Virtual Microscopy with Extended Depth of Field

Andrew P. Bradley
School of Information Technology
and Electrical Engineering
The University of Queensland
a.bradley@itee.uq.edu.au

Michael Wildermoth
School of Information Technology
and Electrical Engineering
The University of Queensland
wildermoth@itee.uq.edu.au

Paul Mills
School of Veterinary Sciences
The University of Queensland
Australia
p.mills@uq.edu.au

Abstract

In this paper, we describe a virtual microscope system, based on JPEG 2000, which utilizes extended depth of field (EDF) imaging. Through a series of observer trials we show that EDF imaging improves both the local image quality of individual fields of view (FOV) and the accuracy with which the FOVs can be mosaiced (stitched) together. In addition, we estimate the required bit rate to adequately render a set of histology and cytology specimens at a quality suitable for on-line learning and collaboration. We show that, using JPEG 2000, we can efficiently represent high-quality, high-resolution colour images of microscopic specimens with less than 1 bit per pixel.

Keywords

Virtual Microscopy; Extended Depth of Field; JPEG 2000.

1   INTRODUCTION

Virtual Microscopy (VM) can be defined as an artificial microscope environment, created on a personal computer, that when presented to a user has the *look-and-feel* of a real microscope. Virtual Microscopy has number of potential advantages over traditional microscopy:

- Digital specimens do not degrade over time [5];
- Browsing, navigation and annotation of specimens is greatly enhanced [4];
- Exactly the same specimens can be distributed to multiple sites instantaneously [5];
- There is no risk of breakage, loss or contamination during transportation; and
- Significantly reduced set-up and maintenance costs for image browsing. However, specimen scanning is still an expensive [7] and time consuming process [8], [9].

For these reasons VM is starting to be used in a number of specific microscopy applications. For example, VM is currently being trialled by the Royal College of Pathologists Australasia for their quality control of pathology laboratories[1] and a number of universities are evaluating the enhanced learning potential that VM can provide (for example, see[2]).

A VM system consists of two main parts:

1. **Scanning system**, which digitises the microscopic specimens, and
2. **Client browser**, which is used to browse and view the digital images of the microscopic specimens.

In this paper, we will outline the VM system we have developed and describe one specific improvement: the addition of extended depth of field (EDF) imaging. We show that EDF imaging improves the quality of the scanned images both in terms of local focus and contrast and improved image mosaicing (stitching). In addition, we investigate the bit rate requirements to adequately render digital images of microscopic specimens suitable for applications such as on-line learning and collaboration.

1.1   Focus in Virtual Microscopy

Currently, high-resolution VM systems are limited in use to relatively thin histological specimens, say less than 5μm. This is because at high magnification (typically x40) the depth of field (focus) of the microscope is only about 2μm and so it is not possible to adequately focus on thicker specimens [9]. This problem can be reduced by using a line scan, rather than an area scan camera, as the automated focus algorithm is able to track changes in focus more precisely [8]. However, the goal of VM should be to mimic the focus abilities of a real microscope [4]. That is, to provide *digital focus* so that the user has the ability to focus on and through the sample. Clearly, this type of multiple focal plane imaging requires a significant increase in both the complexity of the scanning process and the amount of data to be stored. For example, if focal planes are coded independently then a bit-stream containing ten focal planes will be approximately ten times the size of a bit-stream of a single focal plane. To date, this type of functionality is only available for a limited number of pre-specified areas, e.g., the 3D re-visit® system from Aperio technologies[3]. Clearly then, there is a need for a hybrid VM system that digitally extends the depth of field of the optics on the scanning microscope to match the depth of the specimen (say, 20μm). In this way, 2D images are produced with all of the objects, in every field of view, in focus. While the 3D depth information is effectively lost in these images, they have superior image quality, are compatible with current VM browsers and do not produce significantly larger bit-streams.

