Towards Optimal Image Stitching for Virtual Microscopy

Ben Appleton, Andrew P. Bradley, Michael Wildermoth
Electromagnetics and Imaging Research Division
School of Information Technology and Electrical Engineering, The University of Queensland
Queensland 4072 Australia
{appleton, bradley, wildermoth}@itee.uq.edu.au

Abstract

In this paper we present an image stitching method based on dynamic programming and describe its application to automated slide acquisition for Virtual Microscopy (VM). Given a large number of fields of view (FOVs) acquired from a single microscope slide, we composite these images into a single large 'virtual slide' image. The location of each FOV is determined using a new algorithm based on dynamic programming. We compare the performance of the proposed algorithm to an existing greedy algorithm. In a visual trial it is shown that the new algorithm provides a significant improvement in perceived image quality at image boundaries compared to the existing algorithm.

1 Introduction

The aim of Virtual Microscopy (VM) is to replace traditional light microscopes by personal computers in some scenarios. In VM the microscope slide is automatically scanned once and then stored in digital form on a central server. Following this, the slide may be browsed on any personal computer as though it were physically present. Virtual Microscopy has several benefits over traditional light microscopy: a virtual slide can be browsed remotely; specimens do not degrade over time; and slides cannot be broken or lost (they can even be backed up!). In addition to these direct benefits, the widespread acceptance of VM will lead to large digital slide databases for the development of image analysis techniques and also open up a range of applications such as online learing and quality assurance.

One of the central problems of automated VM is the image acquisition stage. Slides are simply too large to be acquired as a single image, so instead it is necessary to capture many fields of view (FOVs) before combining these into a single slide image. Automated microscope stages are used to browse and align these fields of view — unavoidably however these stages include some positioning errors. If left uncorrected these errors lead to large artifacts in the assembled virtual slide. In this paper we present a new image stitching method for Virtual Microscopy based on dyamic programming.

1.1 Image stitching

Szeliski presents an authoritative survey of image stitching (also known as mosaicing) in [14]. The major focus in image mosaicing is the reconstruction of 3D scenes from many viewpoints, beginning with the simple case of planar surfaces imaged from several viewpoints and working toward stereo matching and more general multiple-view reconstruction. As a result this research has focussed mainly on the estimation of rotation, affine and projective transformations using sparse feature-matching methods.

Levin *et al.* [9] investigated the minimisation of gradients across a stitching border to best stitch panoramic images. Panoramic image stitching is similar to the planar image stitching problem considered in this paper. However the focus of their work is on the best combination of two images at their borders by appropriate blending. This step comes after the images have already been aligned using another algorithm.

Duffin and Barrett [5] present a method to simultaneously align a large set of images by solving for the parameters of their projective tranformations. The solution is based on the Levenberg-Marquardt local minimisation technique. As a result a large overlap ($> 50\%$) is required between neighbouring images in order to obtain robust solutions.

The focus in this paper is on image stitching under an unknown set of translations. The translations form a highly constrained class of transformations (having only two parameters) and so we may consider specialised methods which would be inapplicable for the high-parameter transformations encountered in other image stitching or mosaicing applications.

2 Image stitching for virtual microscopy

2.1 Image acquisition

The slide scanning system consists of an AcCell 2000 (AccuMed International, Chicago) automated microscope, which includes an Olympus BX51 microscope plus an automated slide positioning stage. For colour imaging a high-resolution camera (Diagnostic Instruments, Sterling Heights, MI) and a light balanced daylight filter (80A) are used. Further details may be found in [3].

For the specimens examined in this paper, an area of approximately $15\text{mm} \times 10\text{mm}$ is scanned. Slides are scanned using a large number of fields of view — by rows from top-to-bottom, with individual rows being scanned in a zig-zag order (alternately left-to-right and right-to-left). Each field of view uses a $\times 40$ objective and produces a 2048×2048 image. As the camera uses $7\mu\text{m}$ CCD elements this results in an effective pixel resolution of $0.18\mu\text{m}$. Therefore each image covers an area of $400\mu\text{m} \times 400\mu\text{m}$, so at least 1000 fields of view are needed to scan one slide. In practice the fields of view are selected to overlap at their borders so it is necessary to acquire as many as 1500 fields of view. A typical virtual slide contains over 4 giga-pixels, therefore images are stored using the JPEG2000 image compression standard leading to substantial space savings [4].

