Research Article

A vocabulary for the identification and delineation of teratoma tissue components in hematoxylin and eosin-stained samples

Ramamurthy Bhagavatula, Michael T. McCann[1], Matthew Fickus[2], Carlos A. Castro[3], John A. Ozolek[4], Jelena Kovacevic[5,1]

Massachusetts Institute of Technology Lincoln Laboratory, Boston, MA, [1]Department of Biomedical Engineering, Center for Bioimage Informatics, [5]Department of Electrical and Computer Engineering, Carnegie Mellon University, [3]Department of Obstetrics and Gynecology, Magee-Womens Research Institute and Foundation of the University of Pittsburgh, [4]Department of Pathology, Children's Hospital of Pittsburgh of the University of Pittsburgh, Pittsburgh, PA, [2]Department of Mathematics and Statistics, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, USA

E-mail: *Michael T. McCann - mtmccann@cmu.edu
*Corresponding author

Abstract

We propose a methodology for the design of features mimicking the visual cues used by pathologists when identifying tissues in hematoxylin and eosin (H&E)-stained samples. **Background:** H&E staining is the gold standard in clinical histology; it is cheap and universally used, producing a vast number of histopathological samples. While pathologists accurately and consistently identify tissues and their pathologies, it is a time-consuming and expensive task, establishing the need for automated algorithms for improved throughput and robustness. **Methods:** We use an iterative feedback process to design a histopathology vocabulary (HV), a concise set of features that mimic the visual cues used by pathologists, e.g. "cytoplasm color" or "nucleus density." These features are based in histology and understood by both pathologists and engineers. We compare our HV to several generic texture-feature sets in a pixel-level classification algorithm. **Results:** Results on delineating and identifying tissues in teratoma tumor samples validate our expert knowledge-based approach. **Conclusions:** The HV can be an effective tool for identifying and delineating teratoma components from images of H&E-stained tissue samples.

Key words: Automated histology, classification, segmentation

Access this article online

Website:
www.jpathinformatics.org

DOI: 10.4103/2153-3539.135606

Quick Response Code:

INTRODUCTION

Histology is vital to medicine and research as it enables quantitative and qualitative analysis of tissue samples, stained and visualized through microscopes; the most routine and cost-effective of these stains is H&E (see Table 1 for a list of all abbreviations used in this paper, and see Figure 1 for example H&E images). In stem-cell research, an automated tool to identify and delineate tissue types would help enhance our understanding of tissue development in teratoma tumors derived from embryonic stem cells. Current methodologies require visual inspection of each representative tissue section from each block of processed tissue, identification of all the tissue types present, semi-quantification of each type, and reconstruction of teratomas from these estimations, all tedious tasks producing approximate results.

Challenges and Guiding Principle

Histopathological data from teratomas present numerous challenges: The samples consist of different tissues at various developmental stages, tissue boundaries may not be well-defined, and staining techniques can vary from sample to sample and from lab to lab. There is great variability within a single tissue [Figure 1a and b] as well as similarity between tissues [Figure 1c and d]. Moreover, unlike in natural-image segmentation (e.g. identifying trees, grass, or sky), where non-experts can determine the right answer, only trained pathologists can distinguish many of the tissues in teratomas.

In light of these challenges, we adopt the use of expert knowledge as a guiding principle: We distill the visual cues used by pathologists into a histopathology vocabulary (HV), a set of features understood by both pathologists and engineers. We use this vocabulary to create an automatic system for tissue identification and delineation in teratoma tumors. At the same time, the vocabulary framework can be used to automate other histology tasks.

Related and Previous Work

There have been many recent advances in the development of computer-aided diagnostic (CAD) tools for histology.[1] Many of these tools address classification tasks, e.g. distinguishing normal versus diseased presentations of a single type of tissue, or tissues of multiple types.[2,3] These works focus on automated detection or grading of cancer in areas such as the breast, prostate, cervix, brain, and lymphatic system. Other CAD systems focus on the delineation of component biological materials within an image, such as identifying glands in prostate histology. Still others address the normalization or preprocessing of histological images.[4,5] These methods are gaining importance as different pathology labs generate images of the same tissue that are visually different beyond pathological and morphological differences due to variation in staining and tissue-processing methods.

Our previous work in this area focused on demonstrating the feasibility of identifying single tissues in images of H&E-stained samples.[6,7] The algorithm performed a multiresolution decomposition on the input image, then classified in each subspace and combined the subspace decisions through voting. We used Haralick texture features and a set of nucleus-based features specifically designed for this application and reported a best average accuracy of approximately 88%, supporting the feasibility of single-tissue classification.

Gaps to Fill

Given these initial results on single-tissue classification, here we address multi-tissue classification. We move toward a principled method for designing features based on the expert knowledge of pathologists. Parts of the current work were presented at ISBI 2010,[8] including Figure 1. Though we validate our current method on the same dataset as in our previous work, the method are not the same, as the vocabulary in the current work is significantly expanded.

HISTOPATHOLOGY VOCABULARY

A generic classification system consists of feature extraction and classification. Based on our guiding principle, we want to develop a HV as an intuitive feature set that mimics the pathologist's visual cues. It is constructed through an iterative feedback process between pathologists and engineers, resulting in a computational vocabulary to describe tissues in the same way we use vocabulary in a spoken language to describe various concepts.