---

[1] http://www.rcpaqapa.netcore.com.au/vm.html

[2] http://www.path.uiowa.edu/virtualslidebox/
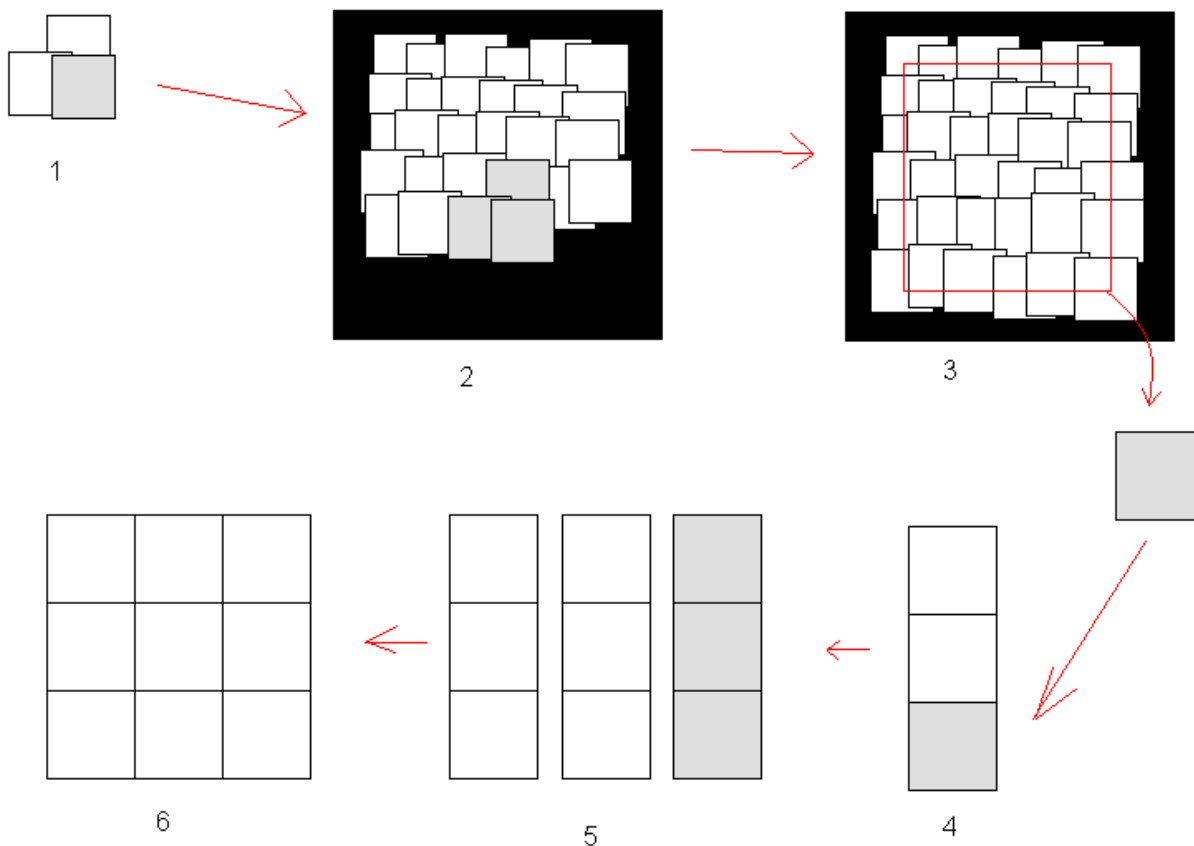
[3] http://www.scanscope.com

Figure 1. Overview of the image stitching process: (1) each FOV is stitched in raster scan order to the FOV above and behind; (2) this locates the FOV on to a 6x6 'canvas' buffer. (3) When complete, this buffer is cropped and pushed to the JPEG 2000 compression engine (4). When an entire column of buffers is complete, it is saved as a JPEG 2000 tile fragment (5). When all tile fragments are complete they are appended to form the complete JPX bit-stream (6).

2 SYSTEM DESCRIPTION

The scanning system consists of an AcCell 2000 (AccuMed International, Chicago) automated microscope (which is based on an Olympus BX51 microscope), a high-resolution 3-shot digital colour camera (Diagnostic Instruments, Sterling Heights, MI) and a light balanced daylight filter (80A) is used for all colour imaging. System software is written in JAVA (Sun Microsystems, Santa Clara, CA) with the computationally intensive image processing operations, such a normalised cross-correlation (NCC) and the discrete wavelet transform (DWT) being written in C and Intel® integrated performance primitives (IPP)[4]. All images are stored using JPEG 2000 (Kakadu Software, NSW, Australia)[5] and so are compliant with the international standard ISO/IEC 15444 and can be browsed interactively via the JPEG Internet protocol (JPIP). JPEG 2000 was chosen because it can efficiently handle large (i.e., Gbyte) images and provides, in a single bit-stream, a wide array of functionalities desirable for VM:

- Progressive transmission by resolution, quality, component, or location;
- Random access; and
- Integrated lossless to lossy compression [2].

