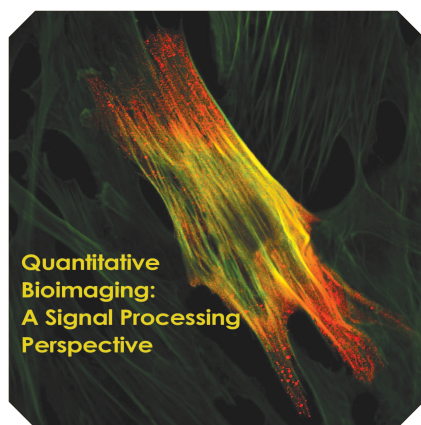


Automated Histology Analysis

[Opportunities for signal processing]

Histology is the microscopic inspection of plant or animal tissue. It is a critical component in diagnostic medicine and a tool for studying the pathogenesis and biology of processes such as cancer and embryogenesis. Tissue processing for histology has become increasingly automated, drastically increasing the speed at which histology labs can produce tissue slides for viewing. Another trend is the digitization of these slides, allowing them to be viewed on a computer rather than through a microscope. Despite these changes, much of the routine analysis of tissue sections remains a painstaking, manual task that can only be completed by highly trained pathologists at a high cost per hour. There is, therefore, a niche for image analysis methods that can automate some aspects of this analysis. These methods could also automate tasks that are prohibitively time-consuming for humans, e.g., discovering new disease markers from hundreds of whole-slide images (WSIs) or precisely quantifying tissues within a tumor.

In this article, we aim to acquaint the signal processing researcher with histology and review the current approaches to the fascinating and important signal processing problems associated with histology image analysis. Throughout, we focus on slides stained with the ubiquitous hematoxylin and eosin (H&E) stain and imaged with brightfield microscopy.



Quantitative
Bioimaging:
A Signal Processing
Perspective

© ISTOCK PHOTO.COM/BEANOS

HISTOLOGY: THE PATHOLOGIST'S VIEW

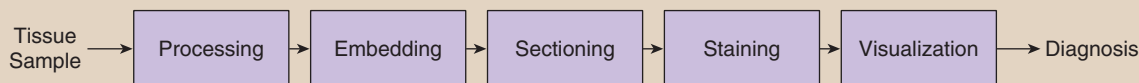
The main goal of the surgical pathologist in a diagnostic practice is to examine tissue and render a correct diagnosis that will ultimately translate to a therapeutic intervention for the patient. The therapeutic response may range from no action, in the case of a diagnosis of normal or unremarkable, to close follow-up, local excision, medical treatment only (benign diagnoses), or radical chemo-

therapy and/or surgery (malignant diagnoses).

Figure 1 presents the pipeline from tissue processing to diagnosis from the perspective of the diagnostic pathologist. Understanding this process will give the reader an appreciation of how images are derived from tissue and the associated sources of variability, noise, and artifacts; this information is critical to designing automated image analysis systems. This portion of the article is based on standard histology texts [1], [2], as well as years of experience in the field of pathology.

TISSUE COLLECTION

The clinical histology process begins when the treating physician, after assessing the patient by history, physical examination, and/or radiographic and laboratory studies, determines that treatment can proceed no further without histology confirmation. The treating physician then must obtain enough good-quality tissue to obtain a diagnosis. There are several possible approaches to tissue collection, including fine-needle aspiration, needle biopsy, excisional biopsy, or excision of the lesion in its entirety. The sensitivity (likelihood of



[FIG1] A block diagram of the histology process.

getting the correct diagnosis) and specificity (likelihood of not getting the incorrect diagnosis) increase from fine-needle aspiration to excision of the entire lesion. This is because the larger biopsies preserve more cellular context and allow the pathologist to examine multiple slides from different areas of the sample.

After biopsy, the pathologist evaluates the tissue on the macroscopic scale, measuring it and recording a description of its color and characteristics. For larger tissues (e.g., tumor resections or colon resections), the tissue must be trimmed to fit into the tissue cassettes (approximately $10 \times 10 \times 3$ mm) that will contain it for the subsequent processing steps (Figure 2).

