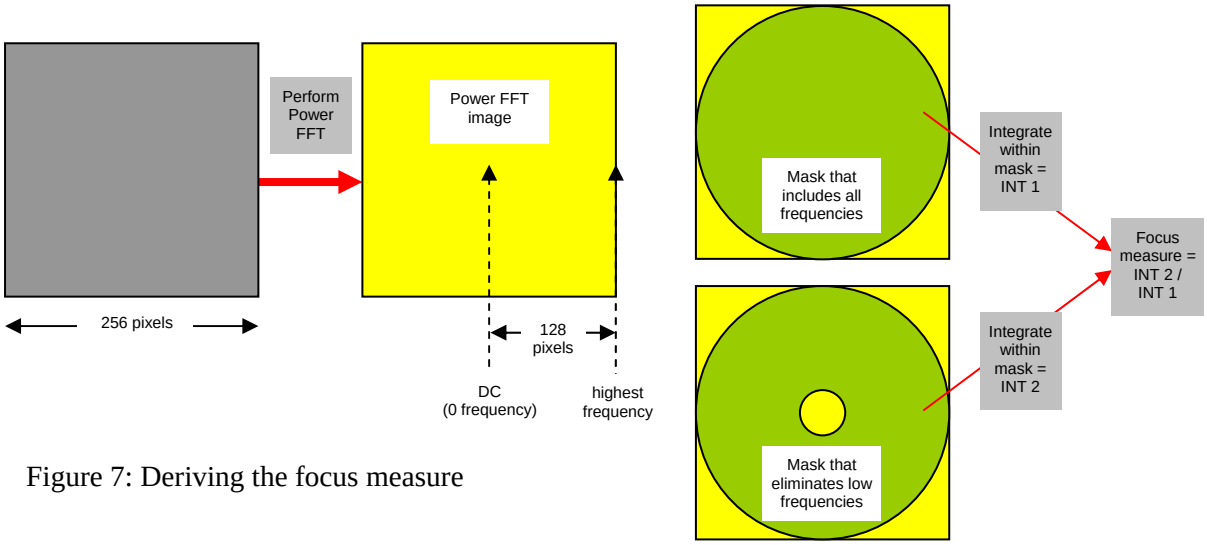

Figure 7: Deriving the focus measure

The object of the exercise is to maximise the 'FFT' focus value calculated above by varying z. Two methods work very well: The simpler 'Polynomial Fit' collects a number of values between two end points and performs a polynomial fit to the resultant z/FFT data. This is somewhat of a sledgehammer approach. The preferred method uses an algorithm to 'home in' on the optimum value of 'z' using a Fibonacci sequence to determine subsequent moves. Starting values that work in specific instances (i.e. with particular objectives) can be easily found empirically. These values are adjusted automatically on a change of objective. Obviously both algorithms fail if started when the image is completely out of focus, since successive focus indicator readings are too similar to perform sensible curve fitting or decide which direction to take next.

The first task for a search method is to find three points a_0, x and b_0 satisfying:

$$a_0 > x > b_0 \qquad \text{and} \qquad f(a_0) < f(x) > f(b_0)$$

This ensures that a maximum of a particular function is found between the two starting points. The principle of the search method is to minimize the uncertainty range in which the maximum is located. The starting uncertainty range is $[a_0, b_0]$. To reduce the uncertainty range we will now take two separate points a_1 and b_1 and compare these two with each other. When $f(a_1) > f(b_1)$ the next uncertainty range will then be $[a_0, b_1]$, otherwise it will be $[a_1, b_0]$. The two intermediate points are chosen in such a way that the reduction of range is symmetrical:

$$a_1 - a_0 = b_1 - b_0 = \varphi\,(b_0 - a_0)$$

After this reduction of uncertainty range the process can be repeated and we will find two similar points, a_2 and b_2. Generally when the process is repeated and in the former iteration $f(a_{n-1}) < f(b_{n-1})$ the new b_n will be set at the former a_{n-1}. If $f(a_{n-1}) > f(b_{n-1})$ it will be the other way around and a_n will be set at the former b_{n-1}.

