FAST LINE-BASED IMAGING OF SMALL SAMPLE FEATURES

M. A. Iwen * *G. S. Mandair*† *M. D. Morris*‡ *M. Strauss*§

University of Michigan
markiwen, gmandair, mdmorris, martinjs@umich.edu

ABSTRACT

This project aims to reduce the time required to attain more detailed scans of small interesting regions present in a quick first-pass sample image. In particular, we concentrate on high fidelity imaging of small sample features via hyperspectral Raman imaging (e.g., small scale compositional variations in bone tissue [4]). The current standard procedure for high quality hyperspectral Raman imaging of small sample features consists of four steps: First-Pass Imaging, Detail Identification, Planning, and finally Detail Imaging. Traditionally, Detail Imaging and Planning have been carried out manually by human personnel—after acquiring some quick low-quality data in First-Pass Imaging, a researcher looks for interesting features (Detail Identification) and decides how to acquire higher-quality data for the interesting features (Planning), which is done in the final Detail Imaging phase. In this paper we will discuss automating the Detail Identification and Planning steps, resulting in a decrease of the procedure's total integration time. We fix an arbitrary way to automate Detail Identification and compare several different Planning methods. Our primary result is a method guaranteed to return a least cost (e.g., minimum integration time/number of scans) Detail Image under a general cost model. Because of their generality, the methodologies developed here may prove widely useful to basic biomedical scientists as well as to researchers in the pharmaceutical industry.

Index Terms— Biomedical imaging, Optimization methods, Raman spectroscopy, Bones, Graph theory

1. INTRODUCTION

Within the last several years many biomedical research groups have begun studying the compositional chemical properties that underlie the mechanical properties of bone. Unlike higher levels of architecture, the compositional level of bone was previously neglected due to the paucity of tools for non-destructive bone composition study. Recently the content and organization of bone at the molecular level has been successfully explored using Raman microspectroscopy and Raman imaging [1, 3, 4, 5]. These studies, as well as others in the literature, have begun to shed light on the molecular mechanisms of bone failure and response under both normal and diseased states.

An important hindrance to spectroscopic studies has been the long data acquisition time required for Raman microspectroscopy and Raman imaging. The time required to acquire a 256×256-pixel Raman image now varies between about 30 minutes and several hours. Reasons for this long imaging time include the tendency for current image acquisition protocols to be simple, manual, and non-adaptive. For example, during sample imaging a constant integration (acquisition) time is traditionally used at every data point despite the fact that there are usually several different optimal integration times for different types of regions.

Currently, small-scale sample features are imaged via Raman spectroscopy in four steps. First, during First-Pass Imaging, a low fidelity neighborhood image is quickly obtained. Then, during Detail Identification, the first-pass image is used to identify small interesting features—this stage is often done manually by a human expert. That expert then plans how to gather data during the fourth step. Finally, during Detail Imaging, the specimen is imaged again according to the plan to gather high quality detail data. In this paper we will propose automating Detail Identification and Planning with the following goals:

- Make Detail Identification more reliable and more repeatable than current manual processes. We expect our proposal to make this stage quicker as well, though we have not investigated this experimentally.

- Make the Planning phase provably optimal or nearly-optimal in the sense of minimizing the time for subsequent Detail Imaging.

2. BACKGROUND AND METHODOLOGY

For the remainder of the paper we will consider each Raman image to be an $n \times m$ array of spectral data. Every image location (i, j) will correspond to a physical location in row i and column j of the sample. Each column of the image is

gathered by one scan. Hence, given that each scan provides n pixels of spectral data, it takes m scans to produce an $n \times m$ image. During each scan, a sample column of data is illuminated with a laser while the induced radiation from each of the sample column's n data points is measured with an EM-CCD detector. In general we'd like to reduce the total imaging integration time not only for increased speed, but also to minimize potential sample damage due to the laser illumination. Hence, given a small collection of interesting sample positions to be imaged with a long integration time, we'd like to minimize the number of long scans required to cover the interesting sample positions.

In this paper, our focus is the comparison of different methods for the Planning phase. To that end, we will fix a method for Detail Identification. We discuss this further in Section 4.

The purpose of this paper is to propose a new method for Raman imaging and give theoretical and proof-of-concept support using a small amount of data. Ultimately, the effectiveness of our methods must be validated using many samples; that will be the subject of future work. We will avoid asking questions that can only be addressed by examining many samples.