PROCESSING

The next step in the diagnostic pipeline is tissue processing, which involves chemically and physically stabilizing the tissue. The tissue is first immersed into a fixative solution that is used to stop cells from breaking down and prevent microorganism growth. In general, tissue is fixed for a few hours (small biopsies) to about 24 hours (large biopsies). Fixation is critical because poorly fixed tissue leads to poor tissue sectioning and poor microscopic morphology.

After fixation, the tissue is physically stabilized by one of several methods (freeze drying, microwave, chemical) with the end goal of preserving the cellular morphology. The most commonly employed method involves the use of alcohols and xylene and is automated in most laboratories: First, the tissue is dehydrated, which clears the water and aqueous fixative from it. Next, the tissue is cleared of the dehydrating agent, leaving the tissue ready for paraffin infiltration. Finally, the paraffin warms in the processor until it is liquefied, infiltrates the tissue under vacuum, and then cools so that the tissue becomes firm. This process takes approximately nine hours and, in many laboratories, is run overnight. One of the end results of processing tissue in this manner is that the tissue section is slightly smaller than the original fresh or fixed tissue prior to processing.

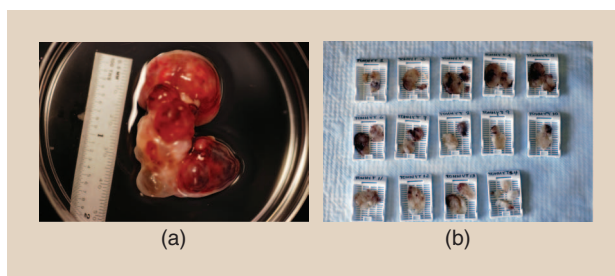
EMBEDDING

After processing, the tissue is embedded in a block of support material, as shown in Figure 3. To achieve this, the tissue is placed on the bottom of a mold and paraffin is poured over it. The original tissue cassette is placed over the mold and then onto a cooling plate to solidify the paraffin. The result is called a *tissue block*, tissue that is impregnated with and surrounded by hardened paraffin. Because the tissue will be sliced parallel to the cassette, orientation of the tissue during embedding is key; see Figure 4 for an example.

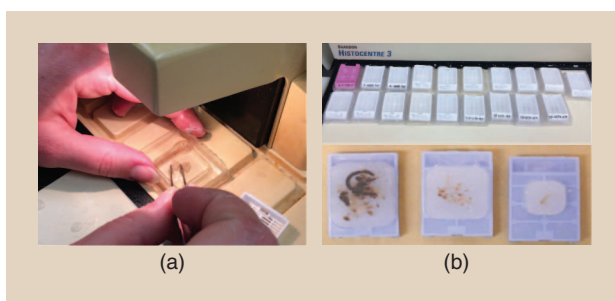
SECTIONING

Sectioning is cutting thin slices of tissue that are mounted on microscope slides (Figure 5). This is achieved with a tool called a

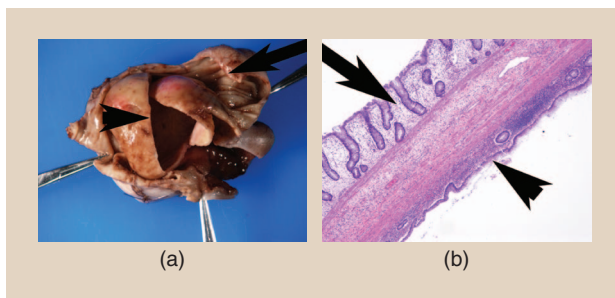
microtome, which operates like a deli slicer. This step can be manual, semiautomated, or automated. For example, the manual rotary microtome advances the block by a set amount with each turn of the wheel and thus produces a ribbon of tissue. Most tissue sections for diagnostic purposes are cut at $3\text{--}4\text{-}\mu\text{m}$ thickness. For certain applications (e.g., silver staining of kidney biopsies) thinner sections are necessary; these are more difficult to obtain without damaging the tissue. Thicker sections tend to make staining dark and obscure nuclear detail.