For a typical specimen, an area of approximately 15×10mm is scanned. With an x40 objective and a camera with 7µm pixels, 100% fill factor, arranged in a 2048×2048 sensor array this relates to an effective camera resolution of 0.18µm per pixel. This then results in a scanned image of approximately 4.6 giga-pixels (83,400×55,600) of the sample. Therefore, typical colour image file sizes will be of the order of: 14GBytes of raw, uncompressed data; approximately 2 GB with truly lossless compression; and ≤ 1GB with visually lossless compression.

Clearly, with such large images it is not feasible to load the whole image in to memory at any one time. Therefore, it is essential to only decode and view the requested area of the sample at the requested resolution. JPEG 2000 provides a highly structured bit-stream, where information is put into packets relating to a specific resolution, spatial area, colour

---

[4] http://www.intel.com/cd/software/products/asmona/eng/perflib/ipp

[5] http://www.kakadusoftware.com

component and quality layer [2]. Therefore, it is ideally suited to this type of application.

The overall scanning process is depicted in Figure 1 and consists of the following three general steps:

1. **Scanning**: images of each FOV are captured and (losslessly) stored to disk;

2. **Stitching**: individual images are aligned in relation to the FOV above and behind and then pasted into a *canvas* buffer. Periodically, this buffer is pushed onto the JPEG 2000 compression engine until an entire (~6 mega-pixel wide) column of the image is complete. It is then saved as a JPEG 2000 tile fragment [2];

3. **Final Bit-stream Creation**: when all of the columns of image have been stitched and compressed the final bit-stream is created by appending these tile fragments into a single JPX bit-stream.

In the following sections each of these steps will be described in more detail.

2.1 Image Scanning

Before FOV scanning can start we first must set the white balance of the camera and acquire a background correction image. Both of these are done by exposing the camera to a blank area close to the specimen on each slide, so that we compensate for any mounting medium. To set the white balance, the exposure settings on each colour channel (RGB) are set, using a golden section search, so that they have equal intensity. The background correction image, which is subtracted from all subsequently acquired images to compensate for uneven illumination across the FOV, is the average of five acquisitions of one background FOV.

To minimise the effect of slack in the stepper motors, samples are scanned from top to bottom in an alternating left to right, right to left raster scan pattern. Each FOV is overlapped with the previous FOV to enable the final image of the specimen to be stitched together. Currently, an overlap of 45% is used to maximise the robustness of the stitching process.

Each FOV is focused by maximising the standard deviation (SD) of the image intensity over the whole FOV. The focusing algorithm performs a binary search, repeatedly halving the focus step size (from 8 to 0.5µm) to find the maxima. However, to improve robustness, when the smallest focal step size has been reached, the algorithm maximises the average of the previous three SDs. In addition, as part of the focusing algorithm a series of images are acquired from both above and below the optimal focal point. These images can then be combined into a FOV with digitally extended depth of field (EDF) [1]. In this way, it is possible to save two versions of each FOV captured:

1. One at the best single focal point (BSFP); and

2. One with extended depth of field (EDF).

The EDF algorithm used here is described in more detail in Section 2.1.1. After each FOV has been focused it is indexed and temporarily stored in Portable Network Graphics[6] (PNG) format.

2.1.1 The EDF Algorithm

The EDF algorithm used here is essentially that described in [1]. In summary, as each image is acquired (as part of the focus measurement above) it is pushed to the EDF algorithm. The EDF algorithm initially performs a seven level DWT on each these images from the focus *stack*, so that it can preserve the maximum amplitude wavelet coefficients, at each resolution, at each spatial location. When focusing has completed and the EDF algorithm has been passed all of the images from the focal stack, then the preserved *maximum amplitude* wavelet coefficients are inverse wavelet transformed (IDWT). The resultant EDF image is then clamped to maintain the dynamic range of the original image stack.

For this specific application, a number of modifications were made to the algorithm described in [1]:

- **Extension to colour**: Images were first converted from the RGB to YCbCr colour-space. Wavelet coefficient selection was then performed on the luminance (Y) component only and the chrominance channels (CbCr) were copied across as appropriate. In addition, clamping of the EDF image was based on the maximum and minimum of the Y channel over the focal stack;

- **Computational complexity**: As a DWT is performed on each colour component of each focal plane, a (conventional) critically sampled DWT was used; and

- **De-noising**: Selecting the maximum amplitude DWT coefficients at each location at each resolution not emphasises edge contrast, but also noise in the background. Therefore, a small degree of wavelet shrinkage [3] was performed to suppress small wavelet coefficients related to camera noise.