In virtual microscopy a single large image, the 'virtual slide', is constructed from many small fields of view. The aim of image stitching is to perform this tiling so as to minimise visual artifacts at the image borders.

There are several sources of visual artifacts in the acquisition process. During slide scanning there may be changes in background lighting between neighbouring fields of view, due to changes in ambient lighting or to fluctuations in the lamp power supply. These can introduce a clearly visible tiled effect when the virtual slide is viewed at low magnification. In addition to lighting errors, geometric warping can occur due to radial distortion in the camera. This warps the individual fields of view which cannot be brought into perfect alignment at their borders. However these sources of error can be ameliorated by appropriate image correction. The dominant source of artifacts in the virtual slide is poor alignment of the images during the stitching phase. In particular, the slide positioning stage used in this work has an open loop controller which is only accurate to $\pm 10\mu m$, *ie.* ± 50 pixels. Even a closed looped controller will typically have $\pm 2\mu m$ (10 pixels) positioning error[1]. As a result the fields of view tend to drift out of alignment as a scan proceeds. This paper addresses the problem of recovering the unknown translation of these images.

As an example, Figure 1 depicts the result of image stitching using an existing 'Greedy' method as well the dynamic programming approach proposed in this paper. We observe a significant horizontal discontinuity running across the image where two fields of view have been stitched at their borders.

[1] http://autoscan.com.au

2.2 Problem formulation

Let us denote the original slide image $\mathcal{I} : \mathcal{Z}^2 \rightarrow \mathcal{C}$, *ie.* a mapping from the discrete 2D domain $\mathcal{Z}^2$ to a colourspace $\mathcal{C}$. The automated slide scanning process samples this image into a collection of subset images $I^{\vec{i}}$ where $\vec{i} = (i, j)$ encapsulates the column index i and row index j of the image in the scan. The goal of an image stitching algorithm is to invert this sampling process in order to construct an approximation $\hat{\mathcal{I}} \approx \mathcal{I}$ from the collection $I^{\vec{i}}$.

We have observed that the dominant source of geometric error is due to errors in translation by the automated slide positioning stage. Therefore in order to perform image stitching it is sufficient to estimate the positions $\vec{x^i}$ of each image $I^{\vec{i}}$ on the slide. Given that we expect some small distortions to remain in the reconstructed image $\hat{\mathcal{I}}$, we consider this problem as the optimisation of a suitably designed objective function. The objective function used in this paper has the form:

$$E[\{\vec{x^i}\}] = \frac{1}{2} \sum_{\vec{i}} \sum_{\vec{j} \in N(\vec{i})} \text{sim}\left(I^{\vec{i}} + \vec{x^i},\, I^{\vec{j}} + \vec{x^j}\right) \qquad (1)$$

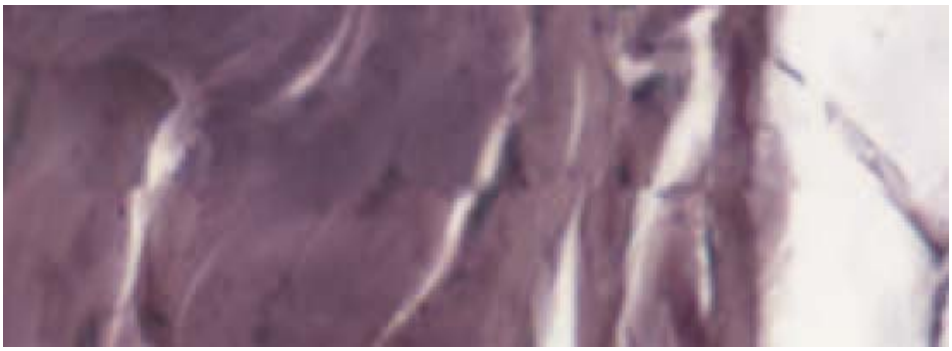
The objective function to be maximised in Equation 1 simply gives the total similarity between all neighbouring images after their translation has been recovered. Here $N(\vec{i})$ denotes the set of neighbouring images from image $\vec{i}$: above, left, right and below. $\text{sim}(\cdot, \cdot)$ denotes the similarity between these two images, while $I^{\vec{i}} + \vec{x^i}$ and $I^{\vec{j}} + \vec{x^j}$ denote the offset of image $I^{\vec{i}}$ (resp. $I^{\vec{j}}$) by the proposed translation $\vec{x^i}$ (resp. $\vec{x^j}$).

