AUTOMATIC LATTICE DETECTION IN NEAR-REGULAR HISTOLOGY ARRAY IMAGES

Brian A. Canada, Georgia K. Thomas, Keith C. Cheng, James Z. Wang, and Yanxi Liu

The Pennsylvania State University, University Park, PA, USA

ABSTRACT

Near-regular texture (NRT), denoting deviations from otherwise symmetric wallpaper patterns, is commonly observable in the real world. Existing lattice detection algorithms capture the underlying lattice of an NRT pattern and all of its individual texels, facilitating an automated analysis of NRT. Many real world images, as in those of zebrafish larval histology arrays, depart significantly from regularity and challenge the current state of the art wallpaper group theory-based lattice detection methods. We propose an alternative 2D lattice detection algorithm that exploits translation and reflection symmetries and specific imaging cues. By outperforming existing methods on histology array images, our algorithm leads us towards complete automation of high-throughput histological image processing while broadening the spectrum of NRT computation.

Index Terms—lattice estimation, biological tissues, biomedical image processing

1. INTRODUCTION AND BACKGROUND

Near-regular texture (NRT) is defined [1] as minor geometric, photometric, and topological deformations from an otherwise translationally symmetric 2D wallpaper pattern. NRT can be ubiquitously observed in both man-made and natural objects and their images, such as a checkerboard pattern on a wrinkled tablecloth, a brick wall with color and texture variations among the individual bricks, the hexagonal cells of natural honeycomb, or the scales that make up the curved surface of a fish or shark.

Significant progress has been made in the automated detection of NRT patterns, whether static [1, 2, 3] or in motion [4]. The detection algorithms used in these works are based on wallpaper group theory [5] in that the topology of such patterns can be characterized completely by a quadrilateral "lattice" composed of fundamental generating units, called *texels* or *textons*. Indeed, regular quadrilateral lattices can be represented sufficiently by only two linearly independent vectors.

B.C. (corresponding author – email: canada@psu.edu) is with the Integrative Biosciences Program and is supported by the Penn State Academic Computing Fellowship. G.T. and K.C. are with the College of Medicine and are supported by NIH, the Life Sciences Greenhouse of Central Pennsylvania, and Pennsylvania Department of Health Tobacco Settlement Funds. J.Z.W. is also with Carnegie Mellon Univ. (CMU) and is supported by NSF. Y.L. is also with CMU, as well as the University of Pittsburgh, and is supported by NSF, NIH, and the PA Department of Health.

Fig. 1. The input zebrafish histology image: The challenge here is to estimate the 2D lattice structure implied in fiduciary marker-free, near-regular, sparse, and noisy zebrafish histology images such as that shown above.

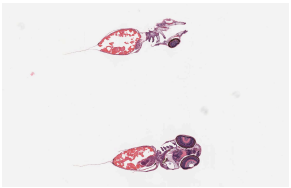

Fig. 2. Selected specimens from Fig. 1 at higher magnification, illustrating the typical differences among "texels" (tiles) of the same 2D "wallpaper" pattern.

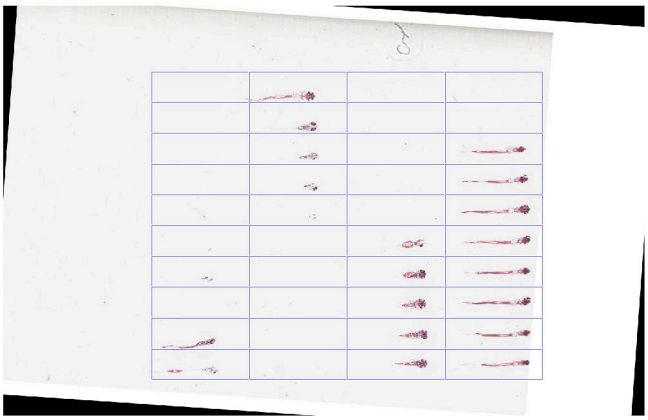

Fig. 3. Output from our algorithm: The lattice overlaid on the rotation-corrected input array image is detected by exploiting the topological 2D lattice structure and maximizing bilateral symmetry of the zebrafish larva in each lattice unit cell.