The Fibonacci-search supposes that the φ-values are allowed to vary at every iteration. To make sure that only one new function evaluation is allowed at each iteration the values are restricted to $0<\varphi_k<1/2$. The strategy for selecting the evaluation points φ_k is as follows:

The purpose is then to minimize the reduction of the uncertainty range. This can be formally stated as:

$$\text{Minimize } (1-\varphi_1)(1-\varphi_2)\ldots\ldots(1-\varphi_N)$$

$$\text{for } k = 1,\ldots.., N.$$

The solution for this optimization-problem is defined on the basis of the Fibonacci Sequence F_1, F_2, F_3,…. The definition of this sequence is as follows for k>=0:

$$F_{k+1} = F_k + F_{k-1}$$

With the starting values to be $F_{-1}=0$ and $F_0=1$:

F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
1	**2**	**3**	**5**	**8**	**13**	**21**	**34**

The solution to the problem of optimization of the φ-value is then:

$$\varphi_1 = 1 - F_N / F_{N+1};$$
$$\varphi_2 = 1 - F_{N-1} / F_N;$$
$$\vdots$$
$$\varphi_k = 1 - F_{N-k+1} / F_{N-k+2};$$
$$\vdots$$
$$\varphi_N = 1 - F_1 / F_2.$$

In the Fibonacci Search Method the uncertainty range is reduced by a factor

$$(1-\varphi_1)(1-\varphi_2)\ldots..(1-\varphi_N) = 1/F_{N+1}$$

In the final iteration in the Fibonacci Search there is an anomaly because:

$$\varphi_N = 1 - F_1 / F_2 = \tfrac{1}{2}$$

This is not an allowed value of φ. To get around this, the reduction for the last iteration is set at:

$$\varphi_N = \tfrac{1}{2} - \varepsilon$$

where ε is a very small number of correction. The whole reduction factor of the Fibonacci Search Method can than be presented as:

$$(1+2\varepsilon) / F_{N+1}$$

This is presented in the table below

Evaluations (N)	Fibonacci ($1/F_N$)
5	0.125
10	0.0112
15	0.00101
20	0.0000914
25	0.00000824

Image compression and display

Now that we have collected a perfect set of images, it soon becomes obvious that we have a huge amount of data, which makes image display slow and/or cumbersome. Fortunately, there are highly effective ways of compressing the data, though lossy compression methods (e.g. methods based on the cosine transform such as JPEG) are definitely to be avoided. Lossless compression methods, such as those using algorithms based on fractals, are best, but are relatively slow and/or expensive to implement. A convenient alternative is to use wavelet compression techniques, which permit compression ratios of as much as 20:1 with negligible loss of information.

Whichever storage method is used, it is inadvisable to perform any of the 'stitching' calculations on the compressed images, unless it is certain that the compression is truly lossless. The practical consequence of this is to ensure that appropriate RAM area is available for the stitching process. Fortunately, the RAM addressing space in modern PCs tends to be budget-limited rather than by technical issues, so it is a good idea to keep track of memory prices and purchase when prices are low: you will always need more!

Disc storage is fortunately cheap and the disc capacities and access speeds keep increasing, so it is wise to always keep as much uncompressed information as possible. You are still faced with an almost bewildering range of formats, each with some form of limitation, either technical or associated with incompatibility with your collaborators. There is no simple solution to this. Bitmaps (*.bmp) or Tagged Image Format (*.tif) are undoubtedly convenient but often limited. In our work, we often deal with images consisting of many more than two dimensions. We need to ensure that images are readily shared and analysed with a variety of in-house developed software packages. We have thus implemented a highly flexible, generic multidimensional file format: Image Cytometry Standard (*.ics) that allows the convenient storage and access of data in excess of the usual three or four dimensions in a single file. Additional data regarding the scale, range and acquisition conditions can also be stored in the file, as well as a descriptive 'history' of the data. Routines can be implemented for the handling and display of multi-dimensional information utilising dynamic data arrays (DAA) that is both memory space and access efficient and flexible. With large multi-dimensional data sets, accessing appropriate 'planes' of data for display and processing in a timely manner becomes a challenge. We have therefore implemented data reformatting routines to enable easy data maintenance in an optimised memory structure. A software application to display any 2D plane from a multidimensional data set has been developed and we are planning to extend this to allow rendering of any 3 dimensions is planned. But dealing with multi-dimensional data sets is another story, best described in another document, so let's get back to the subject matter of this note.

Correlating image edges

It is instructive to remind ourselves about what is meant by correlation. The basic idea is to 'slide' adjacent image edges, in both directions, until the best 'match' of edge features is found. An interesting question arises as to the most optimum pixel resolution to use. If we use too many pixels, we are wasting time correlating, while too few pixels would not allow features to be matched properly. 'Too few' or 'too many' relate to the dimensions of the features we expect to see in our image and this in turn is dependent on focus quality, i.e. on the maximum spatial frequencies present in the image.