Figure 2 gives a visual interpretation of Equation 1. Figure 2(a) depicts a typical image stitching — images are placed in an approximate grid formation with small offsets according to the translational errors of the automated image acquisition system. Figure 2(b) depicts the graph of relationships between images. Images should align well (according to the similarity functional) with their neighbours above, below, left and right.

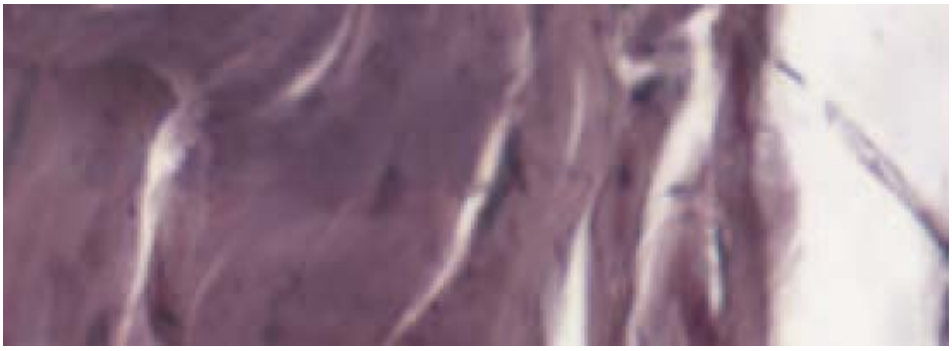
The similarity function used in this paper is the Zero-mean Normalised Cross Correlation (ZNCC) function,

$$\text{ZNCC}(A,\, B) = \frac{\left\langle A - \bar{A},\, B - \bar{B} \right\rangle}{\sqrt{\left\langle A - \bar{A},\, A - \bar{A} \right\rangle \left\langle B - \bar{B},\, B - \bar{B} \right\rangle}}$$

Here A and B are the two images to be compared, $\langle \cdot, \cdot \rangle$ represents the inner product and $\bar{A}$ (resp. $\bar{B}$) denotes the



(a)



(b)

Figure 1. An image stitching artifact. Observe the horizontal line across the middle of the image. (a) The result of stitching using the Greedy algorithm proposed previously. (b) The result of stitching using the Dynamic Programming algorithm proposed in this paper.

mean of A (resp. B). This similarity function is the optimal statistical estimator for translation under the assumption of additivite white Gaussian noise (AWGN) and unknown changes in background lighting and contrast [6]. It may be efficiently computed over a range of translations using the FFT [10].

3 Method

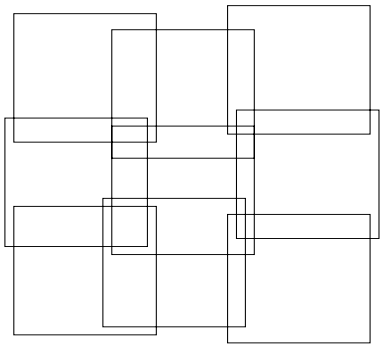
Two methods for image stitching are considered in this section. The first of these, the 'Greedy' method, estimates the location of each image in turn. The second method, the 'DP' method, estimates the location of an entire row of images simultaneously using a dynamic programming algorithm. Figure 3 depicts the two different schemes for stitching images. The Greedy method has been presented in [3] and is included here solely for comparison.