2.2 Image Stitching

The stitching process uses normalised cross correlation (NCC) [6] to determine the location of the image being stitched, relative to images that have already been placed on the canvas. The stitching proceeds in raster scan order: left to right, top to bottom. So in general, we have an existing image that has been placed on the canvas both above and to the left of the image being stitched. Because we know that each FOV has a pre-specified overlap with all eight neighbouring FOVs (currently 45%) we can extract rectangular areas from the current canvas buffer that we now should appear in the current FOV. In this way, we calculate the NCC of two rectangular sub-regions related to the top and left hand side of the current image. We then find the location of the maxima of each NCC matrix and use this to obtain the location of the current image relative to the two images already placed on the canvas. As we store the location on the canvas of each FOV as we perform

---

[6] http://www.libpng.org/pub/png/

the stitching, we can then use the relative locations from the NCC matrices to calculate two absolute canvas locations. We can then combine the two results in a proportional average, based on the actual maximal value of each cross correlation coefficient. That is:

$$\gamma'_t = \gamma_t / (\gamma_t + \gamma_l)$$
$$\gamma'_l = \gamma_l / (\gamma_t + \gamma_l)$$
$$x' = \gamma'_t x_t + \gamma'_l x_l$$
$$y' = \gamma'_t y_t + \gamma'_l y_l$$

Where γ_t and γ_l are the cross-correlation coefficients from the top and the left hand images respectively, (x_t, y_t) and (x_l, y_l) are the relative offsets of the respective NCC maxima. Therefore, if the NCC coefficient from above is higher, we have increased confidence that it provides a good stitch, and therefore should have a greater influence on the final coordinates. We use this process on each FOV in the image, although obviously along the edges of the scan area we have only one FOV to stitch against.

Additionally, two processes are used to minimize the effect of stitch-line artefacts that result from imperfect stitching, some of which are inevitable. Firstly, the edges of the image being pasted down are pixel blended. We currently blend five pixels in from the edge, with the blending proportions evenly distributed as follows:

[1/4:3/4; 2/3:1/3; 1/2:1/2; 1/3:2/3; 3/4:1/4].

The second process is designed to avoid the visually annoying artefacts that occur when two parallel image boundaries are close together (which results in one image above, one below, and a small sliver of the third between them). This occurs because of the overlap between each FOV. To avoid this problem we simply paste only the parts of the tile that aren't overlapping with existing tiles on the canvas (other than an overlap of five pixels required to do the blending). For clarity, the blending process is not depicted in Figure 1.

One last fail-safe is worth mentioning. If a FOV is blank or largely empty of material then the NCC will be either very low (currently, < 0.5) or the stitch results will be far from the expected location (currenty $> \pm100$ pixels (18μm) in either x or y). In these cases we simply ignore the cross correlation results and paste the image down the default position on the canvas. The default position is an offset relative to the image above and was estimated for the current FOV overlap using the average offset calculated on a stitched graticule slide (ignoring FOVs on the top, left and right sides of the scan area which are affected by slack in the stepper motors).

As illustrated in Figure 1, when an area consisting of 6x6 FOVs has been stitched and placed on the canvas, the buffer is cropped and sent to the JPEG 2000 compression engine. Cropping the buffer in this way allows the right hand and bottom edges to for form the initial left hand and top edges respectively of subsequent canvas buffers. Hence, there are no visible boundaries between different canvas buffers, either as you progress down the column of a tile fragment or between tile fragments. Although, in theory JPEG 2000 is capable of coding even giga-pixel images as a single tile [2], in practice we have found it necessary to tile the scan area into vertical stripes in order to control memory usage (particularly the uncompressed canvas buffer on which the stitching is performed). Currently, stripes are approximately 6 mega-pixels wide, being 6 FOVs each with a 45% overlap, and the full height of the sample (which may be many 10s of mega-pixels). This allows an eight level wavelet transform to be performed before the number of wavelet coefficients at the highest level gets smaller than the code block size (which is fixed at 32x32 for all levels).

2.3 Final Bit-stream Creation

When all the tile fragments (stripes) have been stitched and saved as individual JP2 bit-streams, they are finally combined into a single JPX bit-stream that represents the whole specimen. In principle, JPEG 2000 allows for this to be done as a *transcoding* process that does not require each stripe to be decompressed and then recompressed into the final bit-stream. However, a limitation in the current Kakadu implementation of JPEG 2000 means that we can not currently do it this way. Therefore, each stripe is decompressed and the recompressed as a tile fragment.