We stress that the methodology is general and can be used in any application domain; in fact, it has been used to create an active colitis vocabulary,[9] as well as in helping diagnose middle-ear infection by creating an otitis media vocabulary.[10] Some individual features are novel, others are not; we emphasize that while individual features are not necessarily novel, the vocabulary is. The aim is to construct a concise, meaningful set of features that well represents the data in a specific application domain.

Table 1: Summary of abbreviations used in this paper

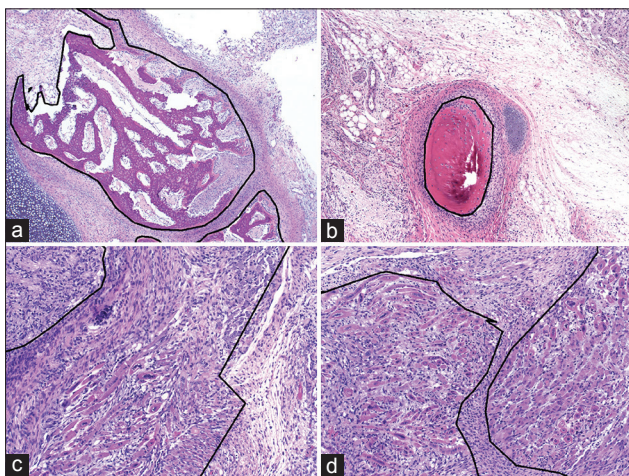| Abbreviation	Definition
H&E	Hematoxylin and eosin
HV	Histopathology vocabulary
CAD	Computer-aided diagnosis
LBP	Local binary patterns
NN	Neural network

Figure 1: Challenges in images of hematoxylin and eosin-stained samples. (a and b) Intraclass variability. (c and d) Interclass similarity

Vocabulary Creation Methodology

To find a set of features emulating the visual cues used by pathologists, we propose the following methodology.

1. Initial set of experts' descriptions: The pathologist provides initial descriptions of those characteristics best describing each tissue in an H&E-stained preparation, e.g. "bone is dark red and has few nuclei, and cartilage is blue-gray." For this work, the pathologist described H&E-stained teratoma tissue types including neuroepithelial, neuroectoderm, immature and mature neuroglial, ganglion, squamous epithelium, skin, immature and mature skeletal muscle, smooth muscle, connective, cartilage, bone, adipose, immature kidney, gastrointestinal, respiratory, pancreatic, liver, and necrotic. Pathologists J.O. and C.C. chose this list of tissues as being representative of what they saw in their teratoma research, but we stress that others could be added, including partially differentiated germ cell components.

2. Identification of features: From this set, the engineer distills the key implementable terms, or features, creating a computational vocabulary. For example, from the description above, the engineer identifies color and nucleus density as the key features distinguishing bone and cartilage.

3. Description translations: The engineer then translates the pathologist's descriptions into statements using the vocabulary. For example, bone might be described as "background color = dark red; nucleus density = low".

4. Verification of translated descriptions: The pathologist then receives the translated descriptions and tries to identify the tissue being described, emulating the overall classification system with translated descriptions as features and the pathologist as the classifier. This step has two possible outcomes

- If the pathologist is unable to recognize a tissue based on its description in terms of the vocabulary, or if any of the terms in that vocabulary are unclear to the pathologist, then the engineer must refine the vocabulary or add additional features to it
- If the pathologist is able to identify each tissue based on its description in terms of the vocabulary, then the discriminative power of the vocabulary is validated.

By using this approach, we arrive at vocabulary of features that are understood by both engineers and pathologists and are sufficient to distinguish the tissues at hand.

Vocabulary

Using the proposed methodology, we have created an initial HV consisting of the eight most discriminative features, listed in Table 2. Figure 2 shows examples of some of the features. These features come from following the above procedure with the pathologists C.C. and J.O. Other pathologists may list different visual cues, resulting in different vocabularies. These could be merged with ours to create a larger vocabulary; consensus among the experts is not required because the classifier will learn which

Table 2: Histopathology vocabulary

Symbol	Name
f_b	Background color
f_c	Cytoplasm color
f_e	Lumen density
f_{nc}	Nucleus color
f_{nd}	Nucleus density
f_{ns}	Nucleus shape
f_{no}	Nucleus orientation
f_{nr}	Nucleus organization

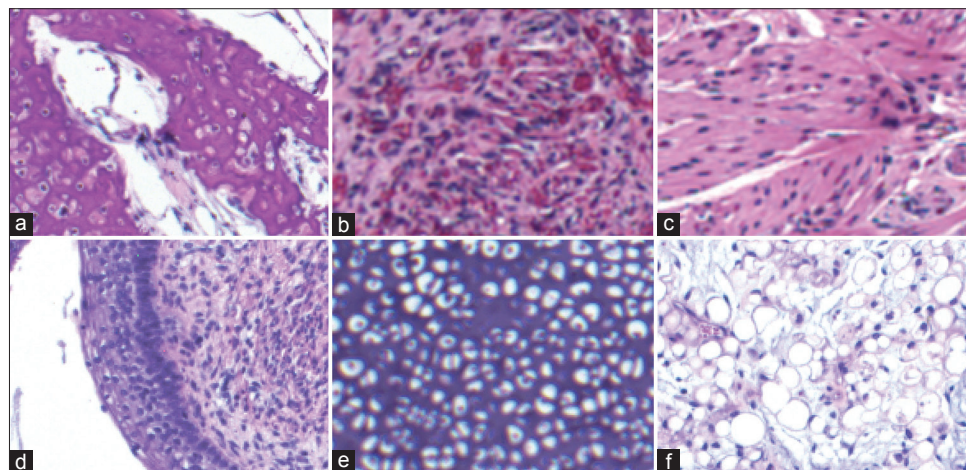These eight terms represent both visual cues used by pathologists and computational features understood by engineers