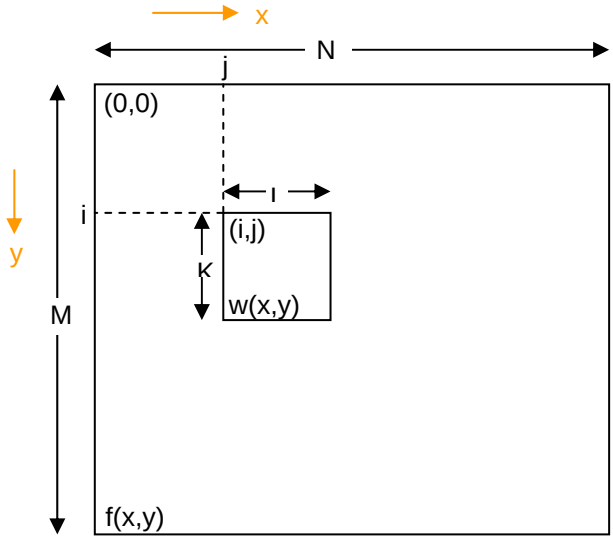

Figure 8: General normalised cross correlation procedure.

The normalised cross correlation coefficient for the case above is defined as:

$$\frac{\sum_{x=0}^{L-1}\sum_{y=0}^{K-1}(w(x,y)-\bar{w})(f(x+i,\,y+j)-\bar{f}(i,j))}{\sqrt{\left(\sum_{x=0}^{L-1}\sum_{y=0}^{K-1}(w(x,y)-\bar{w})^2\right)}\sqrt{\left(\sum_{x=0}^{L-1}\sum_{y=0}^{K-1}(f(x+i,\,y+j)-\bar{f}(i,j))^2\right)}}$$

In the case where we are to stitch images of equal dimensions it is usually sufficient to extract a narrow strip from the edge of one image and to correlate this with a larger rectangle from the other as shown in figure 9.

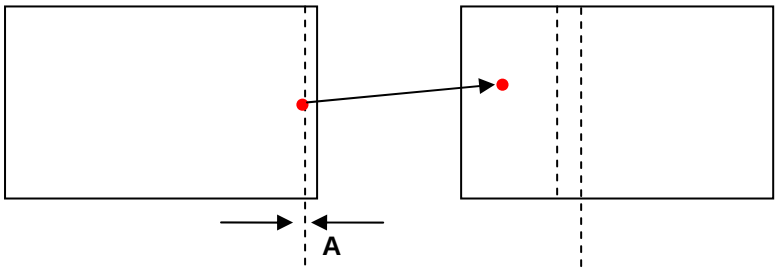

Figure 9: Correlation required to match the edges of two images. If there are features having reasonable contrast, the strip **A** can be as small as 3 pixels wide to obtain an unambiguous point of best correlation.

This obviously gives rise to a lengthy and CPU-intensive series of calculations. However, there are ways in which the process can be optimised. Closer inspection reveals that it is not necessary to recalculate the term, $\sqrt{\left(\Sigma\,\Sigma\,(w(x,y)-\bar{w})^2\right)}$, for every point. Furthermore, if features are large enough, the images may be sub-sampled to obtain a rough idea of the location of the point of best correlation and then a fine search used over a small region about this point.

In the case where the images have been acquired using a motorised stage the overlap between images is known, so it is only necessary to search a small region as shown in Figure 10.

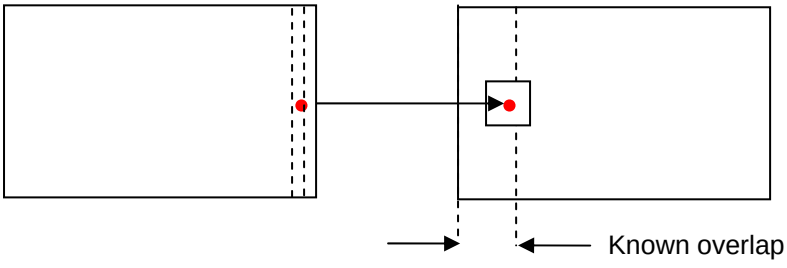

Figure 10: Searching a small rectangular region centred about the theoretical point of best correlation.