3.1 Greedy matching

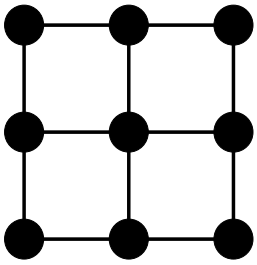
In the 'Greedy' method we place each image in turn in raster scan order. Consider the placement of image $I^{\vec{i}}$ with $\vec{i} = (i, j)$, after having already placed all previous images (i', j') with $j' < j$ or $j' = j \wedge i' < i$. Then we may simply select the location $\vec{x}^{\vec{i}}$ of image $I^{\vec{i}}$ so as to maximise the objective E in Equation 1:

$$\vec{x}^{\vec{i}} = \max_{\vec{y}} \sum_{\vec{j} \in N'(\vec{i})} \mathrm{sim}\left(I^{\vec{i}} + \vec{y},\, I^{\vec{j}} + \vec{x}^{\vec{j}}\right) \qquad (2)$$

where $N'(\cdot)$ denotes only those neighbouring images which have been placed previously (*ie.* the image above and the image to the left). In other words we place the current image simply to best match the neighbouring images which have already been placed, without regard to those images which have yet to be placed. A weighted variant of this method is presented in [3].

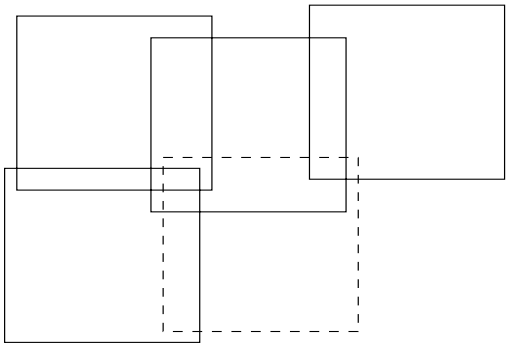


(a)

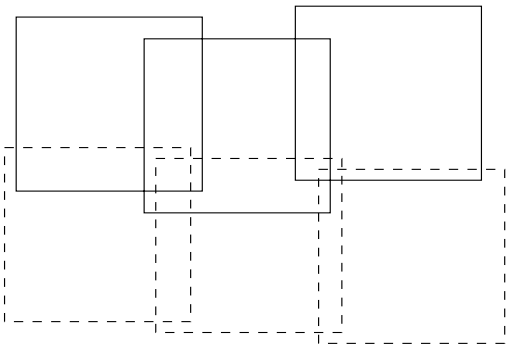


(b)

Figure 2. The image stitching model. (a) Images corresponding to fields of view are stitched into a large image of the entire slide. (b) The stitching grid. Points correspond to images while lines relate neighbouring images.



(a)



(b)

Figure 3. (a) Greedy image stitching places a single image at a time. (b) DP-based image stitching places a whole row of images simultaneously. Solid rectangles denote images placed beforehand, while dashed rectangles denote images being placed by an iteration of the method.

3.2 Matching by dynamic programming

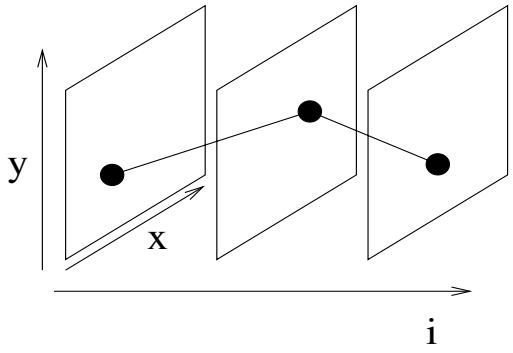
Whereas the Greedy matching method optimised the placement of a single field of view, here we present a method to optimise the placement of an entire row of images simultaneously. Consider the placement of the j^{th} row of images $I^{(\cdot,j)}$ after having already placed all previous images (i', j') with $j' < j$. Then we would like to select the locations $\vec{x}^{(\cdot,j)}$ of these images so as to maximise the objective E:

$$\vec{x}^{(\cdot,j)} = \max_{\vec{y}^{(\cdot,j)}} \sum_{(i',j-1)} \mathrm{sim}\left(I^{(i',j)} + \vec{y}^{(i',j)},\, I^{(i',j-1)} + \vec{x}^{(i',j-1)}\right) + \sum_{(i',j)} \mathrm{sim}\left(I^{(i',j)} + \vec{y}^{(i',j)},\, I^{(i'-1,j)} + \vec{x}^{(i'-1,j)}\right) \quad (3)$$

Here we decompose the sum into two components: the terms contributed by the prior placement of the row above, and the terms contributed by the relative placement of images within the current row. For clarity of expression here we have ignored boundary terms (*ie.* we ignore the fact that there is no 'left' image for the left-most image in this sum).