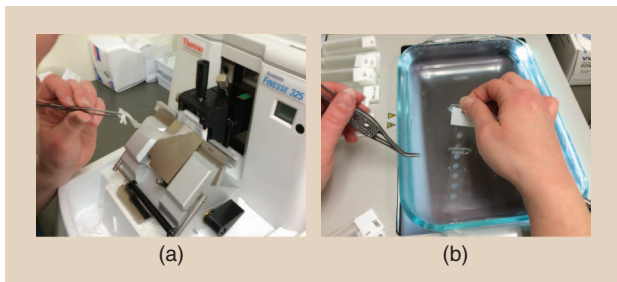
[FIG2] (a) A gross specimen is cut into smaller pieces and placed into (b) cassettes for further processing.



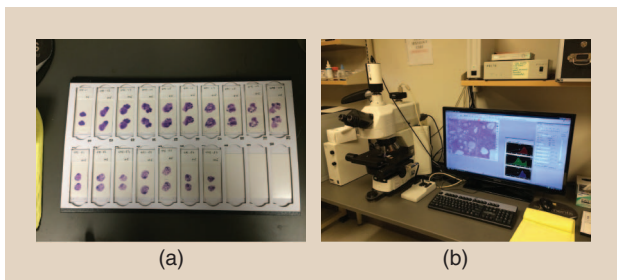
[FIG3] (a) Paraffinized tissue is oriented and embedded into a block of paraffin. (b) Tissue blocks after embedding.



[FIG4] The importance of proper orientation during embedding. (a) A gross specimen shows an opened enteric duplication cyst (arrowhead) intimately joined to the bowel (arrow). (b) A histology image of the cyst with normal bowel (arrow) abutting cyst wall (arrowhead) can only be observed when slicing occurs perpendicular to the wall.



[FIG5] (a) During sectioning, the tissue block is sliced into 3–4- μm sections, which remain connected in a ribbon. (b) Collecting a ribbon from the water bath using a glass slide.



[FIG6] (a) Histology slides after staining. (b) The visualization setup including a microscope with attached camera and a desktop computer.

These thinly cut sections are floated out onto a water bath heated to about 10 °C below the melting point of the paraffin. This allows dispersion of any wrinkles generated by the microtome blade at sectioning. Once the sections are cut and floated, they are placed onto 25 × 75 × 1 mm glass microscope slides.

STAINING

At this point, the tissue slices are nearly invisible under a light microscope, so they must be stained to create contrast. Most staining procedures in the laboratory, aside from immunohistochemical (IHC) ones (antibody based), use chemicals or dyes that will bind or have affinity for certain components of the cells and extra-

cellular components. The chemical properties of these dyes produce the visual appearance that is seen under the microscope.

The most widely used stains for both diagnostic and research histology are H&E. Hematoxylin stains nucleic acids and appears blue/purple, while eosin stains proteins and appears pink/red when visualized under a brightfield microscope. So for most tissues, cell nuclei are blue, while cytoplasm can vary from clear to red to purple depending on its constituents. The reason that H&E staining of tissue has persisted for decades as the primary tissue stain in diagnostic and research pathology is that these stains attach themselves to almost every cellular component, allowing for the visualization of whole cells and all tissue components. Another reason is that these stains provide excellent contrast between cellular constituents by having chemical properties that produce colors at opposite ends of the visual spectrum. These color perceptions are helpful in diagnosis, though not entirely necessary, since, even in grayscale, distinctions between and within tissues can be made and diagnoses rendered.

VISUALIZATION

Once stained, the slides must be visualized (Figure 6). A growing trend in pathology is to digitize slides so that pathologists can make diagnoses based solely on the digital image [3], [4]. The advent and refinement of whole-slide scanners have made rapid scanning and high-resolution WSIs commonplace. These systems are now offered by many companies and offer spatial resolutions using the 40 × objective of approximately 0.23–0.25 $\mu\text{m}/\text{pixel}$. Storing digital images would be an attractive alternative to storing glass slides, since glass slides take considerably more space, can be damaged or lost, and fade over time, but currently institutions must keep their glass slides and tissue blocks for at least ten years.

