Optimization of complex cancer morphology detection using the SIVQ pattern recognition algorithm

Jason Hippa,1, Steven Christopher Smitha,1, Jerome Chenga, Scott Arthur Tomlinsa, James Monacob, Anant Madabhushib, Lakshmi Priya Kunjua and Ulysses J. Balisa∗

a Department of Pathology, Medical Science, University of Michigan, Catherine, Ann Arbor, MI, USA
b Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

Abstract. For personalization of medicine, increasingly clinical and demographic data are integrated into nomograms for prognostic use, while molecular biomarkers are being developed to add independent diagnostic, prognostic, or management value [1]. In a number of cases in surgical pathology, morphometric quantitation is already performed manually or semi-quantitatively, with this effort contributing to diagnostic workup. Digital whole slide imaging, coupled with emerging image analysis algorithms, offers great promise as an adjunctive tool for the surgical pathologist in areas of screening, quality assurance, consistency, and quantitation. We have recently reported such an algorithm, SIVQ (Spatially Invariant Vector Quantization), which avails itself of the geometric advantages of ring vectors for pattern matching, and have proposed a number of potential applications. One key test, however, remains the need for demonstration and optimization of SIVQ for discrimination between foreground (neoplasm- malignant epithelium) and background (normal parenchyma, stroma, vessels, inflammatory cells). Especially important is the determination of relative contributions of each key SIVQ matching parameter with respect to the algorithm’s overall detection performance. Herein, by combinatorial testing of SIVQ ring size, sub-ring number, and inter-ring wobble parameters, in the setting of a morphologically complex bladder cancer use case, we ascertain the relative contributions of each of these parameters towards overall detection optimization using urothelial carcinoma as a use case, providing an exemplar by which this algorithm and future histology-oriented pattern matching tools may be validated and subsequently, implemented broadly in other appropriate microscopic classification settings.

1. Introduction

For personalized medicine, clinical and demographic data are being integrated into nomograms for prognostic use, while molecular biomarkers are being developed to add independent diagnostic, prognostic, or management value [1]. In the setting of diagnostic surgical pathology of several diseases, morphometric quantitation of key histologic characteristics is performed manually or semi-quantitatively on H&E sections, though such efforts remain plagued by significant inter- and often intra-observer variability [2–5]. To address these concerns, several adjuncts, availing themselves of advancing digital slide technologies, have emerged to provide greater standardization and precision, including implementations, reaching FDA approval for clinical use (such as Ventana VIAS Image Analysis System™, Aperio’s Digital IHC solution™, Bioimagene’s PATHIAMTM system with iSCAN™, and Bioview’s Duet™). However, with the increasing acceptance of digital slide workflow, and improvement in image quality and scanning speed, the application of H&E image analysis...
algorithms may offer important opportunities for screening, quantitation, and standardization to practice while sparing generation and testing of novel hypotheses regarding relationships between histologic features and disease classification and risk stratification.

Image processing techniques, such as pattern recognition algorithms, offer the potential to rigorously and consistently identify and quantify morphologic and architectural features unique to a diagnosis [6–10]. However, the technical demands of image sampling in the x and y translation of freedom, 360-degree rotation, and mirror symmetry, in conjunction with morphologic heterogeneity of tissue (due to biological heterogeneity and tissue processing artifacts), yield astronomical combinations of image features prohibiting efficient computational analysis such as might have real-time utility in a complex diagnostic setting.

One example of a disease process that is characterized by a constellation of highly variable morphologic features is urothelial carcinoma (UC), the most common type of urinary bladder cancer in Western populations. With numerous described histologic variants characterized by important prognostic and therapeutic implications [11], UC constitutes an epitome of morphologic heterogeneity. The micropapillary variant of UC (MPUC) [12], in particular, demonstrates a wide spectrum of architectural and cytologic features presenting a complex diagnostic problem for expert observers [5], while its natural history of early and extensive invasion often predates aggressive, early surgical management [13].