In many real world image patterns, we have observed significant departures from regularity, with large gaps in the NRT pattern that give it a "sparse" appearance (Fig. 1) or severe differences among texels of the same NRT (Fig. 2), such that the current state of the art automatic lattice detection algorithm [2, 3] appears inadequate to extract the implied complete lattice (Fig. 3). On the other hand, certain application-specific cues can be used to enhance the reliability of automatic lattice extraction. Here, we introduce a new method for 2D lattice detection that accounts for sparse and irregular NRT while taking advantage of unique spatial constraints in array images of zebrafish histology.

1.1. High-throughput zebrafish histology imaging

Histology—the microscopic study of biological tissues—is a highly sensitive method of observing mutant phenotypes (observable traits resulting from an induced genetic mutation) at fine levels of detail, on the order of 1 μm or less. Because of their small size, zebrafish larvae are particularly amenable to "high-throughput" histology (Fig. 4), a procedure developed by Cheng and colleagues [6] in which arrays of up to 50 zebrafish larvae (10 rows by 5 columns) at different stages of development are fixed to preserve morphology, embedded in agarose gel, processed into tissue blocks by paraffin infiltration, and sliced into thin sections. The sections are individually mounted on glass microscope slides for staining and then scanned to produce high-resolution "virtual slides" for scoring and annotation of mutant phenotypes.

The phrase "high-throughput," as used in this context, does not currently apply to all steps in the workflow, however. While multiple larvae may be processed and digitized "in parallel," the process of scoring each image is rate limiting. Comprehensive annotation of the phenotype requires scoring several levels (planar sections) of each larva. While this process is manageable when manually annotating a small number of larvae, the task becomes increasingly impractical as the number of larvae rises.

Although we have developed prototype systems for automatically scoring histological abnormalities in zebrafish organ images [7, 8], the manual image pre-processing required to prepare such images is arguably even more time-consuming than manual image annotation. The first (and slowest) step in pre-processing involves the extraction of specimens from the original virtual slides and then matching each specimen to its corresponding entry in the laboratory's embedding record, which keeps track of the age and general classification (normal, mutant, or unknown) of the specimens embedded in a particular tissue block (Fig. 5). Certainly, an automated method for specimen extraction would improve the rate of overall throughput in the histology workflow, but the nature of the slide preparation process introduces a number of obstacles that such a method must account for.

1.2. Challenges to automatic lattice detection

In the ideal case of a virtual slide with a perfectly horizontally-oriented rectangular array of 50 visible specimens and

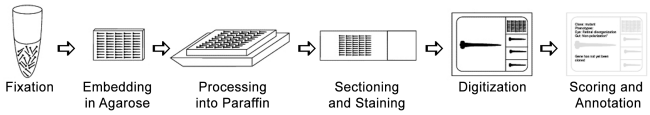

Fig. 4. Current laboratory pipeline for "high-throughput" histology [6]. The steps up to and including digitization can be automated and/or conducted in parallel, but the overall process is rate-limited by the tedious process of scoring and annotation.

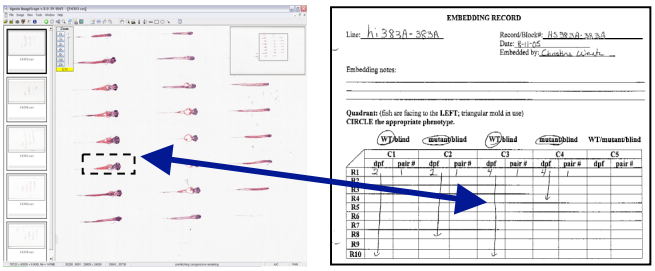

Fig. 5. The most time-consuming task in the manual pre-processing of histology array images is the extraction of individual specimen images from the original virtual slides and matching them to their corresponding positions in the laboratory embedding record. Our algorithm enables the automation of much of this task.

no artifacts or background noise, the extraction process would simply involve locating a rigid 10 row by 5 column lattice with each lattice unit cell bound to its corresponding position on the embedding record. Because of variability in both experimental design and in the handling and preparation of glass slides prior to digitization, such an ideal case is rarely achieved. Sections are often mounted on the slide at a slight angle, dust and other artifacts may be present, and since tissue block embedding depths often vary from specimen to specimen, a particular section may not appear to contain specimens in all positions, giving the array an irregular or sparse structure (Fig. 1).