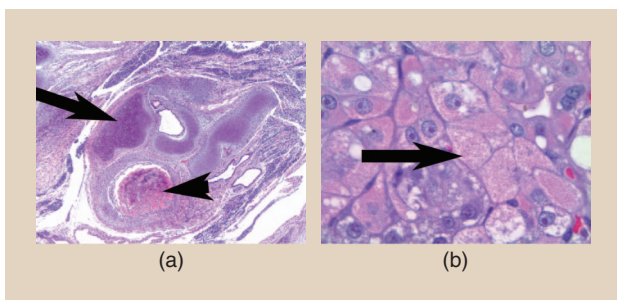
Despite these advantages, in most medical centers, the pathologists still rely on visualization of the slide through a microscope. According to pathologists, a microscope offers faster panning, faster focusing, and an intangible sense of being closer to the tissue. This last advantage is more difficult to explain, but the sense is that the eye can capture greater detail through the microscope than from a digital image on a monitor. Even so, studies have shown no appreciable difference in diagnoses rendered by pathologists using digitized images compared to diagnoses rendered using a microscope [5]–[8].

ANALYSIS

The goals of the pathologist in the clinical domain versus the research domain can be very different; we discuss each separately.

CLINICAL PRACTICE

In the clinic, the goal of the pathologist is to render accurate and timely diagnoses. For a given slide, they analyze a wide variety of characteristics including tissue architecture, cellular color and texture (Figure 7), and cellular/nuclear morphometry (Figure 8) to produce diagnoses. Computer algorithms may automate some of these tasks. For example, in cytopathology, automated methods for screening Pap smear slides limit the number of cytotechnologists needed to complete these screenings. In the future, we could



[FIG7] Some examples of important visual cues in histology. (a) In this teratoma section, color makes distinguishing cartilage (blue/gray, arrow) from bone (pink/red, arrowhead) easy, even at low magnification. (b) In this liver section of a child with mitochondrial disorder, texture is important. Under high magnification, cell borders are accentuated and cytoplasm shows tiny red granules representing abnormal mitochondria (arrow).

hope to automate or semiautomate histology screenings, e.g., of the gastrointestinal tract. Such automation could greatly reduce health-care costs and potentially provide pathologists more time for challenging cases and important research.

RESEARCH

In the research domain, pathologists may aim to quantify differences between histology samples in terms of a variety of parameters including cellular/nuclear morphometry, amount of stroma (the connective tissue cells that support the function of cells around them) present, types of tissue present, etc. Although some basic analyses can be readily performed using available image analysis software (e.g., Photoshop, ImageJ, MetaMorph), most quantification in pathology remains semiquantitative: staining intensity may be rated as low, moderate, or strong; the amount of a certain cell type may be visually estimated as 0–25%, 25–50%, or >50% of the total population; and morphometric descriptions of cells are limited to semantic descriptions such as larger, thickened, pleomorphic, or cellular.

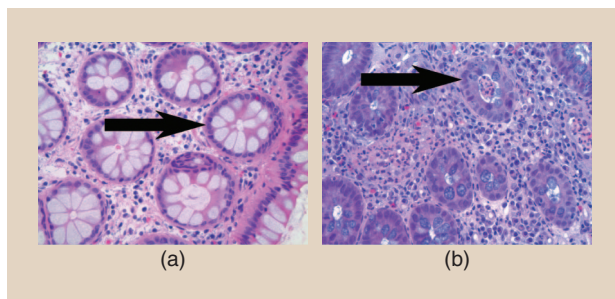
For some studies, this type of analysis is not accurate enough, because many biologically or clinically relevant features cannot be easily captured and processed by the human visual system. For example, given two tumors, how would a pathologist support the claim that the average nucleus size is different between them? Similarly, how can pathologists quantify, by eye, complex patterns such as chromatin distribution? Signal processing solutions to these quantification challenges would be extremely useful technology for researchers in almost any area of investigation that analyzes and quantifies observations from tissue specimens.