3. OPTIMAL COLUMN/ROW SCANNING

In this section, we assume that Detail Identification has been performed, resulting in a set P of interesting pixels in the $[n] \times [m]$ grid. We address the Planning stage.

Traditionally, only columns are scanned. Once the sample is fixed, imaging only takes place by acquiring frames (scanning columns) from left to right. However, it is generally possible to rotate the specimen by $90°$. We therefore consider the more general problem of minimizing the number of long column and/or row scans required to cover a small number of interesting sample points.

Definition 1 *Given a set $P \subseteq [n] \times [m]$ of p interesting pixel locations, a set $U = C \cup R$ is a* feasible cover *of P if $C \subseteq [m]$ is a set of columns and a set $R \subseteq [n]$ of rows such that, for every $(i, j) \in P$, either $i \in R$ or $j \in C$.*

A feasible cover U of P is optimal *if it has the minimum size of all feasible covers.*

The set P is typically derived from quick First-Pass Imaging. See the 4×3 rectangular image in Figure 1 for an example problem.

In the Figure 1 example image we'd like to scan the five black pixels. Hence, our set of interesting pixels is $P = \{(1,1), (1,2), (1,3), (3,3), (4,3)\}$. Our task is to find the minimum number of columns and/or rows to scan in order to image all 5 black pixels.

We next compare three methods for obtaining feasible covers. They all take a set P of p interesting pixels, and return a set of columns C and/or rows R to be scanned in order to cover P. The three methods are:

3.1. Push Broom

Let $x = min\{j \mid \exists i \in \mathbb{N} \text{ with } (i,j) \in P\}$ and $y = max\{j \mid \exists i \in \mathbb{N} \text{ with } (i,j) \in P\}$. Scan $C = \{x, x+1, \ldots, y-1, y\}$ and $R = \emptyset$.

The Push Broom method is essentially the current standard method for scanning a small number of interesting pixels. After quickly obtaining a low fidelity first-pass image, a set of interesting pixels is obtained. The entire region from leftmost to rightmost column containing interesting pixels is then rescanned from left to right with a higher integration time.

3.2. Optimal Columns

Scan column set $C = \{j \mid \exists i \in \mathbb{N} \text{ with } (i,j) \in P\}$ and row set $R = \emptyset$. In effect, scan every column containing an interesting pixel.

3.3. Optimal Rows + Columns

Scan any cover of P that is Optimal.

It is straightforward to implement the Push Broom and Optimal Columns methods. Algorithms for Optimal Rows + Columns have been known [6]; we include a brief discussion for completeness and to illustrate the computational cost.

Algorithm 2 *The FF algorithm for computing an Optimal Rows + Columns cover of input P is as follows.*

1. *Construct a **scan graph** for P. The scan graph of P is a directed weighted graph, G, with node set $\{s,t\} \cup \{1, 2, \ldots, n\} \cup \{1, 2, \ldots, m\}$ and edge set $\{(s,i) \mid 1 \le i \le n\} \cup P \cup \{(j,t) \mid 1 \le j \le m\}$. All edges from the source node s and into the termination node t have a weight of 1. All remaining P edges are given a weight of ∞.*

2. *Use the Ford-Fulkerson method [2] to find a minimum cut of G.*

3. *Using the final resulting residual network we let C be the set of columns reachable from s and R be the set of rows not reachable from s.*

We omit the proof of the following.

Theorem 3 *Algorithm 2 produces an Optimal Rows + Columns cover of its input, P.*

Example 4 *Recall the Figure 1 example image. Figure 1's middle graph is the scan graph for the 4×3 image with $P = \{(1,1), (1,2), (1,3), (3,3), (4,3)\}$. Figure 1's rightmost graph gives the* residual network *that arises using the Ford-Fulkerson algorithm for a minimum cut in the scan graph. In the rightmost graph all gray nodes are reachable from the source node s. All white nodes are unreachable*
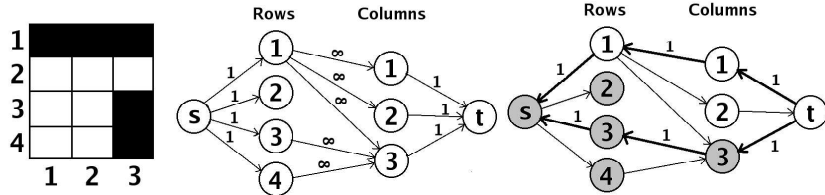
Fig. 1. An Example Problem, The Problem's Related Scan Graph, and a Scan Graph Solution

from s. Note that the gray(reachable) column 3 and white(not reachable) row 1 nodes provide us with an Optimal Rows + Columns cover of P. By inspecting the example image we can see that scanning row 1 and column 3 is indeed a minimal way of imaging P. Furthermore, we can see that if we only use columns or rows alone it will require 3 scans to cover P as opposed to only 2 scans.