The optimisation problem posed by Equation 1 involves placing all images simultaneously. Here we have constructed a sub-problem (3) corresponding to placing an entire row of images simultaneously. This sub-problem is amenable to a solution by Dynamic Programming. Dynamic Programming (DP) is a template for a broad class of algorithms consisting of two parts [11]:

1. A recursive formulation of the problem



(a)

Figure 4. The trellis which forms the basis for image stitching by dynamic programming. i denotes the image index within the current row, while (x, y) denotes the position of the image. Overlaid is a maximum path corresponding to the optimal set of positions for the images in the current row.

2. A caching scheme to avoid solving the same problem twice

For problems which may be expressed using trellises, Dynamic Programming is more often known as Viterbi's algorithm [15].

To maximise Equation 3 we construct a trellis. A path through a trellis consists of a sequence of vertices connected by edges. The score for a path is the total of the scores for the vertices and edges along that path. The number of columns in our trellis will be the number of images in the row to be stitched. We equate vertices with potential positions for each field of view, so that a path through the constructed trellis then assigns positions to each image in the row.

Now by assigning suitable scores to the vertices and edges in this trellis, we may equate the score of any path with the objective function for row-wise stitching presented in Equation 3. In particular, the vertex scores in the trellis are simply the first component in Equation 3 while the edge scores are the second component in Equation 3. This is a common framework within image analysis and has been applied to stereo matching [8, 7, 12] and image segmentation [13, 1] for example. Further details of the transformation from Equation 3 to a trellis, or the consequent optimal path extraction may be found in [8] for example.

Figure 4 depicts this trellis for a row consisting of 3 images. A maximum path is overlaid which corresponds to the optimal placement of images within this row.

4 Results

4.1 Example

Figure 5 depicts an example of a stitching artifact at the vertical border between two images. Presented here are the results from Greedy matching (a) and from DP matching (b). In this example, the dynamic programming based algorithm produces a visually superior result to the Greedy method.

4.2 Visual trial

A visual trial was conducted so as to evaluate the results from the proposed DP method. Results from the Greedy method are used as a reference. The trial took the form of a two alternative forced choice (2AFC) test in which subjects are presented with stitching results from the same field of view, one computed using the Greedy method and the other using the DP method. Subjects were instructed to select the result which is 'visually superior'. While making their decision they were able to toggle between the two views, however a blank grey image is interposed for 2 seconds so that insignificant differences are not highlighted.

Twenty-six images containing stitching artifacts were selected from slide images stitched with both the Greedy and the DP methods beforehand. These images were taken from a collection consisting of a single cytology and three histology slides. Images were selected independently for both Greedy and DP. They were selected as being representative of the worst stitching artefacts present for each method on each slide. Six subjects were selected who had good or corrected vision; three were experienced in light microscopy and three were not. The images were presented on a standard 19 inch TFT monitor in a darkened room to optimise viewing conditions.

Table 1 presents the results collated by subject. It lists, for each observer, the number of preferences in favour of the DP method and the Greedy method. We now statistically analyse this data.

Our null hypothesis is that the two methods are indistinguishable, while the alternative hypothesis is that the DP method produces visually superior results to the Greedy method. According to the null hypothesis we expect to make six observations of a Binomial process with 26 data points and probability $p = 0.5$ in favour of the DP method. We recall that a Binomial process has mean $\mu = Np$ and standard deviation $\sigma = \sqrt{Np(1-p)}$. Pooling the results from the six subjects we obtain $N = 156$, $\mu = 78$ and $\sigma = 6.2$. Totalling the entries from Table 1 we obtain a score of 92:64 (DP:Greedy). Given the large ($N = 156$) number of observations, we may apply the theorem of de Moivre-Laplace to approximate this Binomial distribution

Table 1. Two alternative forced choice — results per viewer

Viewer	DP	Greedy
B1	18	8
A1	17	9
A2	16	10
S1	15	11
E1	13	13
S2	13	13
Total	92	64

Table 2. Two alternative forced choice — results per image.