Figure 2: The appearance of the histopathology vocabulary in different tissues. (a and b) Background color for bone, Is bright pink, while it is dark pink for pseudovascular tissue, which contains many small capillary-type vessels. (c and d) Nucleus orientation and organization for smooth muscle. Is unstructured, while it is structured for skin. (e and f) Clear areas (lumen) for cartilage. Are less dense than for fat

terms are most useful during training. Here, we outline the implementation of these features. Figure 3 shows an overview of the system. For more mathematical detail, refer to the appendix.

Segmentation

We aim to compute each feature in a small neighborhood around each pixel in the image. Because the HV refers to four specific biological components (background, cytoplasm, lumen, and nuclei), we first identify each pixel in the input image as belonging to one of these components. To identify both nuclei and lumen, we perform an initial segmentation based on a stain-separation method.[4] Following this initial segmentation, the nonnucleus and nonlumen regions are further refined into cytoplasm and background regions through a simple thresholding of region homogeneity with respect to the stain contributions, the intuition being that cytoplasm is more consistent in composition compared to background. The result of the segmentation is four sets of pixels locations corresponding to background pixels, cytoplasm pixels, lumen pixels, and nucleus pixels.

Features

For each HV feature, we compute a value for each pixel in the image using the pixels in a small region around it. Therefore, an important parameter of the method is the size of this region. In this work, we use circular regions with radii of 4, 8, 16, 32, and 64 pixels. We now describe each HV feature.

Background Color

Background color characterizes the collective color of the tissue components aside from nuclei, lumen, and cytoplasm. As these components represent a large portion of almost any given tissue, their appearance is highly indicative of a tissue's identity. To implement this feature at a given pixel, we calculate the average color of the nearby pixels that have been segmented as background.

Cytoplasm Color

Cytoplasm color is calculated similarly, as a local average of the color of pixels segmented as cytoplasm.

Lumen Density

Lumen density characterizes those regions where the H&E stain was not absorbed, such as lumen (inside space of a cellular component) or lack of tissue (e.g. at the edge of a sample). We calculate lumen density as the fraction of the local area that is covered by lumen pixels.

Nucleus Color

Nucleus color is typically blue/purple as nuclei primarily absorb hematoxylin. To extract nucleus color at pixel, we use the local average of the color of the nucleus pixels.

Nucleus Density

Nucleus density is highly reflective of not only tissue identity but also specific pathologies. To calculate it, we count the number of nucleus pixels in a local region, then divide by the number of tissue (i.e. nonlumen) pixels in the region. The effect is to calculate the local fraction of the tissue that is covered by nuclei.

Nucleus Shape

Nucleus shape can be a strong indicator of tissue identity as it often correlates strongly with nucleus orientation and organization. To quantify the local nucleus shape, we calculate the eccentricity of each segmented nucleus. We then calculate the local entropy of these eccentricity values. Thus, if all the nuclei around a given pixel are the same shape, then the value for nucleus shape is zero; on the other hand, if the nuclei around a given pixel have a wide variety of shapes, then the value for nucleus shape is one.

Nucleus Orientation

Nucleus orientation characterizes their orientation with respect to each other, partially quantifying the overall organization of nuclei within a tissue. For example, nuclei can be parallel to each other, perpendicular, or anywhere in between. Similarly to the nucleus-shape feature, we first compute the orientation of each nucleus. We then characterize the nucleus orientation at pixel as the local entropy of these orientations. If all the nuclei around a given pixel are oriented in the same direction, then the value for nucleus orientation is zero (very structured
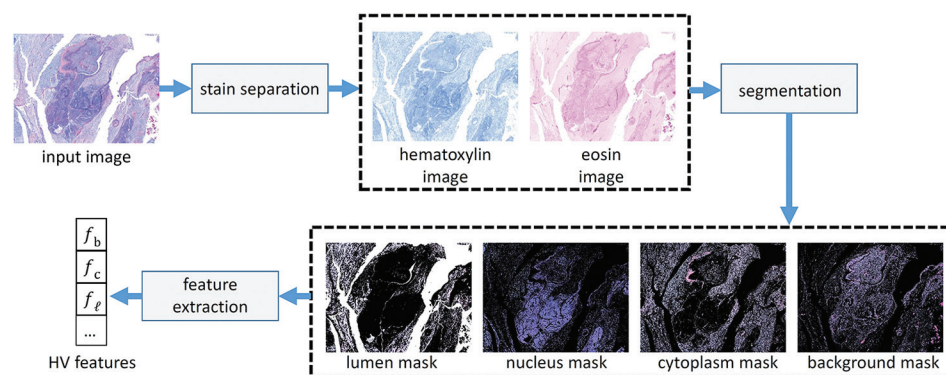

Figure 3: Illustrative block diagram of the histopathology vocabulary feature extraction process

organization); on the other hand, if the nuclei around a given pixel are oriented randomly, then it is one (unstructured organization).