The computational cost to run Algorithm 2 is polynomial in the size of the input, P. Note that the size of P is at most the total number mn of *possible* pixels; in the context where this algorithm is used, we expect that $|P| \ll mn$. For a 256×256-pixel image, we expect that the time to compute an Optimal Rows + Columns cover of P will be less than the time to acquire data in the Detail Imaging step. In any case, our focus in this paper is minimizing the data acquisition time, which we equate with sample damage; we mention that computation time is acceptable.

4. EMPIRICAL EVALUATION

We compare the performance of Push Broom, Optimal Columns, and Optimal Rows + Columns on two test problems. For both test problems we assume that scanning any row and/or column is just as costly as scanning any other. All non-P scan graph edges are given a weight (cost) of one.

See Figure 2 for the first test image and results. For our first test we let I be the noisy Figure 2 "HELLO" image and let the set of interesting pixels, P, be the lightest p pixels in I. Note that this first test contains a variety of both horizontal and vertical bands of light (i.e., interesting) pixels. As a result we can see in Figure 2's results graph that the Optimal Rows + Columns method requires substantially fewer columns and rows than the other two methods to cover the lightest $p \leq 30\%$ of I's pixels. Between the Optimal Columns and Push Broom methods we can see that the Optimal Columns method outperforms the Push Broom method for covering a very small (i.e., less than about 2%) number of the lightest pixels. However, both Push Broom and Optimal Columns are about the same cost for larger p.

See Figure 3 for the second test image. In Figure 3 our image I is a first-pass Raman image of test sample consisting of mouse bone embedded in PMMA plastic. Here the lighter

pixels correspond to bone while darker pixels correspond to PMMA. Gray pixels indicate bone covered by a thin layer of PMMA. Here our pixels of interest, P, are the p boniest (lightest) pixels in I. Here we *assume* that choosing the p boniest pixels, for various p, according to the low-fidelity First-Pass image is a good way to do Detail Identification; properly addressing this question is beyond the scope of this paper.

Figure 3's first-pass bone + PMMA image, I, was produced by scanning each of the 60 image columns with a 1 second integration time. We would like, however, to scan each bony (interesting) pixel for 8 seconds. Hence, Figure 3's result graph reports $60 + 8(\#$ columns/rows to cover $P)$ seconds for each method. There we can see that both the Optimal Columns and Optimal Columns + Rows methods outperform Push Broom for scanning the lightest $p \leq$ about 15% of I's pixels.

Note that Figure 3's first-pass bone + PMMA image, I, is biased toward a strong Optimal Columns performance. Not only does each of I's columns cover more than three times as many pixels as each row, but all of I's boniest (i.e. lightest) features are aligned vertically. Generally this will not be the case. However, even for this difficult test image, Optimal Rows + Columns still requires less scan time for most small $|P|$ values (i.e. $< \approx 5\%$ pixels scanned).

5. GENERALIZATIONS AND FUTURE WORK

In the Optimal Rows + Columns method there is some flexibility with respect to the edge weights assigned in the scan graph. Although all P pixel edges should always be given a weight of ∞, the remaining edges from the s node and into the t node need not all have weight 1. In general the weight assigned to an edge (s, i) should correspond to the cost of scanning row i. Likewise, the weight assigned to an edge (j, t) should correspond to the cost of scanning column j. If, as above, all non-P edges are assigned the weight 1 it means that all rows and columns require the same unit of cost to scan. However, each non-P column/row scan graph edge can indeed be given any desired positive real cost. This leaves the user a good deal of flexibility in assigning row and column costs based on the first-pass image I.