We have previously shown that the use of ring vectors can address many of the limitations due to intrinsic incorporation of x and y translation, rotational, and its mirror symmetry within the vector itself, which we have implemented in our pattern recognition image analysis program, the “wobble” feature, using a high-throughput pattern recognition image analysis algorithm, Spatially Invariant Vector Quantization (SIVQ) [14]. Briefly, SIVQ is a pattern recognition algorithm that can identify morphologic and architectural features on H&E and special stained digital slides [14, 15]. Initially, one (or a small number of) distinct predicate image candidate feature(s) is identified by the user. This feature in turn is utilized to perform an exhaustive search of the entire surface area of the digital slide, resulting in the generation of a statistical probability heatmap of quality of matches.

SIVQ is unique in that it uses a set of rings instead of a block. A ring is the only geometric structure in two-dimensional space besides a point that has continuous symmetry. Thus, using a series of concentric rings, one can convert this two-dimensional orientation problem into a linear pattern matching of rotational sampling where each ring would match a series of points on a circle and then rotate thru the complete 360 degrees, then move to the next x, y coordinate. Video demonstrations on multiple uses of the SIVQ algorithm can be viewed on our institution’s YouTube channel [16]. Most recently, we have added tools to automatically generate the optimal vector and integrated it into MATLAB™ [17] as well as demonstrated its potential as a tool for user-directed automated tissue microdissection for molecular studies [15].

However, as we have reported before [14], the SIVQ algorithm requires input of multiple key parameters, including ring vector selection, ring diameter, number of sub-rings, and inter-ring wobble. Briefly, a predicate ring vector is created by adjusting the size and number of sub-rings and selecting the feature of interest, capturing the pixels values (described further below). These series of concentric rings are overlaid onto a x, y pixel coordinate, as shown in Fig. 1. This example contains 10 rings where the number and spacing of the dots along the ring corresponds to sampling at twice Nyquist frequency, ensuring complete sampling. The pixels values surrounding each dot in the ring are assessed via bicubic interpolation and compared to the predicate ring vector pixel values. The rings are then systematically rotated and the pixels values are calculated and compared for each point, including its mirror symmetry by flipping the ring.

Anecdotally, we have observed that the pattern recognition performance of the algorithm, in multiple “proof of principle” use cases, is highly variable as a function of ring size, sub-ring number and inter-ring wobble (unpublished observations). Importantly, in no use case have we systematically evaluated the relative contribution of each of these parameters to the performance of SIVQ, leaving open the question of how to implement and optimize its performance for future application.

Thus, proceeding from a diagnostically challenging and clinically critical [5] use case of MPUC, herein we ascertain the relative contributions of SIVQ algorithm parameters to detection of this complex and heterogeneous cancer morphology. Not only do these findings support the argument that SIVQ has the potential to provide a useful adjunctive tool to the surgical pathologist by demonstrating excellent performance when
Fig. 1. Schematic demonstrating a SIVQ ring vector. For demonstration purposes, an representative example ring vector, sized 10 pixels in diameter and with 9 sub-rings, was superimposed on a pixel-level image. This image shows the relationship between the number of points on the ring that are sampled for analysis, as well as the orientation, sizes, and maximal number of subrings. For adequate sampling, Nyquist frequency theory dictates that sampling should occur at twice the maximum frequency, thus requiring sampling at every half pixel. For clarity, in this image the number of sampling points was determined by the size of the outer ring and held constant for the inner rings, though in general SIVQ recalculates the number of sampling points was calculated by Nyquist frequency theory at each ring size.

optimized for detection of MPUC, but more importantly, the use case provides an exemplar for optimization of SIVQ ring parameters that is broadly applicable to future implementations.

2. Methods

2.1. Digital slides

A de-identified hematoxylin and eosin (H&E) stained section of a transurethral resection specimen of MPUC was scanned into a digital format using the Aperio XTi\textsuperscript{TM} whole slide imaging scanner at 40×. The digital slide has been made publicly available on our WSI repository website (www.WSIrepository.org).

Using Microsoft Paint, the surface area of tumor was painted in consensus by pathologists (SCS, JH, SAT, and GU expert LPK) to generate a ground “truth” map (Fig. 2B).