Nucleus Organization

Nucleus organization further quantifies the organization of nuclei in a tissue. For example, in skin tissue, nuclei are arranged in smoothly changing collinear layers of common orientation. To capture this behavior, we find the local center of nucleus mass for each point in the image, and the distance to that point. The local average of this distance at a pixel gives an indication of the uniformity of the arrangement of nuclei in that region, forming one feature. We compute a second feature at each pixel by computing the orientations of the vectors to the local centers of mass and calculating their local entropy with the same approach as above. These features give a low-level description of the size and orientation of the macro shape formed by the nuclei. While more complex organizations may require a more refined approach, the simplicity of this approach provides robustness.

EXPERIMENTS

We now describe experiments comparing the HV to commonly-used generic texture-feature sets on the task of identifying and delineating the tissues in images of teratoma tumors.

Generic Texture-Feature Sets

In many histopathology classification tasks,[11-13] generic texture information was chosen as the primary descriptor of class identity, as it is both intuitively and empirically a vital descriptor. We therefore compare our HV to commonly-used texture descriptors: Gabor filter banks, local binary patterns (LBP), and textons.

Gabor Filters

Gabor Filters have been used in image analysis for a long time, including in histopathology applications.[2,3,11,12] They are designed to respond to textures at specific orientations and scales. We use Gabor filters at eight evenly spaced orientations (0, 22.5, 45,…, 157.5°) over five scales (4, 8, 16, 32, 64 pixels) for a total of 40 filters. Each of these 40 filters is applied to each red, green, and blue channel of the image. The final features are formed by computing the local mean, standard deviation, and mode of each filter's response, yielding a total of $40 \times 3 \times 3 = 360$ Gabor features per pixel.

Local Binary Patterns

Local binary patterns[14] are another powerful method for texture characterization; LBP methods have been used in many applications, including automated histology.[13] They describe a local texture using simple spatial operators and encoding methods.[14,15] For our comparison, we use an LBP operator with a circular neighborhood of radius one pixel with eight evenly spaced (angularly) points.

The resulting binary vectors are aggregated into local histograms; these histograms are used as feature vectors.

Textons

Textons[16,17] create a prototype library of textures with which to both classify and generate textures. They have been used in a host of applications, including biomedical ones.[18-20] We use an implementation consisting of isotropic Gaussians and Laplacians of Gaussians in addition to oriented Gabor filters as before, for a total of 50 filters. Following the filter phase, we apply a K-means clustering using a Euclidean distance metric to learn a fixed number (we chose 100) of textons. We then compute local histograms of texton occurrence; these local histograms become the feature vectors.

Pixel-Level Classification

We use pixel-level classification (making a decision about the identity of each pixel individually) to identify and delineate tissues in multi-tissue images.

A two-layer neural network (NN) is used as the classifier. The input layer consists of as many nodes as the length of the feature vectors with hyperbolic tangent sigmoid activation functions. The output layer has as many nodes as the number of tissues/classes of interest with linear activation functions. Our choice of network design is motivated by our past success with it in a variety of applications including histopathology.

We expect this approach to yield relatively smooth delineations. However, junctures between tissues will lead to regions of confusion. To address such issues, we apply a local refinement in the form of a weighted local voting to the initial set of labels. In short, the final label at each pixel is calculated as the most frequent label in a small area around that pixel. Again, we leave the details of this calculation for the appendix.

Dataset

The dataset consists of 36 images of H&E-stained samples of teratomas, derived, serially sectioned, and imaged at ×4 magnification.[6] The images have a size of 1600 ×1200 pixels and a pixel size of 1.6125 μm. The total number of teratomas represented in the dataset is 10; 15 tissues appear in these images although the number of images in which each tissue appears varies greatly. We choose to work with this magnification as it is the most useful from the pathologist's point of view as well as because it allows for the greatest number of multi-tissue images. All images were of sufficient clarity for the pathologists C.C. and J.O. to delineate component tissues of the teratomas.

The ground truth (definitive diagnosis) for each image is obtained by expert pathologist hand segmentation and labeling of tissues (C.C. and J.O.). Regions of uncertain identity, due to either lack of information or artifacts, are ignored and not used for evaluation.

Experimental Setup

We use bone (B), cartilage (C), immature neuroglial (I), neuroepithelial (N), and fat (F) tissues to make our feature comparison. These tissues were selected because they commonly occur in a variety of presentations in our data, allowing the creation of a large and varied training set of these tissues.

For each tissue, we randomly select 50% of the images containing that tissue and add them to our training images, while sequestering the rest for testing; we ensure that no image appears in both the training and testing sets. We then choose 1% of available training pixels to compose our training set. We weight this selection so pixels near the edges of training regions are chosen less frequently, as they can confuse the classifier.