Perhaps the problem could be solved through the use of fiduciary markers, which if embedded properly could provide a built-in set of interest points or landmarks to assist the computer in detecting the rectangular boundary around the larval array. For example, a certain type of hair or other thin filamentous material could be embedded at each corner of the tissue block to demarcate the array boundary, which would be especially helpful for highly irregular sections. However, the manual process of preparing and slicing a tissue block is already error-prone; including an *additional* step of embedding fiduciary markers so that they remain normal to the slide surface and do not shift position or cause tearing or wrinkling during the sectioning process is not only impractical but further increases the likelihood of error.

In order to develop an automated, fiduciary marker-free method for precise placement of a rectangular lattice over a zebrafish array, we are forced to work only with those features that already exist in the array images. At high magnification, it is clear that individual larvae in a zebrafish histology

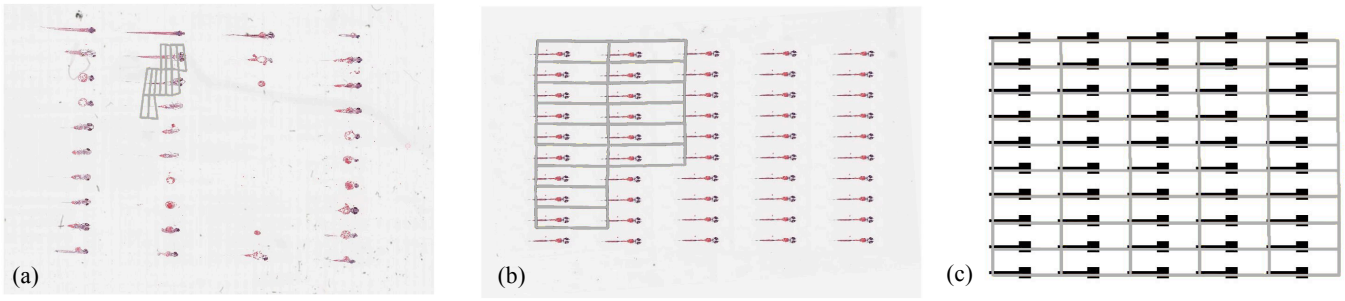

Fig. 6. Results of lattice detection in selected real and simulated zebrafish array images using the Hays *et al.* algorithm [2, 3]. Image (a) is taken from an actual virtual slide, (b) is a "synthetic array" of 50 identical larvae, and (c) is an array of 50 "simulated" larvae.

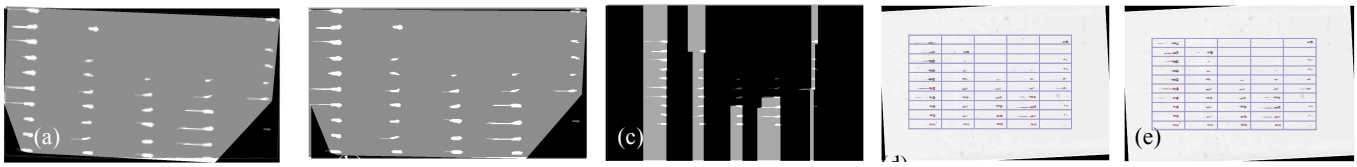

Fig. 7. Illustration of our new lattice detection algorithm. In (a), the convex hull around the array specimens is rotated until the fraction of the bounding box area occupied by the convex hull is maximized as in (b), noting that the bounding box "shrinks" to fit the convex hull as it rotates. In (c), the number of columns is determined by a morphological closing operation. The bounding box is then divided into the initial lattice (d), which is optimized by varying the lattice height until the reflective symmetry in each cell is maximized (e).

array vary significantly in their morphological appearance (Fig. 2). On the other hand, a lower magnification will blur these distinctions and give a more regular appearance to the array, owing primarily to the fact that the larvae are all oriented in the same direction, they are well-aligned and are generally equidistant from one another, and they are more or less bilaterally symmetric. As a result, we can treat each array image as a 2D NRT deviating from a regular "wallpaper" pattern and capture its near-regularity quantitatively via lattice detection.