Note to enable efficient parsing of large JPEG 2000 images we use the following settings:

- 10 quality layers are assigned in approximately logarithmically spaced bit rates from 2 bits per pixel (bpp), as described in Section 3.1;

- The code-block size is set to 32x32 and the precinct sizes are 256x256 at the two highest resolutions and 128x128 elsewhere;

- The bit-stream progression order is specified to be resolution; position; component; and then layer (RCPL). This optimises access to resolution first, position second etc;

- Packet length information is inserted into the bit-stream for faster bit-stream parsing. Tile-parts are also used which group code-blocks from each resolution together.
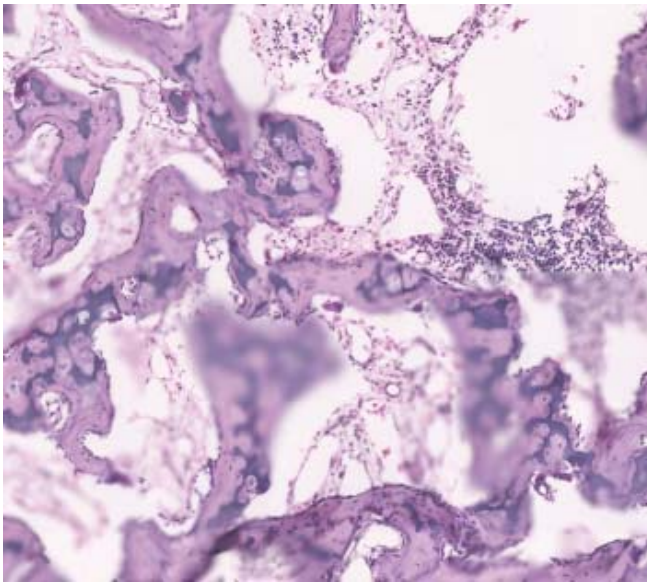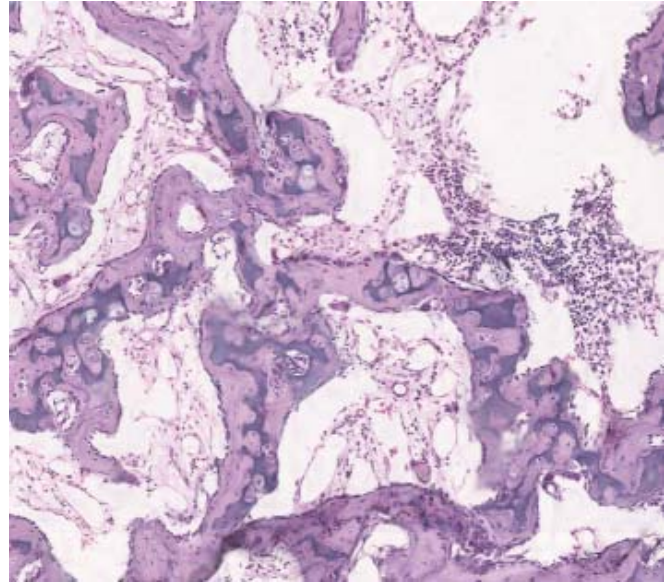
(a)   (b)

Figure 2. A 6x6 FOV of slide H89 (a) BSFP and (b) EDF ±18µm.

2.4   VM for Online Learning

Currently, the primary application of this project is enhanced online learning of anatomy and neuroanatomy courses for veterinary science students. Therefore, we have set up a JPIP server that hosts a set of JPEG 2000 images of histological samples used as part of a course. These images have been scanned with the system described in this paper and have had manually entered metadata descriptions and annotations added to each virtual slide [2]. We have also developed VM image browser that is effectively a JPIP client and allows students to quickly and easily browse and interact with these images as they would on a conventional microscope.

The VM browser displays two views of the sample, firstly a *macroscopic* view of the whole sample, as is shown in Figure 3; and a *microscopic* view where the user examine specific areas of the images in detail. For simplicity, the magnifications available for the microscopic view are those directly available in the JPEG 2000 bit-stream, i.e., dyadic sub-multiples of the native x40 magnification used for scanning, e.g., x20, x10, x5 etc. In addition, there is also a capability to digitally zoom to x80 magnification to allow for closer examination of very small structures. The user is able to navigate around the sample using either the arrow keys on the keyboard, the mouse pointer or by clicking on a location in the macroscopic view. At all times the extent of the current microscopic view is displayed on the macroscopic image, which provides valuable context information.

3   EXPERIMENTAL EVALUATION

In this section we outline some formal and informal evaluations that we have carried out on our VM system.