2.2. SIVQ

The use of SIVQ image analysis has been previously described elsewhere by Hipp & Cheng et al. [14, 15]. An additional feature, referred to as “wobble,” was added to the algorithm that allows for rings to independently rotate amongst each other. The degree to which the inner ring rotates amongst its best match of the outer ring directly outside it is defined by the user and ranges from 1–180°.
To improve the speed of the algorithm, alternate numbers of rings up to the total number of sub-rings was used for these analyses. For example, if a ring with a diameter of 11 pixels had 10 subrings, then ring numbers 11, 9, 7, 5, 3, and 1 were analyzed.

2.3. Statistical comparisons of SIVQ ROC and AUC

The utility of the “wobble” parameter was tested directly across various numbers of sub-rings (0, 2, 6, 12) and various angles of “wobble” (1–180 degrees) by analysis of the area under the curve (AUC). Briefly, all areas of malignancy in both fields of interest were identified and painted in consensus by pathologists (SCS, JH, SAT) in consultation with a genitourinary pathology expert (LPK). This image was used as gold standard for calculation of the AUC by comparing distributions of SIVQ scores for pixels within and outside of these curves.

The same vector was selected for each set of experiments with the various parameters being systematically held as constants or variables. The output of the SIVQ analysis resulted in numeric values at each pixel location, which were then converted to a grayscale value of 0–255. The area corresponding to these pixel locations were dilated to the size of the ring...
Fig. 3. ROC curves and AUCs for adjusting the number of sub-rings. For investigation of the contribution of number of sub-rings to overall performance, a ring vector of 13 pixels in diameter was selected. SIVQ analysis was performed by altering the number of sub-rings: 0 (blue), 2 (red), 6 (green), and 12 (black), skipping every other ring. The panel to the left shows incremental improvement in the ROCs as the number of sub-rings is increased. The right panel demonstrates incremental improvements in the AUCs as the number of sub-rings are increased. The maximal AUC was identified with the maximal number of 12 sub-rings and was associated with an AUC of 0.85.

vector that was used for the analysis. These values were compared to a ground truth map to calculate a ROC curve and its AUC.

3. Results

3.1. Histopathologic interpretation

SIVQ allows image analysis at multiple length scales [14]. As screening histologic sections for UC, including its variant MPUC, is undertaken at low power [5], we chose a length-scale of 4× magnification at which to select vectors. Figure 2A shows the representative field of a digital slide of MPUC, stained with standard H&E, which we selected for testing of SIVQ. Small, irregular, and infiltrative nests of malignant urothelial cells show areas of identifiable ‘inverse polarization’ of peripherally oriented nuclei and prominent retraction artifact (Fig. 2A). However, the morphologic heterogeneity of this tumor is demonstrated by areas of multiple nests within the same lacuna, and by epithelial ring forms also indicated, as reported recently. Figure 2B shows the “ground truth” map of UC, painted in consensus by the authors.

3.2. Effect of the number sub-rings (0 vs 2 vs 6 vs 12)

The total number of sub-rings can vary from 0 (minimum) to 1 = Diameter (maximum size in pixels). Conceptually, increasing the number of sub-rings increases the specificity while decreasing the sensitivity because more rings requires more pixel values within the feature to be matched. At 4× magnification, using a standard ring vector size of 13 pixels in diameter, SIVQ analysis was performed with the following number of sub-rings: 0, 2, 6, and 12 (while skipping every other ring). The SIVQ results were then compared to the ground “truth” map and ROC curves were generated (Fig. 3A). Figure 3A shows incremental improvement in performance as one moves from the lowest to the greatest number of sub-rings. The AUC was also calculated demonstrating a maximal performance with 12 sub-rings (0.85, Fig. 3B).