Finally, due to the unequal number of training pixels for each tissue, we further sample the training pixels so that each class contains the same number of training pixels. This sampling removes the variation in training-set size that can bias the classifier during training.

After training, we apply the classifier to the images in the testing set. For each pixel in each image, it decides whether the pixel is in a region of bone, cartilage, immature neuroglial, neuroepithelial, or fat tissue. After classification, we compare the classifier's decisions with the true labels given by the pathologist. Each pixel that the classifier and the pathologist agree on is counted as correct. Each pixel that the pathologist labels as one of the aforementioned five tissues, but that the classifier labels as a different one is counted as incorrect. Any pixel that the pathologist labels as a tissue other than the five aforementioned tissues is not scored. We calculate the pixel-level accuracy (number of correct pixels divided by the number of correct plus incorrect pixels) separately for each tissue. We repeat the process with a new random selection of training images 10 times and report the average result. The pixel-level accuracy for each tissue is an estimate of the sensitivity of the classifier to that tissue; e.g. an accuracy of 70% for bone means that 70% of pixels that should be labeled bone are labeled so by the classifier.

RESULTS AND DISCUSSION

Table 3 compares the accuracy of the classifier using the HV (first column) to that of the classifiers using the Gabor (second column), LBP (third column) and texton (fourth column) feature sets. For examples of images automatically labeled by our method, see Figure 4. The HV outperforms the generic features for all tissues and scales. The Gabor features have their best performance on cartilage; this coincides with our perception that texture is an important descriptor of cartilage [Figure 2e]. Similarly, the relatively poor performance on bone tissue reflects a lack of color and architectural information that is more indicative of this tissue. The LBP features perform better than the Gabor features. We again see that while texture description is vital, it is not sufficient. In particular, color information and organization of components are needed. Note that the LBP features perform better at smaller scales; this fits our intuition on the power of LBPs in quantifying micro-textures, e.g. cartilage. Textons perform somewhere between LBP and Gabor features, and, as the others, are inferior to HV. Note that here, commonalities between the texton and Gabor features become apparent; the improved performance at larger scales can be attributed to increased robustness to noise. Moreover, the superior performance of textons when compared with Gabor features lends credence to the notion that the increased conciseness of the texton representation improves discrimination.

The consistently superior performance of the HV over the generic feature sets validates the vocabulary approach to this problem. The HV is powerful because it encodes valuable pieces of domain knowledge, including the segmentation of the tissue into background, cartilage, lumen, and nuclei, and focus on the number and shape of nuclei, information that is unlikely to be extracted be a generic feature set. One limitation of the approach is that tissues that are similar with respect to the defined vocabulary will often be confused during classification. In this case, it may be necessary to add additional terms to the vocabulary to distinguish these

Table 3: Comparison of average accuracies when segmenting bone, cartilage, immature neuroglial, neuroepithelial, and fat with features computed scales of 4, 8, 16, 32, and 64 pixels

Scale (pixels)	Accuracy (HV/Gabor/LBP/textons) (%)				
	B	**C**	**I**	**N**	**F**
4	72/35/62/40	72/51/65/60	66/44/50/54	73/33/60/39	65/44/62/51
8	72/38/61/41	71/48/66/52	71/43/52/48	72/37/61/43	67/40/64/45
16	73/36/65/41	73/50/65/53	70/46/56/50	70/35/62/40	69/40/62/42
32	75/38/58/42	76/53/60/56	71/48/50/52	70/38/59/42	71/42/60/45
64	70/39/57/45	74/55/60/58	69/50/51/54	71/40/54/45	70/43/55/43

Each entry shows the accuracy for our proposed HV versus the comparison methods of Gabor, LBP, and texton feature sets on teratoma data. The best accuracy for each tissue is bolded. For each scale and tissue, the HV features perform the best. HV: Histopathology vocabulary, LBP: Local binary patterns, B: Bone, C: Cartilage, I: Immature neuroglial, N: Neuroepithelial, F: Fat

tissues. For example a feature describing fiber length might be necessary to distinguish between mature and immature skeletal muscle. Also note that the current work validates the HV on a test set including only five