While it is possible to correlate edges of images in the same order as they were acquired, there are 'better' ways to proceed. Acquisition is most readily (and efficiently) performed in a snake-like pattern, as shown in Figure 11. When correlating however, it is better, though less convenient, to perform this in a spiral-like fashion, from the centre outwards, as shown in Figure 12. Histology sections tend to be round; hence the only information available for correlation at the corners of the area scanned is due to dirt on the slide. By starting in the middle, greater weight is given to the area containing information of interest. This method also lends itself better to optimised correlation of sets of images that have been acquired manually, i.e. without the use of a motorised stage, where the overlaps are unknown and uneven. In this case images at the corners of the rectangular region may not even have been acquired.

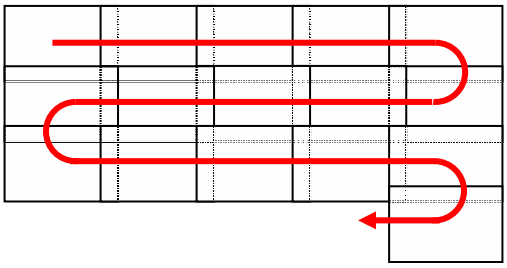

Figure 11: Example of sequence of stage moves to 'scan' a large area, with minimum times from frame to frame

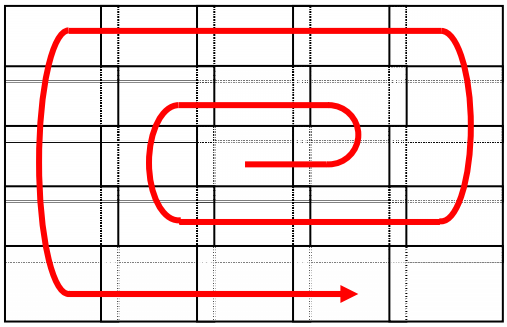

Figure 12: Example of optimal sequence to correlate adjacent images such that two edges at right angles are most often 'available'

It may often be the case, especially when working with weakly fluorescing samples, that a good point of correlation cannot be found. In this case a 'hole' is left in the mosaic image. When all possible images have been placed, it is likely that there will be more information available for correlation, e.g. the hole may now have an image on all four sides which can be used to find a best match. A second pass of the spiral will generally fill all such holes successfully. At the corners of the region there may be no information at all if the slide was very clean and we are dealing with a circular or convex sample. In this case there is little choice but to assume the position of the images based on the known stage position at the time of acquisition. In any case, such featureless regions will probably be of little interest or importance.

Feathering or blending image edges

There are instances where image edges show a degree of intensity variations, which are unacceptable and show up as distinct steps in intensity, even when correlation is near-perfect. A useful approach to eliminate such visually undesirable effects is to 'feather' or blend the image edges. However it should be noted that any quantitative analysis, which uses intensity as a measure, is then rarely, if ever, valid. Instead of using information from only one of the overlap regions, and 'discarding' the other overlap region, use is made of both sets of information. The application of blending methods thus improves the overall perceived image quality by making boundaries invisible. Each image set consists of a number of overlapped images that are initially placed onto a grid, using visual inspection (in the case of manually acquired images) or their relative positions are known (in the case of automatically acquired images, i.e. with a stage). Images are then sorted so that they can be brought together in the spiral-like pattern as described previously. We create an empty 'canvas' of appropriate size and place the first, central image in the middle of this. Each image is taken in order and placed on the canvas; its position is determined by edge correlations to the neighbouring images; this ensures 'best' placement. Edge blending is then performed and the process repeated for all other images. This is shown in simplified form in Figure 13.