3.3. Large versus small ring vectors

SIVQ works optimally when an archetypal image feature, specific to and pervasively characteristic of the disease process of interest is identified as its exemplar for pattern recognition. For example, a small ring vector captures a cellular feature particular for an individual cell. However, when a small ring vector (3–5 pixels in diameter) is selected for an individual UC cell, the inflammatory infiltrate and stromal cells are also identified due to the similar nuclear/cytoplasmic textures (Supplementary Figure 1). In contrast, selection of a larger vector (11–13 pixels in diameter) identified from within one of the micropapillary nests captures the circular distribution of the tumor nuclei and adjacent cytoplasm (Supplementary Figure 1). Use of this
larger vector identified less background cells (inflammatory and stromal cells) with many more hits specific to nests.

To characterize the effect of ring size and identify the optimal vector, a point within the center of a nest was selected and every odd-numbered ring vector from 3–25 pixels in diameter was collected (while holding the number of sub-rings to a constant, at maximal value) and used to search the field of view (Fig. 2A). The results are displayed in Fig. 4A and B. Proceeding from 3 to 11 pixel ring diameters, we identified an improvement in the ROC curves up to a size of 11 pixels in diameter; subsequently the performance degrades up to 25 pixel diameters. The AUC confirms the best performing vector size was 11 with a value of 0.84 and the corresponding heat map is shown in Fig. 4B.

3.4. Effect of the “wobble” angle

We hypothesized that increasing the angle of “wobble” between the rings would identify more micropapillary nests by increasing the number of potential matches of the heterogeneous cellular/nuclear morphologies and distributions. We analyzed every other ring size from 3 to 23 pixels in diameter; however, we added an extra dimension of “wobble”: no wobble (1), 60 degree, 120 degree, and maximal wobble (180 degrees). Three dimensional ROCs are shown in Fig. 5. The wobble angle had a minor increase on the ROC curves for those large rings between 0 and 120 degrees of “wobble”.

Based on the data above, we further investigated the effect of “wobble” at the lower degrees of freedom. Using a ring vector of 11 pixels in diameter with 10 sub-rings we analyzed the effect of the “wobble” angles of 1, 3, 6, 12, 22, 45, 60, 120, and 180 degrees. Increasing the angle from 1 to 120 degrees improved the AUC by approximately 0.006 (Fig. 6).

3.5. Effect of a benign mimic of urothelial carcinoma on the ROC and AUC

To further assess the ability of ring vectors to separate MPUC from background, we employed the same ring vectors from Fig. 4 using an additional field of view, which contained reactive urothelium (a benign mimic of in situ UC), a slightly different stromal texture and dilated blood vessels which served as a negative control (Fig. 7). This field of view, which contained no neoplasm, was designated as ground truth “negative” and the results were combined together with the results from the ground “truth” of Fig. 1. The ROC curves that contain both fields of views are shown in Fig. 8A.

Fig. 4. Effect of the ring size. A similar point within the nest of cancer cells was selected to generate rings ranging in size from 3–25 pixels in diameter (with the maximal number of sub-rings, sampling every other ring). The panel to the right shows an incremental improvement in the ROC curves as the size of the ring was increased from 3–11 pixels in diameter, followed by an incremental degradation in the ROCs from 11–25. The panel to the left demonstrates the corresponding AUCs with vector size of 11 demonstrating the greatest AUC, 0.84.
Fig. 5. Effect of ring size and wobble. Similar to Fig. 4, we explored the effect of ring size (from 3 to 23 pixels in diameter); however, we added an extra dimension of inter-ring “wobble”: no wobble (1), 60 degree, 120 degree, and maximal wobble (180 degrees). A ROC was plotted across both dimensions of these parameters, resulting in a three dimensional curve, shown for clarity, here, with four rotational views (A–D). The bottom right figure shows best that the wobble angle had a minor increase on the ROC curves for those large rings between 0 and 120 degrees of “wobble”.

Fig. 6. Effect of wobble. Using a ring vector of 11 pixels in diameter with 10 sub-rings, numerous “wobble” angles of 1, 3, 6, 12, 22, 45, 60, 120, and 180 degrees were analyzed. Increasing the angle from 1 to 120 degrees had little improvement on the AUC ($\sim 0.006$). was also an improvement in the maximal AUC at ring vector size 11 to 0.88.