tissue types. In real tumors, these tissues may or may not appear, and many other tissues may be present. The classifier does not depend on these five tissues being present (e.g. if there is no bone in a given image, the
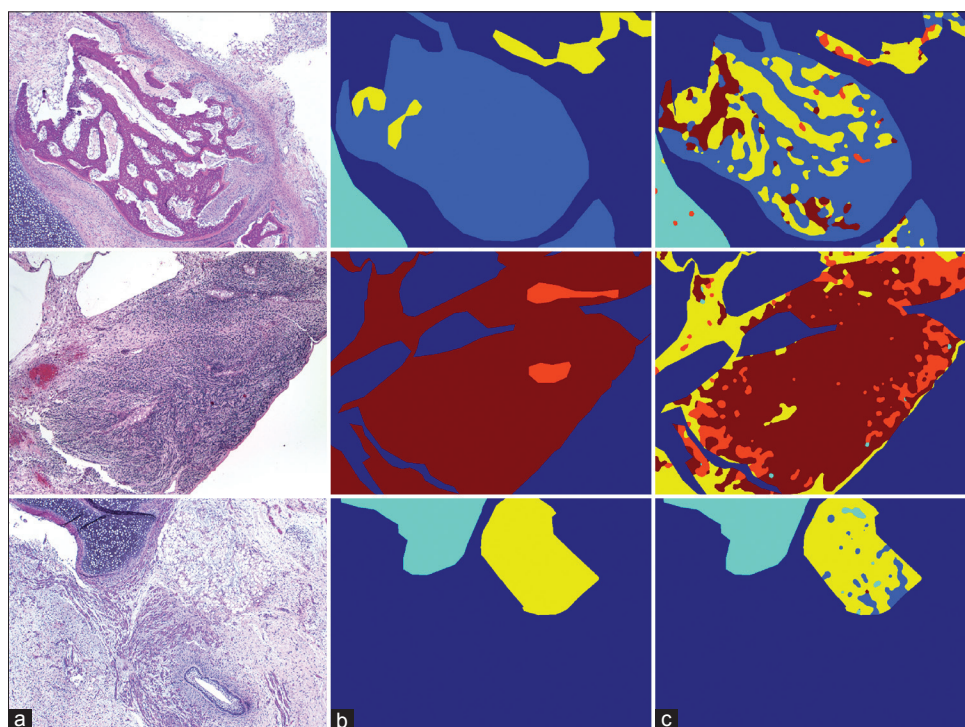

Figure 4: Example identification and delineation of bone (light blue), cartilage (cyan), immature neuroglial tissue (maroon), fat (yellow), and neuroepithelial tissue (orange), with background (dark blue) regions ignored. (a) Original image. (b) Ground truth. (c) Our result.
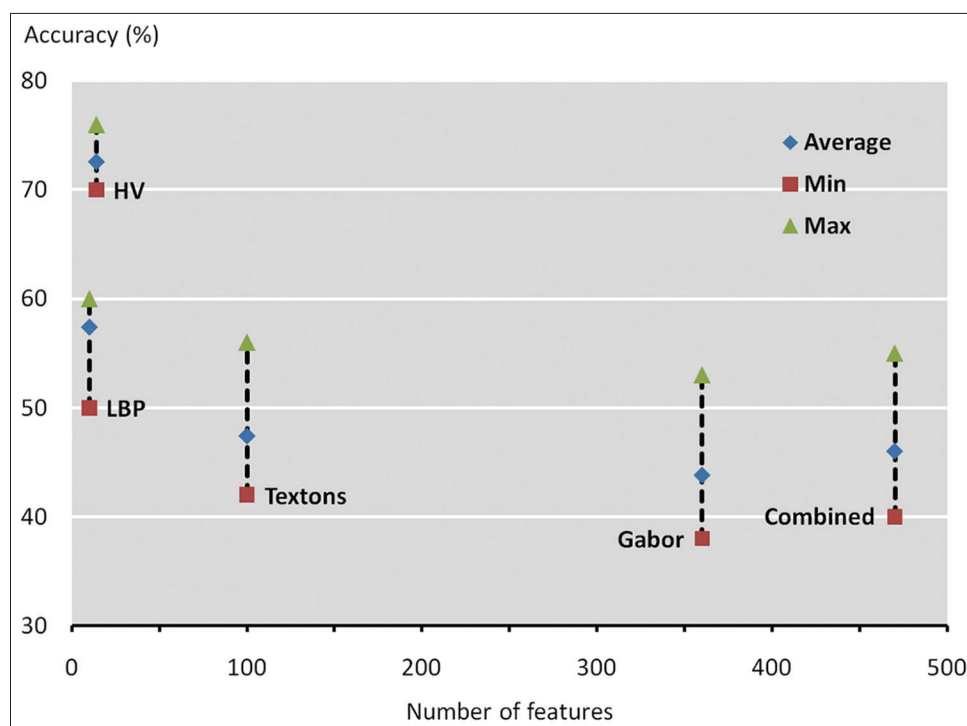

Figure 5: Comparison of results with the combined set consisting of Gabor, local binary patterns and texton features. The blue square represents the overall average accuracy across all five tissues while the red and green squares represent the minimum and maximum individual tissue accuracies respectively. All results are given for the best-performing scale of 32 pixels

classifier should not mark any pixels as bone), but it does need training images for each new tissue type it aims to identify.

Figure 5 shows a comparison of results shown in Table 3 with an additional, combined set, consisting of Gabor, LBP and texton features. This table shows another benefit of the HV: It is succinct while still providing high accuracy.

CONCLUSIONS AND FUTURE WORK

In this work, we have proposed a systematic methodology for the creation of an HV, a concise and effective set of features understood by both pathologists and engineers. We compared the performance of the HV to that of Gabor filters, LBPs, and textons; the HV outperformed all of them, validating the use of domain-specific features that aim to extract expert knowledge. Current and future work includes applying the vocabulary concept to other medical image-processing applications, as well as refining our implementation of the HV features and expanding our experiments to include more tissue types.