SOURCES OF VARIABILITY

There are three main sources of variability in a histology-based diagnosis: biological variability, interobserver variability, and technical variability. Biological variability encompasses the normal variability among people and the myriad of pathological processes that can affect any tissue group. Due to biological variability, slides generated from the same tissue in different patients can look different. Interobserver variability contends that two pathologists can look at the same tissue and render different interpretations. Finally, technical variability is the variability in a slide's appearance due to how it was prepared. We focus here on the details of technical variability.

In the best case, the slide that is generated from the above tissue processing pipeline shows tissue that is properly oriented, sectioned, stained, and coverslipped. Unfortunately, each of these steps can introduce variability into the final product, and differences in protocol between labs can greatly alter the appearance of even biologically similar tissue samples. Some variables such as fixation, specimen orientation in the block, and microtome sectioning are heavily dependent on human skill, and even though tissue processing, staining, and coverslipping are largely automated, they still depend on human monitoring, machine maintenance, and solution preparation.

For automated analysis systems, each source of variability presents a significant obstacle. Tissue that is poorly fixed will not



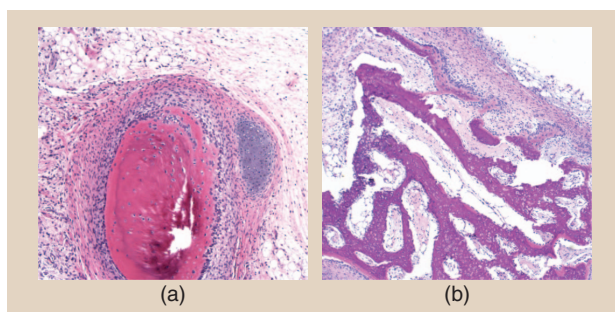
[FIG8] Some examples of the importance of nuclear shape and distribution in colon screening. (a) In normal colon tissue, nuclei are regular in size and distribution around the colon gland (arrow). (b) In colitis, nuclei become much less regular (arrow).

cut well and has a blurry appearance, removing important cues such as edges. Tissue that has been dried out will be shrunken and have poor morphology and stain contrast, similar to poorly fixed tissue. Sectioning artifacts are some of the most commonly encountered artifacts and can produce folds in the tissue, chatter artifacts from a dull blade (seen as alternating light and dark regions), or missing pieces of the tissue.

Staining is a critical source of variability because it produces the color and contrast on the slide (Figure 9). Many different formulations of H&E exist, each producing a slightly different appearance. The stain can also be applied in different manners. In progressive staining, the sample remains in the hematoxylin solution for a specified amount of time to render appropriate staining. In regressive staining, the sample remains in the hematoxylin long enough to overstain, and is then destained back to the desired contrast with an acid alcohol solution. Furthermore, both H&E solutions can have their staining capabilities altered by prolonged storage, contamination by other reagents or water, precipitation, and changes in pH. Staining artifacts can include light staining with either hematoxylin or eosin, precipitated hematoxylin (seen as blue chunks under the microscope), or lack of staining with either hematoxylin, eosin, or both. That these artifacts are well understood by pathologists is another reason why H&E remains so popular.

AUTOMATED HISTOLOGY IMAGE ANALYSIS

In this section, we survey the current methods and signal processing challenges in automated histology image analysis. We aim to



[FIG9] An example of staining variability. Both (a) and (b) show images made of bone, but the color of the bone varies from (a) pink to (b) purple due to staining variability. Such color variations present a challenge for automated analysis systems.

cover a broad range of signal processing topics, but we restrict the review to slides stained with the common and inexpensive H&E stain and imaged with brightfield microscopy. This is because H&E images are prevalent in clinical and research settings and because this focus allows us to survey a more coherent group of methods; we argue that the approaches to a single task, e.g., nucleus detection, vary considerably across staining and imaging modalities. This section is organized around a generalized block diagram of a hypothetical histology analysis workflow, shown in Figure 10. For each block in the diagram, we describe the signal processing challenge and review the current approaches in the literature.