4. Discussion

To identify a particular morphologic or architectural feature for SIVQ, there are critical ring vector parameters one has to optimize, including the ring diameter, number of sub-rings (which rotate along with its outer ring), and inter-ring wobble. Thus, proceeding from a diagnostically challenging and clinically critical [5] use case of MPUC, herein we ascertain the relative contributions of each of these parameters towards overall detection optimization. We believe these findings provide an exemplar by which this algorithm and future histology-oriented pattern matching tools may be optimized, validated, and, it is our hope, subsequently implemented broadly as adjunctive tools in other appropriate microscopic classification settings.
Fig. 7. Analysis of a 4× field of reactive urothelium for comparison. To further assess the ability of ring vectors to separate MPUC from background, we employed the same ring vectors from Fig. 4 and used an additional field showing reactive, hyperplastic urothelium. This additional field of view of reactive urothelium, a benign mimicry of in situ UC, shows a slightly different stromal texture and dilated blood vessels and served as a negative control.

Fig. 8. SIVQ analysis of MPUC and reactive urothelium. Similar to the analysis in Fig. 4, an additional field of view (Fig. 7) was added to the analysis and new ROC curves and AUCs were calculated. There was a very similar pattern with an improvement in performance from 3–11 followed by a degradation from 11–25 (left panel). The right panel shows an improvement in the maximal AUC, 0.88.

MPUC presents a complex diagnostic problem for expert observers [5], and its natural history of early invasion of the muscularis propria of the bladder with frequent lymphovascular invasion and metastatic nodal colonization often predicates aggressive, early surgical management [12, 13]. Thus, the creation of computer-aided detection tools to assist in its identification would be of material benefit to patients and exceed any additional time or costs associated with a digital workflow.

To summarize the findings reported above, ring vectors function like a “safe combination”, each one
rotating autonomously for the correct “match”. This results in highly specific, homogeneous feature identi-
fications. However, to relax this feature and create an increased degree of inde-
pendent rotation. Adjusting the diameter of the rings to capture the optimal image predicate to account for the size and variability of the feature of interest, such as cellular or nuclear size, is of critical importance. Once the optimal ring size is selected, increasing the number of sub-rings greatly improves the specificity and sen-
sitivity. While in this use case adjusting the “wobble” angle has little impact on the results, while making the analysis more computationally expensive, we do note a small improvement in performance and cannot exclude that in other use cases wobble might result in a greater effect.

Since this is the first image analysis report regard-
ning MPUC, and since a computer aided detection of MPUC might have a significant impact on patient care, we have made all of our primary data reported here publicly available on our WSI reposi-
tory site (www.WSIrepository.org). Here, others will have access to review the digital slide, the TIFF images used for the analysis, and the ground truth maps. It is tempting to speculate that a computer aided detec-
tion solution for MPUC is an achievable goal, and we expect that it would entail a multi-algorithm approach. Such efforts are ongoing in our group, and we hope other investigators can build upon this foundation.

In conclusion, not only do the findings reported here extend the case that SIVQ has the potential to provide a useful adjunctive tool for the surgical pathologist by demonstrating excellent performance when optimized for detection of MPUC, but more importantly, this chal-
 lenging use case provides an exemplar for optimization of SIVQ ring parameters in a manner that is broadly applicable to other malignancies.

Supplementary Material

Supplementary Figure 1. Vector images for the various ring sizes. At ring size greater than 15 pixels in diameter, one can observe both the micropapillary nest and adjacent stroma in the chosen vector predicate image. A vector size of 3 or 7 pixels includes only a subset of the nest. A vector size of 11 demonstrates a generic feature motif that encompasses multiple nuclei within the nest of cancer cells.