ACKNOWLEDGEMENTS

REFERENCES

1. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: A review. IEEE Rev Biomed Eng 2009;2:147-71.
2. Zhao D, Chen Y, Correa H. Statistical categorization of human histological images. Proc IEEE Int Conf Image Process 2005;3:628-31.
3. Zhao D, Chen Y, Correa H. Automated classification of human histological images, a multiple-instance learning approach. Proceedings of IEEE Life Sci Syst Appl Workshop; 2006. p. 1-2.
4. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, *et al.* A method for normalizing histology slides for quantitative analysis. Proceeding IEEE International Symposium Biomedical Imaging, Chicago, IL; 2009. p. 1107-10.
5. Mete M, Topaloglu U. Statistical comparison of color model-classifier pairs in hematoxylin and eosin stained histological images. Proceeding IEEE Symposium Computational Intelligence Bioinformatics Computational Biology; 2009. p. 284-91.
6. Chebira A, Ozolek JA, Castro CA, Jenkinson WG, Gore M, Bhagavatula R, *et al.* Multiresolution identification of germ layer components in teratomas derived from human and nonhuman primate embryonic stem cells. Proceedings IEEE International Symposium Biomedical Imaging, Paris, France; 2008. p. 979-82.
7. Chebira A, Barbotin Y, Jackson C, Merryman T, Srinivasa G, Murphy RF, *et al.* A multiresolution approach to automated classification of protein subcellular location images. BMC Bioinformatics 2007;8:210.
8. Bhagavatula R, Fickus M, Kelly W, Guo C, Ozolek JA, Castro CA, *et al.* Automatic identification and delineation of germ layer components in H&E stained images of teratomas derived from human and nonhuman primate embryonic stem cells. Proc IEEE Int Symp Biomed Imaging 2010;2010:1041-14.
9. McCann MT, Bhagavatula R, Fickus MC, Ozolek JA, Kovačević J. Automated colitis detection from endoscopic biopsies as a tissue screening tool in diagnostic pathology. Proceedings IEEE International Conference Image Processing, Orlando, FL; 2012. p. 2809-12.
10. Kuruvilla A, Li J, Hennings Yeomans P, Quelhas P, Shaikh N, Hoberman A, *et al.* Otitis media vocabulary and grammar. Proceedings IEEE International Conference Image Processing, Orlando, FL; 2012. p. 2845-8.
11. Doyle S, Agner S, Madabhushi A, Feldman M, Tomaszewski J. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. Proceedings of IEEE International Symposium Biomedical Imaging, Paris, France; 2008. p. 496-9.
12. Huang PW, Lee CH, Lin PL. Support vector classification for pathological prostate images based on texture features of multi-categories. Proceedings of IEEE International Conference System, Man Cybernetics; 2009. p. 912-6.
13. Qureshi H, Sertel O, Rajpoot N, Wilson R, Gurcan M. Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. Proceedings of International Conference Medical Image Computer Computer-Assisted Intervention; 2008. p. 196-204.
14. Ojala T, Pietikäinen M, Harwood D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. Proc IEEE Int Conf Pattern Recogn 1994;1:582-5.
15. Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002;24:971-87.
16. Julesz B. Textons, the elements of texture perception, and their interactions. Nature 1981;290:91-7.
17. Leung T, Malik J. Representing and recognizing the visual appearance of materials using three-dimensional textons. Int J Comput Vis 2001;43:29-44.
18. Bosch A, Munoz X, Oliver A, Marti J. Modeling and classifying breast tissue density in mammograms. Proc IEEE Int Conf Comput Vis Pattern Recogn 2006;2:1552-8.
19. Khurd P, Bahlmann C, Maday P, Kamen A, Gibbs-Strauss S, Genega EM, *et al.* Computer-aided gleason grading of prostate cancer histopathological images using texton forests. Proc IEEE Int Symp Biomed Imaging 2010;2010:636-69.
20. Chatzistergos S, Stoitsis J, Papaevangelou A, Zografos G, Nikita KS. Parenchymal breast density estimation with the use of statistical characteristics and textons. Proceedings of International Conference Information Technology and Applications Biomedice; 2010. p. 1-4.

APPENDIX

In this appendix, we present the mathematical expressions for each term in the HV.

Definitions

Let $I : Z^{M \times N} \to Z_{256}^3$ be an RGB image, and let A be a set of pixel locations, $(m,n) \in Z^{M \times N}$, in the image. Define the restriction operator, 1_A, such that

$$1_A I_{m,n} = \begin{cases} I_{m,n}, & \text{for} (m,n) \in A; \\ 0, & \text{otherwise.} \end{cases}$$

We form a binary mask of the set A as

$$\mathcal{A} = 1_A U \, ,$$

where U is a size $M \times N$ image of all 1s.

After the segmentation step, we have four sets of pixels locations, B, C, L, N, and their corresponding mutually exclusive binary masks, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{L}$, $\mathcal{N}$, representing background, cytoplasm, lumen, and nuclei, respectively.

Features

To compute the HV, we use these masks along with a local averaging filter, $w_{m,n}$. In this work, we choose $w_{m,n}$ to be a flat circular averaging filter and experiment with radii of $s = 2^k$, $k = 2, \ldots, 6$ pixels.