2. METHODS AND RESULTS

Virtual slides of zebrafish arrays at 0.4× magnification (with an approximate resolution of 1300 × 900) were generated as described in [7] and converted to JPEG format. It should be noted that, at a typical resolution of approximately 60,000 × 40,000, the original virtual slides at full (20×) magnification are too large to be processed in any reasonable amount of time. It makes more sense to detect the lattice shape and position at a lower magnification and then scale the lattice to full magnification for further analysis of the extracted specimens.

2.1. Evaluation of current state of the art

Estimation of lattice patterns in static NRT images has been studied previously by Hays *et al.* [2, 3]. In that work, the patterns exhibit relatively smooth deformations from one lattice unit to the next; these deformations could be due to affine transformations, changes in linear perspective, or perhaps more complex physical and geometric distortions. Treating images of zebrafish histology array as "distant relatives" of near-regular textures, it is logical for us to first evaluate the Hays *et al.* algorithm's potential for detecting the implied 2D lattice in a zebrafish larval array.

Three images were used in testing; the first of these was an arbitrarily chosen representative of typical zebrafish array appearance. The second image was a "synthetic" array image of 50 identical copies of the same specimen produced using Adobe Photoshop. A third image, generated using Adobe Illustrator, contained 50 identical block-like shapes designed to "simulate" the silhouette of a developing zebrafish embryo. As Fig. 6 indicates, the Hays *et al.* algorithm was unable to completely detect the proper lattice even for those images that were specifically *designed* to have a perfectly repeating wallpaper-like pattern, though it did come closest when processing the array of "simulated" fish. Computation time ranged from 30-60 minutes per image on a 1.66 GHz Intel Core Duo processor. Consequently, we chose to design a new algorithm that accounts for sparsity, noise, and other near-regularities encountered in zebrafish histology arrays and similar images.

2.2. Our lattice detection algorithm

Our algorithm proceeds in three stages. In stage 1 (Figs. 7a,b), the rotation offset of the array image, if any, is corrected so that all specimens are oriented horizontally. In stage 2 (Figs. 7c,d), we use the detected number of cells in the array to divide the bounding box circumscribing the specimens into an initial lattice, taking advantage of the upper bounds of 5 columns and 10 rows imposed by the histology array apparatus design. In practice, we have observed that is it sufficient to assume that the lattice will contain ten rows, even if some rows are partially or totally unoccupied. The number of columns in the field of view, however, will vary from 1 to 5 depending on the design of that particular histology experiment. Finally, in stage 3 (Fig. 7e), we optimize the initial lattice by varying its height until we maximize the symmetry of each lattice cell image about its horizontal reflection axis. The algorithm pseudocode is provided as follows:

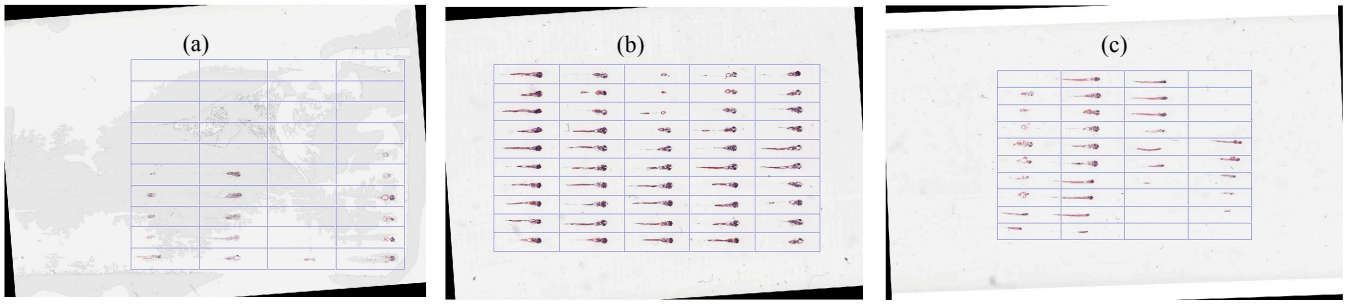

Fig. 8. Selected output images as processed by our lattice detection algorithm. In the vast majority of tests, as with (a) and (b), the algorithm detected lattices with 100% accuracy. Note that the barely-visible specimen in the upper right corner of (a) was detected even with significant image noise. Lattice detection was suboptimal for (c) as a result of specimen shifting due to improper sectioning.