3.1   Rate Distortion Trade-offs

Although the images are hosted and viewed on a single university intranet, it is still desirable to make the files as small as possible so that they can be downloaded and displayed as quickly as possible. However, clearly the smaller we make the files the more we have to quantise the coefficients and hence reduce the reconstructed image quality. Therefore, we designed a simple experiment to estimate the minimum bit-rate at which "acceptable" quality is observed. Note: our definition of acceptable here is an image quality that shows the structures of interest in the necessary detail for learning.

For this trial there were four virtual slides of complete specimens available (three histology and one thin-layer cytology, as shown in Table 1). Four observers, who are the coordinators of the courses in which VM is planned to be used, were then instructed to perform the following:

1. Load the virtual slide so that the low resolution image is displayed;

2. Navigate so that a pre-specified feature of interest in the centre of the field of field;

3. Select x40 magnification to zoom in on this feature (Note: some minor navigation may be required to centre or see the whole feature of interest);

4. Wait until the full quality image has downloaded;

5. Gradually reduce the number or rendered quality layers until the image has an UNACCEPTABLE quality, i.e., you notice blurring or jagged edges. Note: this functionality is not available to the students;

6. Gradually increase the number of rendered layers until the image quality is again ACCEPTABLE;

7. Record of the slide number, region of interest and minimum number of layers showing ACCEPTABLE image quality;

8. Move to the next region of interest or slide and repeat the steps 1 through to 7.

As 10 layer bit-streams were generated in an approximately logarithmic spacing from 2bpp, each layer relates to approximately the following bit-rates:

[2; 1; 0.5; 0.25; 0.125; 0.063; 0.031; 0.016; 0.008; 0.004].

Table 1. Number of layers showing acceptable image quality for all slides in the study

Slide Name	Acceptable Quality Layer			Blank Area
	Min	Median	Max	
H28: Artery1	5	6	7	45%
H85: Artery2	5	6	6	59%
H89: Ear	5	6	7	40%
C101: Cytology	4	5	6	38%
All slides	4	6	7	45%

In Table 1 the amount of blank area on each slide was estimated by first manually selecting an intensity threshold between 238 and 250, performing a morphological closing and then counting the proportion of white pixels in the resulting binary image.
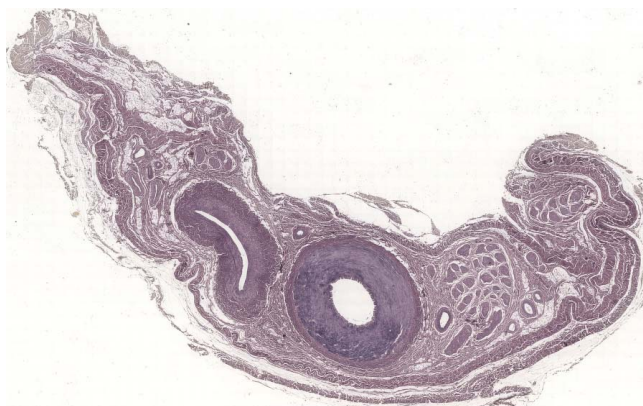

Figure 3. H28: 40,097 x 57,280 histology specimen.

Table 1 shows that there is a broad agreement between observers on the number of layers (and hence the bit rate) required for adequate quality. Even the maximum bit-rates required for adequate quality for this application are extremely low (0.25bpp). As a result a typical 4.5 giga-pixel image requires significantly less than 1 Gbyte of storage. These low bit rates can be explained by considering a number of factors specific to our application:

1. Even the tiled images used here are very large in size and so an DWT with many levels can be performed and the arithmetic coder has lots of data on which to estimate state probabilities;

2. As Table 1 and Figure 3 show there are large areas of the scanned images that are effectively blank and these

areas are very easy to compress. Essentially, most histology samples are likely to be highly irregular in shape and so scanning a rectangular area will in general result in a large number of the FOVs being blank.

3.2   Effect of EDF on Local Image Quality

This visual trial was based upon twenty fields of view (FOV) from a selection of the slides used in the previous study. These FOVs were chosen to have a reasonably varied content, consisting of both isolated and clumped material. The purpose of the visual trial was to directly compare images captured at the best single focal plane (BSFP) and those with extended depth of field (EDF). A two alternative forced choice (2AFC) methodology was selected because of its simplicity, i.e., the observer views the two images (BSFP and EDF) and then selects the one preferred.

Table 2. Observer FOV preferences between best single focal plane (BSFP) and extended depth of field (EDF). FOV Number is also shown when BSFP is selected.