References

[1] E.E. Walk, Improving the power of diagnostics in the era of targeted therapy and personalized healthcare, *Curr Opin Drug Discov Dev* 12(2), 226–234.
[2] C.R. King, Patterns of prostate cancer biopsy grading: Trends and clinical implications, *Int J Cancer* 99(6) (2000), 305–311.
[3] R. Montironi, R. Macracchi, M. Scarpelli, A. Lopez-Beltran, G. Fellegara and F. Algaba, Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: Contemporary approach, current clinical significance and sources of pathology discrepancies, *BJU Int* 95(8) (2005), 1146–1152.
[4] A. Tabesh, M. Terevskov, H.Y. Pang, V.P. Kumar, D. Ver-
 bel, A. Kotsianti, et al., Multifluorescence prostate cancer diagnosis and Gleason grading of histological images, *IEEE Trans Med Imaging* 26(10) (2007), 1366–1378.
[5] A.R. Sangoi, A.B. Beck, M.B. Amin, L. Cheng, J.I. Epstein, D.E. Hansel, et al., Interobserver reproducibility in the diag-
osis of invasive micropapillary carcinoma of the urinary tract among urologic pathologists, *Am J Surg Pathol* 34(9) 1367–1376.
[6] S. Doyle, M. Feldman, J. Tomaszewski and A. Madabhushi, A boosted bayesian multi-resolution classifier for prostate cancer detection from digitized needle biopsies, *IEEE Trans Biomed Eng*.
[7] M.N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Ruigrok and B. Yener, Histopathological image analysis: A review, *IEEE Rev Biomed Eng* 2 (2009), 147–171.
[8] A. Madabhushi, Digital pathology image analysis: Opportu-
nities and challenges, *Imaging in Medicine, [Editorial]* 1(1) (2009), 4.
[9] A. Madabhushi, S. Agner, A. Basavanhally, S. Doyle and G. Lee, Computer-aided prognosis: Predicting patient and disease
outcome via quantitative fusion of multi-scale, multi-modal data, *Comput Med Imaging Graph*.

[10] J.P. Monaco, J.E. Tomaszewski, M.D. Feldman, J. Hagemann, M. Moradi, P. Mosuavi, et al., High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models, *Med Image Anal* 14 (6) (2010), 617–629.

[11] M.B. Amin, Histological variants of urothelial carcinoma: Diagnostic, therapeutic and prognostic implications, *Mod Pathol* 22 (Suppl 2) (2009), S96–S118.

[12] M.B. Amin, J.Y. Ro, T. et-Sharkawy, K.M. Lee, P. Troncoso, E.G. Silva, et al., Micropapillary variant of transitional cell carcinoma of the urinary bladder, Histologic pattern resembling ovarian papillary serous carcinoma, *Am J Surg Pathol* 18 (12) (1994), 1224–1232.

[13] I.A. Ghoneim, R. Mincinovic, A.J. Stephenson, J.A. Garcia, M.C. Gong, S.C. Campbell, et al., Neoadjuvant systemic therapy or early cystectomy? Single-center analysis of outcomes after therapy for patients with clinically localized micropapillary urothelial carcinoma of the bladder, *Urology* 77 (4) (2011), 867–870.

[14] J.D. Hipp, J.Y. Cheng, M. Toner, R. Tompkins and U. Balis, Spatially invariant vector quantization: A pattern matching algorithm for multiple classes of image subject matter—including pathology, *J Pathol Inform* 2 (2011), 13.

[15] J. Hipp, J. Cheng, J.C. Hanson, W. Yan, P. Taylor, N. Hu, et al., SIVQ-aided laser capture microdissection: A tool for high-throughput expression profiling, *Journal of Pathology Informatics* (Original Article) 2 (1) (2011), 19.

[16] U.J. Balis and J.D. Hipp, Algorithm will help search for cancer cells, *UM Health System* (2011) Available from: <http://www.youtube.com/watch?v=O5IJXqkb9SQ&feature=relmfu>.

[17] J.D. Hipp, J. Cheng, J. Monaco, D.R. Lucas, A. Madabushi and U.J. Balis, Automated vector selection of SIVQ and parallel computing integration MATLAB. Innovations supporting large-scale and high-throughput image analysis studies, *Journal of Pathology Informatics*, In press.