Angles other than $90°$ can be considered as well. If each pixel is in more than two possible frames (horizontal and ver-
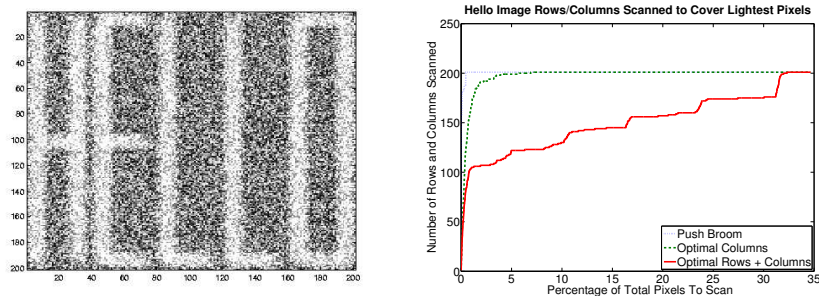
Fig. 2. Test Image Along with the Number of Rows+Columns Required to Cover Its Lightest Pixels
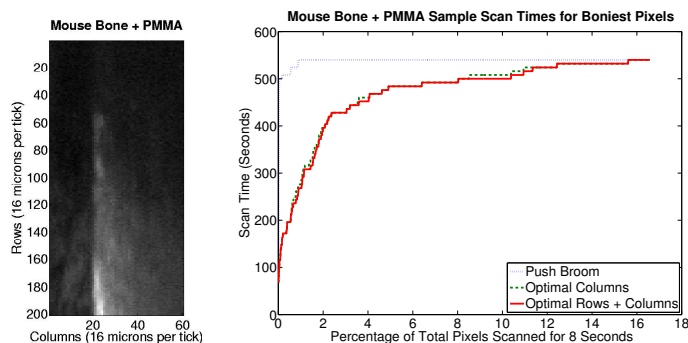

Fig. 3. Bone + PMMA Image, and The Total Time Required to Image Its Boniest (Lightest) Pixels

tical), we know of no efficient computation of an optimal cover. There are, however, fast approximate algorithms [2] for the set-cover problem, including a greedy algorithm, with an approximation ratio of $\ln(\max(m,n))$. There is also inherent approximation involved in using data from one pass in order to predict the outcome of a second pass rotated by an angle that is not a multiple of $90°$. In particular, if the pixels are square, pixels of one pass do not line up exactly with pixels of the second pass. We do not discuss that further here.

Jitter and hysteresis effects on the scanner realignments necessitated by the Optimal Columns and Optimal Rows + Columns methods should also be more thoroughly investigated. However, we don't expect these effects to be important. The spectrometer used to produce the Figure 3 test image utilizes a mirror which can be positioned to better than 0.2 microns [3] (small in comparison to Figure 3's 16 micron length scale). Stages exist with similar precision. Furthermore, hysteresis effects can be mitigated by beginning Detail Imaging behind each column/row starting point and then progressing with column/row scans in only one direction.

6. CONCLUSION

In this paper we demonstrated that two proposed scanning methods, Optimal Columns and Optimal Columns + Rows,

may be useful in decreasing the total integration time required to rescan a small set of interesting image pixels.

7. REFERENCES

[1] A. Carden, M. D. Morris, R. M. Rajachar, and D. H. Kohn. Ultrastructural changes accompanying the mechanical deformation of bone tissue: A raman imaging study. *Calcified Tissue International*, 72(2):166–175, 2003.

[2] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. Introduction to algorithms. *2nd Edition*, pages 643–664, 2001.

[3] K. Golcuk, G. S. Mandair, A. F. Callender, N. Sahar, D. H. Kohn, and M. D. Morris. Is photobleaching necessary for raman imaging of bone tissue using a green laser? *Biochimica et Biophysica Acta*, 1758(7):868–873, 2006.

[4] M. D. Morris, W. F. Finney, and R. M. R. et al. Bone tissue ultrastructural response to elastic deformation probed by raman spectroscopy. *Faraday Discussions*, 126:159–168, 2004.

[5] C. P. Tarnowski, M. Ignelzi, and W. W. et al. Earliest mineral and matrix changes in force-induced musculoskeletal disease as revealed by raman microspectroscopic imaging. *Journal of Bone and Mineral Research*, 19(1):64–71, 2004.

[6] D. Wagner. Efficient algorithms and intractable problems, April 2003. UC Berkeley CS 170 Handout 20.