Observer	BSFP	EDF
APB	0	20
PFW	1 (15)	19
MJN	0	20
PCM	1 (11)	19
SL	3 (11,12,16)	17
LA	0	20
All Observers	5	115

There were six observers (4 male and 2 female) all with good, or corrected, vision and all were experienced users of conventional light microscopes. The viewing distance was approximately 40cm (i.e., a normal PC viewing distance) and the image pairs were viewed one at a time in random order. The observers were free to view the images multiple times before making a decision and a blank mid-grey image was shown between each image (for 2 seconds) to prevent observers switching between the two images to find insignificant differences. Each image pair was viewed once which means that each observer took approximately 10 minutes to view all 20 FOVs. Images were viewed on a 15.1" Thin Film Transistor (TFT) display, in a darkened room (i.e., daylight with drawn curtains). The test images were displayed on a mid-grey background at a size of 760x1024 pixels. Prior to the start of the visual trial all observers were given a short period of training on the usage of the software and were told to select they image they preferred in terms of contrast and morphology of the specimen.

Table 2 shows that for all six observers there was a clear preference (96% or 115/120 FOVs) for the EDF images over BSFP. Of the 4% (5/20) of preferences given to BSFP these are spread over 4 images. Three of these images (11, 15 and 16) show fairly isolated cells where either the difference between EDF and BSPF is very small (11 and 15) or the out of focus objects in BSPF are towards the edges of the image (16). It is not clear why FOV 12 was selected by observer SL as this is an image of a clump of

observer SL as this is an image of a clump of cells for which BSFP leaves a number of nuclei clearly out of focus.

3.3 Effect of EDF on Image Stitching

In this section we will present an informal evaluation of the improvement in image stitching due to EDF imaging. We shall do this by showing in Figure 4 a selection of image stitches that lie on the boundary of objects of interest, in this case cervical cells. The examples shown clearly show that as EDF improves the edge contrast of objects and increases the amoint of high frequency information in the images, it also produces more pronounced maxima in the NCC matrix and this produces more accurate image stitching. It is worth noting that to date we have been unable to find a single example where the stitching using EDF is visually inferior to the stitching with BSFP.

4 CONCLUSIONS

We have developed a VM system that captures images from multiple focal planes and then combines them into a single *all in focus* image. These extended depth of field images have superior image quality when compared to conventionally captured images and also improve the robustness of the image stitching (mosaicing). In addition, we have shown that images can be compressed to less than 1 bpp and still have adequate quality for online learning and collaboration.

Future work will consist of improving background correction and mosaicing to make it globally optimal and also perhaps take consideration of heuristic constraints, such as the (approximately) known shift between FOVs. We will continue to investigate the feasibility of digital focus, however, this requires the implementation of the multi-component transform (MCT) described in part 2 of the JPEG 2000 standard. In addition, to improve image browsing efficiency across the internet requires the implementation of a JPIP proxy server.

5 REFERENCES

[1] Bradley, A.P., and Bamford, P.C., "A One-pass Extended Depth of Field Algorithm Based on the Over-complete Discrete Wavelet Transform," Image and Vision Computing '04 New Zealand (IVCNZ'04), pp 279-284, Akaroa, NZ, 2004.

[2] Christopoulos, C., Skodras, A, and Ebrahimi, T., "The JPEG 2000 Still Image Coding System: An Overview," IEEE Transactions on Consumer Electronics, 46(4): 1103-1127, 2000.

[3] Donoho, D.L., "De-noising by Soft-thresholding," IEEE Transactions on Information Theory, 41(3): 613-627, 1995.

[4] Kumar, R.K., Velan, G.M., Korell, S.O., Kandara, M., dee, F.R., Wakefield, D., "Virtual Microscopy for Learning and Assessment in Pathology," *Journal of Pathology*, 204 (5): 613-618, 2004.

[5] Leong, F.J. and McGee, O.D., "Automated Complete Slide Digitization: a medium for simultaneous viewing by multiple pathologists" Journal of Pathology, 195 (4): 508-514, 2001.

[6] Lewis, J.P., "Fast Template Matching," Vision Interface 1995, Quebec, Canada, pp 120-123, 1995.

[7] Romer, D.J., Yearsley, K.H. and Ayers, L.W., "Using a Modified Standard Microscope to Generate Virtual Slides, The Anatomical Record, 272B: 91-97, 2003.

[8] Soenksen, D.G., "Fully Automatic Rapid Microscope Slide Scanning," US Patent 6,711,283, 2004.