DP:Greedy	Count
6:0	1
5:1	4
4:2	10
3:3	5
2:4	5
1:5	1
0:6	0

by a normal distribution with the same mean and variance. Subsequently we obtain a one-sided score $z = \frac{92-\mu}{\sigma} = 2.25$, which equates to a 98.8% probability that the alternate hypothesis should be accepted. Under the given testing conditions then, we may confidently state that the DP method produces visually superior results to the Greedy method. It is also worth noting that results for the three observers experienced in light microscopy (B1, A1 and A2) show this trend more strongly than for the other three observers.

Table 2 presents the results of DP vs. Greedy on the 26 independent images. Although there are only 6 observations per image, we may still analyse the resulting distribution. The null hypothesis, that there is no visible difference between the two methods, indicates that we should expect the results on each image to be Binomial distributed with $N = 6$ and $p = 0.5$. The probability of 5 or 6 viewers voting in favour of DP is then $\frac{7}{2^6} = 0.11$, ie. a 90% one-sided confidence interval requires that approximately 5 of the 6 viewers vote in favour of DP to obtain a statistically significant result. This is the case on 5 of the 26 images; on only one of the 26 images does Greedy achieve this (Figure 6). Therefore it is clear that the DP method provides an improvement over the Greedy method in the majority of cases.

4.3 Discussion

During the course of the visual trial the subjects noted that the major criterion for evaluating the quality of the stitching process was the presence (or absence) of discontinuities across the stitching border. This suggests that the use of the ZNCC similarity function may be misguided; perhaps in future work we could consider a function which more directly penalises these discontinuities.

We have presented a method to place a single image at each step (the Greedy method) as well as a method to place an entire row at each step (the DP method). It might seem that we could extend the current method to simultaneously optimise the placement of the entire image set. However dynamic programming can only be applied to optimise 1D sequences (such as rows or columns). In fact there is good reason to believe that obtaining the global optimum over all images is NP-hard [2].

We have noted that both matching methods seem to be fairly reliable when the images overlap significantly. Currently images overlap by 45% of their width and height at each border. However, such a large overlap necessitates a longer scan time — more than three times greater than if no overlap was required. In future we will be comparing the two stitching methods on scans with significantly less overlap. It is hoped that the DP method will be more robust than the Greedy method which will allow us to reduce the scanning time, which can be up to a day for a large slide at present.

5 Conclusion

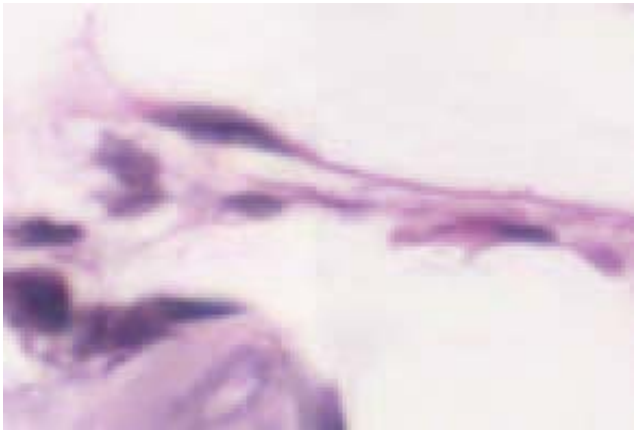
We have developed an image stitching algorithm based on dynamic programming and compared its performance to a simpler existing algorithm. The results demonstrate that the proposed method produces visually superior images, reducing the number and extent of stitching artifacts. Future work will use the new method to reduce the required overlap between images which has the potential to vastly reduce the slide scanning time.

References

[1] B. Appleton and C. Sun. Circular shortest paths by branch and bound. *Pattern Recognition*, 36(11):2513–2520, November 2003.

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. PAMI*, 23(11):1222–1239, November 2001.