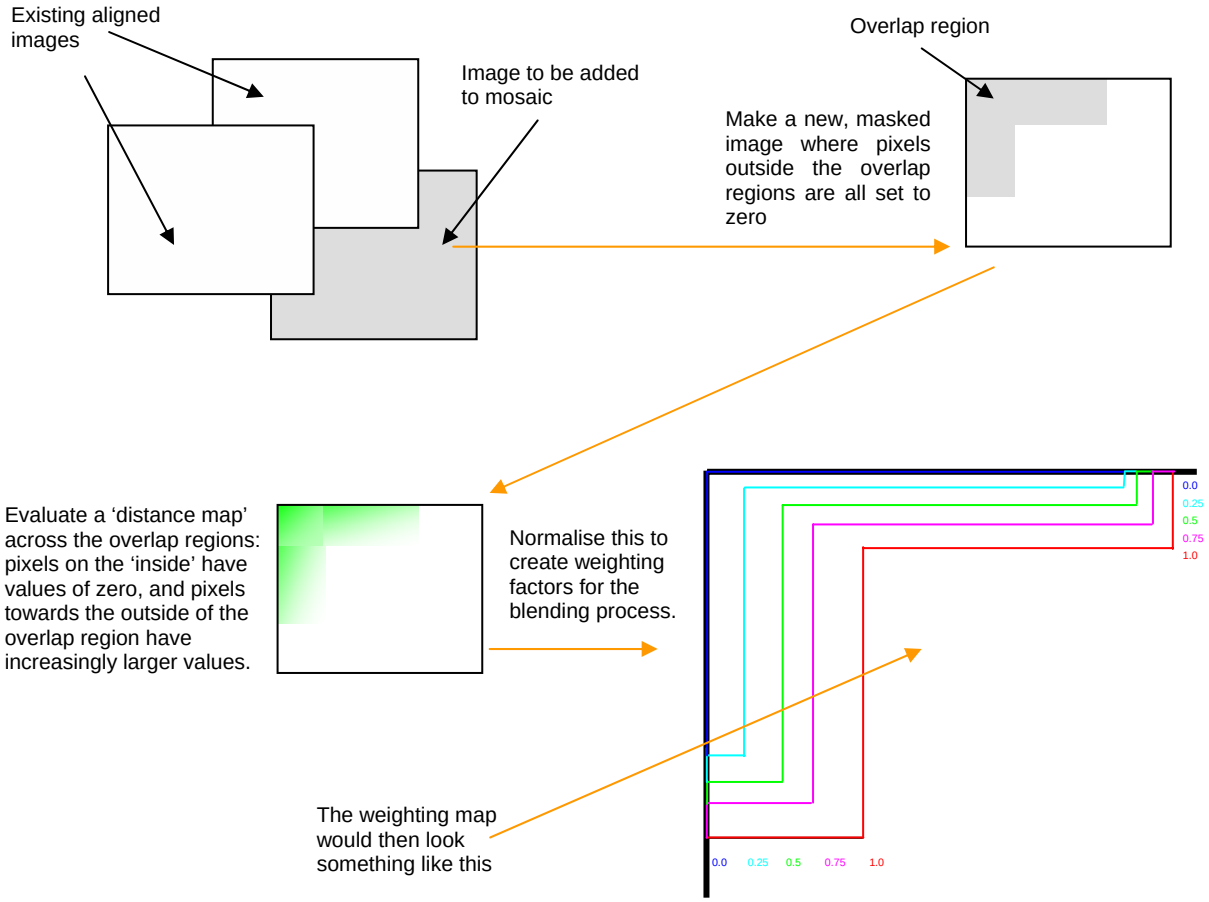

Figure 13: Main steps involved in blending intensities of adjacent images.

The image blending algorithm is created using a 'distance map' approach. The basic idea is to first generate a distance map of every pixel in the overlap region. This is then modified and normalised to generate a 'weighting map' that defines what proportion of the intensities of the two overlap regions is used in the final image. This weighting factor (α) tells us this proportion. The pixels in the original images are correspondingly weighted by a factor $\beta = 1 - \alpha$.

The blended image thus consists of pixels $N_{(x,y)} = \alpha \, . \, I_{(x,y)} + \beta \, . \, M_{(x, y)}$

where $M_{(x, y)}$ is previous image (mosaic) pixel, $I_{(x,y)}$ is image pixel from the new, added image and N(x,y) is new image (mosaic) pixel.

This approach thus minimises the perceived effects of inherent intensity variations and apparent intensity variations resulting from any remaining geometric misalignments between the constituent images.

Examples of stitched and blended images can be found in the document entitled 'Examples of stitching.doc'

Useful addresses of suppliers

Edmund Scientific
Tudor House, Lysander Close
York, YO30 4XB
Tel: 01904-691-469
Fax: 01904-691-569
http://www.edmundoptics.com

Pyser-SGI Ltd
Fircroft Way, Edenbridge
Kent, TN8 6HA
Tel: 01732 864-111
Fax: 01732-865-544
http://www.pyser-sgi.com

Technical Video Ltd.
P.O. Box 693
Woods Hole, MA. 02543, USA
http://www.technicalvideo.com