Stage 1: Global orientation correction

1 A_{convhull} = area of convex hull around array specimens
2 A_{box} = area of bounding box around convex hull
3 $\theta_{\text{corrected}}$ = argmax ($A_{\text{convhull}} / A_{\text{box}}$)

Stage 2: Rigid lattice placement

4 Perform morphological closing operation
5 n_{columns} = number of connected components
6 n_{rows} = 10 (assumed)
7 L_{initial} = result of dividing bounding box around foreground pixels into lattice with ($n_{\text{rows}} n_{\text{columns}}$) equally sized cells

Stage 3: Deformable lattice optimization

8 For each grayscale cell image C in lattice L
9 $S = corr(C, mirrorimage(C))$
10 L_{final} =argmax(S) for all C

2.3. Summary of lattice detection results

Our algorithm was tested on a set of twenty different virtual slide images. These were chosen to be representative of the types of variation one can expect to observe, including differences in overall array shape and orientation, number and sparsity of specimens, quality of sections, as well as background noise and other image artifacts such as dust particles. In comparison to the 30-60 minutes required for processing by the Hays *et al.* algorithm [2,3], the computation time for a typical 1300 × 900 image using our method was less than 5 minutes using a 1.66 GHz Intel Core Duo processor. However, reducing the image dimensions by 50% (from 1300 × 900 to 650 × 450, for example) yielded a substantially improved computation time of about one minute or less, although results can vary depending on the step sizes used in the rotation correction and lattice optimization routines as well as the number and size of individual cells detected in the lattice. Example outputs are shown in Fig. 8.

100% lattice detection accuracy was achieved in 19 of the 20 images tested—even those with significant noise and artifacts (*e.g.,* Fig. 8a). One image yielded 85% accuracy (Fig. 8c), which we attribute to the translational symmetry within the wallpaper pattern being distorted, in that some larvae appear to have shifted from their original positions due to improper tissue block preparation and sectioning.

3. CONCLUSIONS

Our proposed algorithm represents an effective new computational method for lattice-based segmentation in histology or other types of images obtained by array imaging. Practically, our algorithm is a key step forward towards complete automation of image processing in the high-throughput histology pipeline, since one of the most laborious steps in such a workflow involves the extraction of individual specimens from the original virtual slides and matching them to the lab's embedding record. No fiduciary markers are needed for our algorithm to detect the desired lattice patterns with reasonably high accuracy, and it outperforms the current "state of the art" computer vision algorithms for lattice detection in both accuracy and speed. Given the effective computational treatment of histology arrays or other types of array-like images with significant sparsity and irregularity, our algorithm extends the current "near-regular texture" spectrum to a new horizon.

4. REFERENCES

[1] Y. Liu, R. Collins, and Y. Tsin, "A Computational Model for Periodic Pattern Perception Based on Frieze and Wallpaper Groups," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, No. 3, pp. 354 – 371, 2004.

[2] Y. Liu, W. Lin, and J.H. Hays, "Near Regular Texture Analysis and Manipulation," *ACM Transactions on Graphics (SIGGRAPH 2004)*, Vol. 23, No. 3, pp. 368-376, 2004.

[3] J.H. Hays, M. Leordeanu, A.A. Efros, and Y. Liu, "Discovering Texture Regularity as a Higher-Order Correspondence Problem," *9th European Conference on Computer Vision (ECCV)*, 2006.

[4] W. Lin and Y. Liu, "A Lattice-based MRF Model for Dynamic Near-regular Texture Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 29, No. 5, pp. 777-792, 2007.

[5] B. Grünbaum and G.C. Shephard, *Tilings and Patterns*, New York: W. H. Freeman and Company, 1987.

[6] N.A. Sabaliauskas, C.A. Foutz, *et al.*, "High-throughput Zebrafish Histology," *Methods*, vol. 39, pp. 246-254, 2006.

[7] B.A. Canada, G.K. Thomas, *et al*., "Automated Segmentation and Classification of Zebrafish Histology Images for High-Throughput Phenotyping," *Proc. of the 3rd IEEE-NIH Life Science Systems and Applications Workshop*, pp. 245-248, 2007.

[8] B. A. Canada, G. K. Thomas, *et al.*, "Towards Efficient Automated Characterization of Irregular Histology Images via Transformation to Frieze-Like Patterns," *Proc. of the ACM Int'l Conference on Image and Video Retrieval (CIVR),* July 2008.