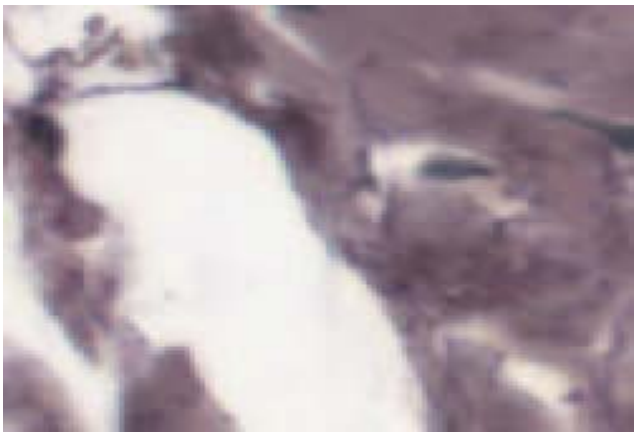
[3] A. P. Bradley, M. Wildermoth, and P. Mills. Virtual microscopy with extended depth of field. In *Digital Image Computing: Techniques and Applications*, Cairns, Australia, December 2005. Submitted.



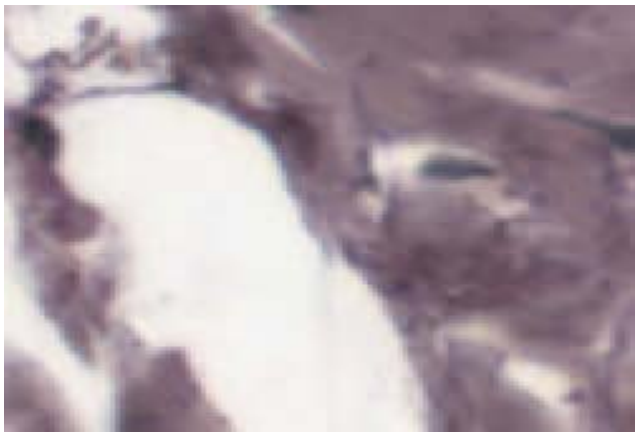
(a)

(b)

Figure 5. An example of a vertical discontinuity. (a) depicts the Greedy result; (b) depicts the DP result. Here the linear tissue structures toward the middle of the image are offset in the Greedy result. The result for this image was 5:1 in favour of the DP method.



(a)



(b)

Figure 6. A case where the Greedy algorithm demonstrated a statistically significant majority (1:5) over the DP algorithm. (a) depicts the Greedy result; (b) depicts the DP result. The tissue boundary shows a clear discontinuity in the DP result.

[4] C. Christopoulos, A. Skodras, and T. Ebrahimi. JPEG 2000 still image coding system: An overview. *IEEE Trans. Consumer Electronics*, 46(4):1103–1127, April 2000.

[5] K. L. Duffin and W. A. Barrett. Globally optimal image mosaics. In *Graphics Interface*, pages 217–222, 1998.

[6] S. Haykin. *Communication Systems*. Wiley, 3 edition, 1994.

[7] C. Leung, B. Appleton, B. C. Lovell, and C. Sun. An energy minimisation approach to stereo-temporal dense reconstruction. In *International Conference Pattern Recognition*, volume 1, Cambridge, United Kingdom, August 2004.

[8] C. Leung, B. Appleton, and C. Sun. Fast stereo matching by iterated dynamic programming and quadtree subregioning. In S. B. A. Hoppe and T. Ellis, editors, *British Machine Vision Conference*, volume 1, pages 97–106, Kingston, United Kingdom, September 2004.

[9] A. Levin, A. Zomet, S. Peleg, and Y. Weiss. Seamless image

stitching in the gradient domain. Technical Report tr:2003-82, Hebrew University, 2003.

[10] A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-Time Signal Processing*. Prentice-Hall, second edition, January 1999.

[11] R. Sedgewick. *Algorithms in C++*. Number 1–4. Addison-Wesley, third edition, 1998.

[12] C. Sun. Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques. *International Journal of Computer Vision*, 47(1/2/3):99–117, May 2002.

[13] C. Sun and B. Appleton. Multiple paths extraction using a constrained expanded trellis. *IEEE Trans. PAMI*, 2005. Accepted.

[14] R. Szeliski. Image mosaicing for tele-reality applications. Technical Report CRL 94/2, Digital Equipment Corporation, Cambridge Research Lab, May 1994.

[15] A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. on Information Theory*, IT(13):260–269, 1967.