[9] Steinberg, D.M. and Ali, S.Z., "Application of Virtual Microscopy in Clinical Cytopathology," Diagnostic Cytopathology, Vol 25(6): 389-396, 2001.
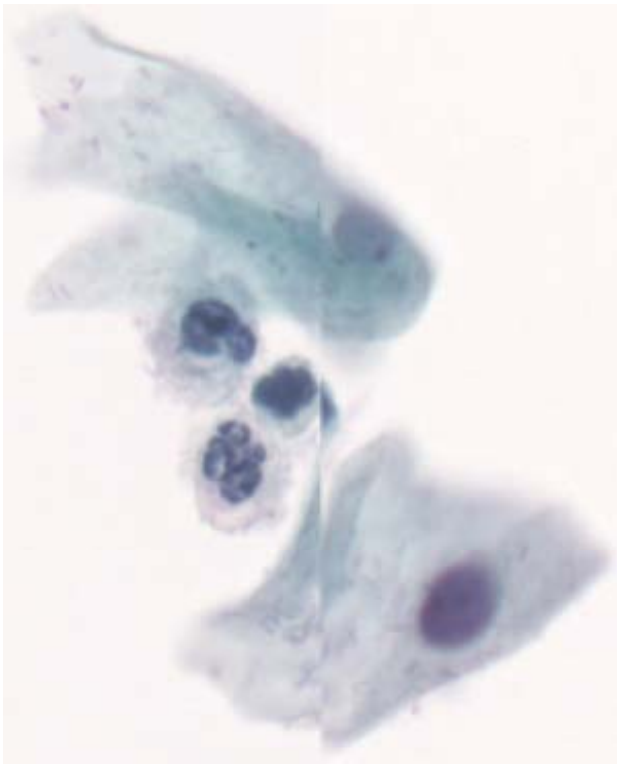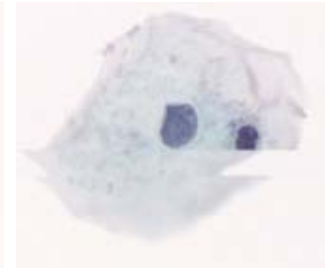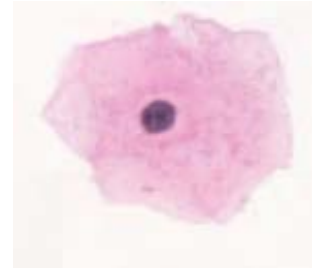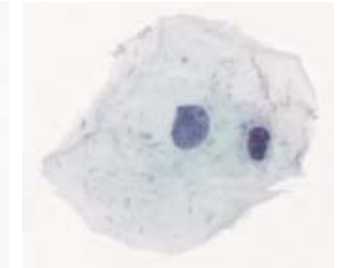
(a.1)
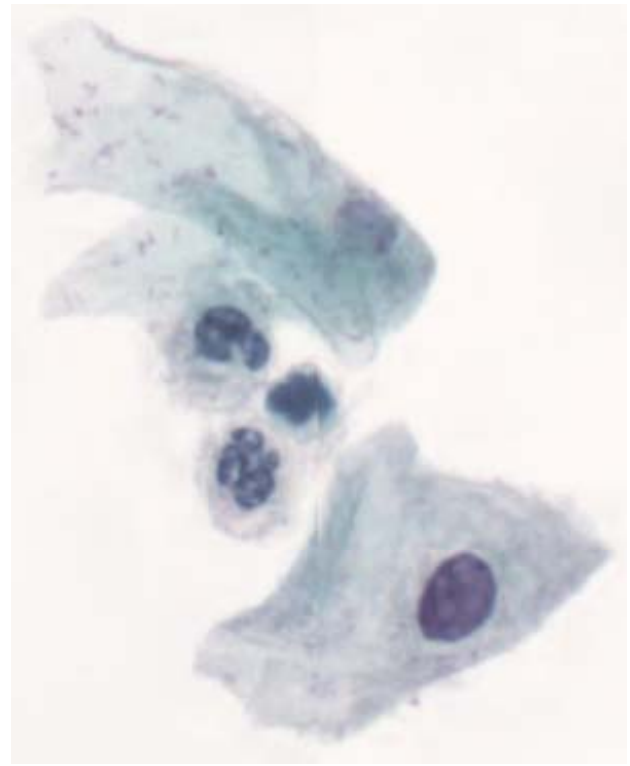
(b.1)

(a.2)

(a.3)

(b.2)

(b.3)

Figure 4. Three examples of stitch lines through cytology cells (a.1-3) BSFP (b.1-3) EDF.