Background Color

We calculate the average local background color at pixel (m,n) as

$$f_b = \frac{(w \times 1_B I)_{m,n}}{(w \times \mathcal{B})_{m,n}}$$

Cytoplasm Color

Cytoplasm color is calculated similarly,

$$f_c = \frac{(w \times 1_C I)_{m,n}}{(w \times \mathcal{C})_{m,n}}$$

Lumen Density

We calculate lumen density at pixel (m,n) as

$$f_\ell = (w \times \mathcal{L})_{m,n}$$

Which gives the fraction of the local area that is covered by lumen pixels.

Nucleus Color

To extract nucleus color at pixel (m,n), we use the same approach as with the other color features,

$$f_{nc} = \frac{(w \times 1_N I)_{m,n}}{(w \times \mathcal{N})_{m,n}}$$

Nucleus Density

Nucleus density is calculated at pixel (m,n) as

$$f_{nd} = \frac{(w \times \mathcal{N})_{m,n}}{(w \times [\mathcal{B} + \mathcal{C} + \mathcal{N}])_{m,n}}$$

In other words, f_{nd} counts the number of nucleus pixels in a local region, then divides by the number of tissue (i.e. nonlumen) pixels in the region. The effect is to calculate the local fraction of the tissue that is covered by nuclei.

Nucleus Shape

To quantify the local nucleus shape, we compute the centroid and eccentricity of each connected region (blob) in $\mathcal{N}$ and create a new binary mask $\mathcal{N}_c$, corresponding to pixels at centroid locations; each blob should correspond to an individual nucleus. We quantize the eccentricities to K values, $\{e_0, e_1, \ldots, e_{K-1}\}$ and identify the pixels of $\mathcal{N}_c$ corresponding to each eccentricity, forming a new mask for each value, $\mathcal{E}_0, \mathcal{E}_1, \ldots, \mathcal{E}_{K-1}$. We then characterize the nucleus shape at pixel (m,n) as the normalized local entropy of these eccentricities,

$$f_{ns} = -\frac{\sum_{i=0}^{K-1} P_{m,n}(e_i) \log P_{m,n}(e_i)}{\log K}$$

Where

$$P_{m,n}(e_i) = \frac{(w \times \mathcal{E}_i)_{m,n}}{(w \times \mathcal{N}_c)_{m,n}}$$

At each pixel, $P_{m,n}$ is the local probability mass function for nucleus eccentricity and f_{ns} calculates its entropy, then normalizes it to lie between 0 and 1.

Nucleus Orientation

Similarly to the nucleus-shape feature, we first compute the orientation of each blob in $\mathcal{N}$. We quantize the orientations to K directions, $\{\theta_0, \theta_1, \ldots \theta_{K-1}\}$ and create binary masks for each value, $\Theta_0, \Theta_1, \ldots, \Theta_{K-}$. We then characterize the nucleus orientation at pixel (m,n) as the normalized local entropy of these orientations,

$$f_{no} = -\frac{\sum_{i=0}^{K-1} P_{m,n}(\theta_i) \log P_{m,n}(\theta_i)}{\log K}$$

Where

$$P_{m,n}(\theta_i) = \frac{(w \times \Theta_i)_{m,n}}{(w \times \mathcal{N}_c)_{m,n}}$$

Nucleus Organization

We define a new local averaging filter, h and create first-order local moment filters, $h^{1,0}$ and $h^{1,0}$ from it via

$$h_{m,n}^{p,q} = h_{m,n} \, m^p n^q$$

These filters allow us to calculate the coordinates of the local center of nucleus mass for each point in the image,

$$\bar{x}_{m,n} = \frac{(h^{1,0} \times \mathcal{N})_{m,n}}{(h \times \mathcal{N})_{m,n}}, \bar{y}_{m,n} = \frac{(h^{0,1} \times \mathcal{N})_{m,n}}{(h \times \mathcal{N})_{m,n}}$$

and the distance to that point,

$$d_{m,n} = \sqrt{\bar{x}_{m,n}^2 + \bar{y}_{m,n}^2}$$

The local average of this distance at a pixel (m,n) gives an indication of the uniformity of the arrangement of nuclei in that region, forming one feature

$$f_{\text{nr}_a} = (w \times d)_{m,n}$$

We compute a second feature at pixel (m,n) by computing the orientations of the vectors to the local centers of mass and calculating their local entropy with the same approach as above,

$$f_{\text{nr}_b} = -\frac{\sum_{i=0}^{K-1} P_{m,n}(\theta_i) log P_{m,n}(\theta_i)}{log K}$$

where

$$P_{m,n}(\theta_i) = (w \times \Theta_i)_{m,n}$$

Label Refinement

For the NN classifier, label masks $\mathcal{K}_0, \mathcal{K}_1, ..., \mathcal{K}_{L-1}$ are accompanied by corresponding confidences $c_{m,n} \in (0,1]$. Given this set of initial labels, their confidences, and a finite weighting function $w_{m,n}$ that sums to 1, we refine the label at the point (x,y) as follows:

$$\ell = \underset{i=0,...,L-1}{\text{argmax}} \sum_{m,n} w_{m,n} \ \mathcal{K}_{i,x+m,y+n} \ c_{x+m,y+n}$$

Confidences $c_{m,n}$ as well as the weighting enforced by $w_{m,n}$, reflect our intuition that more confident labels and certain neighbors are likely to be the same label as a given pixel.