We are aware of three previous reviews of automated histology analysis. The first, [9], deals with analysis of histology images including H&E images as well as fluorescence and multispectral images. It covers preprocessing; segmentation of glands, nuclei, and other subcellular components; feature extraction; dimensionality reduction; and classification. The second, [10], tackles the broader field that the authors term *computational pathology*, which includes histology as well as cytology analysis. It discusses issues of data and ground truth collection including variation among experts and publicly available data sets and describes automated analysis primarily from a statistical pattern recognition viewpoint. The third, [4], discusses histology WSI informatics, including quality control during image acquisition, feature extraction, region of interest (ROI) detection, and visualization.

While our survey does not include every paper on automated histology, we aim to give the reader a sense of what has been tried for the various histology analysis tasks we present. We have given priority to recent journal papers except where necessary. Also, note that we do not focus on results of or comparison between methods. This is because, at this stage, automated histology research is diffuse: most methods are tailored to private data sets and there is no consensus on what quantitative metrics should be reported. Moving toward shared data sets and metrics will be a critical step forward in the field.

PREPROCESSING

Histology images exhibit the same types of artifacts and noise as any digital microscopy image, as well as some novel ones introduced by tissue processing. These can affect analysis unless removed with appropriate preprocessing. A good overview of noise and artifacts in digital microscopy and methods to correct them

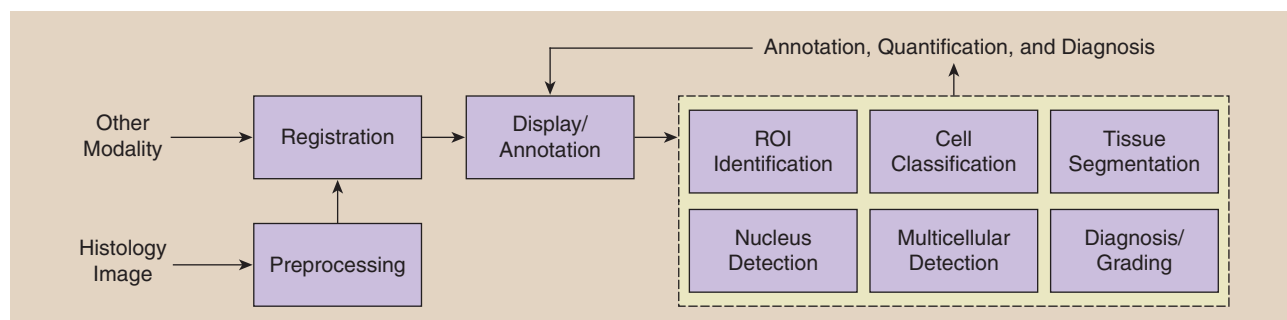
are discussed in [11]; in this section we focus on the issues specific to H&E images of tissue processed as described above.

STAIN NORMALIZATION

As discussed in the section “Sources of Variability,” staining variations affect the appearance of histology images; such variations are problematic for automated analysis because color is a critical feature in histology. Stain normalization is the process of taking two H&E images that have staining variation between them and removing this variation. Doing so has been shown to improve histology image segmentation [12], [13]. One approach to stain normalization is to use color normalization techniques from photography such as histogram equalization; e.g., in [12], the rank statistics of the input image are scaled to match those of the reference image separately in each color channel. The method has been successfully used in other automated histology work, including [14]. Such methods are especially suited for cases where the images to be normalized show approximately the same tissue, e.g., in the case of serial sections such as in [15].

An approach more specifically tailored for H&E image normalization is to first separate the image into H-only and E-only images (sometimes called *color deconvolution*), then normalize these images separately and recombine. This approach better handles cases where, e.g., the hematoxylin stain is too intense but the eosin stain is too weak. When the stain colors are known, the color deconvolution method in [16] solves the separation. When they are not known, the problem becomes more challenging. The approach taken in [13], [17], and [18] is to estimate the stain colors and deconvolve as before. In [17], the stain colors are estimated using the fact that all stained pixels will lie on a wedge in color space; this is a special case of the nonnegative matrix factorization problem studied elsewhere. Reference [18] uses expectation-maximization to find clusters in chromaticity space that correspond to stain colors, and [13] finds pixels stained with only one stain via supervised classification and uses their mean color to estimate the stain colors. A different approach is to estimate the single-stain images directly; e.g., [19] searches for an H-only image that removes most of the contrast from the red channel of the input, based on the assumption that most this contrast comes from the nuclei, which are stained only by hematoxylin.

These separation methods all produce qualitatively fair results, but it is unclear which comes closest to the correct separation.



[FIG10] A block diagram of a generic automated histology analysis workflow. Most current works address only one or a few of these blocks.

We have recently released a stain separation benchmark data set based on chemical destaining [20]; our comparison indicates that the method of [17] is superior to that of [19], but we did not evaluate any of the more recent approaches.

TISSUE DEFORMATION

Because the tissue slices are very thin, they can fold over on themselves during processing, creating a tissue area that is doubly thick. The authors of [21] proposed detecting these folds by their high color saturation, and the authors of [22] devised a method for selecting a good saturation threshold for this detection, but, to our knowledge, no work addresses correcting tissue folds digitally. In the same vein, differences in the water content of different tissues can cause them to pull apart when the tissue is dehydrated during processing; the result is white cracks that are not biologically meaningful. Again, we are not aware of any work that addresses correcting these artifacts.

STITCHING

High-resolution WSIs are often acquired by imaging several strips or tiles separately, which then must be stitched together to create the final image. Because the offsets between the strips or tiles are known, adequate stitching results are often provided by the microscope/scanner software or can easily be achieved with simple compositing techniques. A more challenging stitching problem occurs when a large gross specimen must be sectioned and imaged as several pieces. In [23], the authors describe a GUI that allows stitching of high-resolution images of such tissue fragments via hand-selected control points and linear transformation. We are not aware of a work that addresses histology image stitching with automatic control point selection or explores a richer set of transforms.

REGISTRATION

There are a variety of H&E histology image registration tasks, depending on the image modalities involved. In this section, we review three common registration tasks involving H&E images.

H&E TO H&E

Given serial sections of a tissue sample, each stained with H&E, one aim is to register them to create a tissue volume. This three-dimensional (3-D) reconstruction can more fully show the extent of a pathological process or show relationships of tissue types to one another. This is a difficult registration task: each slice undergoes nonrigid deformation during processing and may exhibit cracking or folding artifacts. The standard approach is to register in a coarse to fine manner as in [15], where the authors register images of serial tissue slices to explore the 3-D shape of cervical tumor fronts using a series of three registrations: 1) a rigid registration using a frequency domain method, 2) a polynomial registration using control points automatically selected with correlation matching on small patches, and 3) a registration consisting of unconstrained local displacements regularized by local curvature. An approach more tailored to histology is to use specific anatomical landmarks; e.g., [24] registers based on blood vessels.

H&E TO ANOTHER STAIN

One may also want to register images of two adjacent tissue slices that use different stains; e.g., H&E to IHC stains. This task is difficult because, by design, the stains will give contrast to different structures. One way to overcome this obstacle is with the selection of an appropriate pixel-wise similarity measure, e.g., the authors of [25] register H&E images to a variety of IHC images using mutual information and the elastix software package [26]. Another approach is given in [27], where distinctive landmarks such as blood vessels are segmented and used to compute the registration.

H&E TO MRI

Compared to histology, magnetic resonance imaging (MRI) has low contrast and resolution, but has the advantage of being noninvasive; registering histology images to magnetic resonance (MR) images could help train radiologists, provide better noninvasive diagnoses, and enable the development of MRI-based computer-aided design tools [28]. This task is difficult because MR images are 3-D while histology images are two dimensional, MR images are lower resolution than histology images, and because the contrast in MR and histology images is generated in different ways. When considering MR images collected in vivo, registration is even more difficult because surgical extraction and histology processing can greatly deform the tissue. In [29], histology images are registered to in vivo MR images via two intermediates, the block face photo and the ex vivo MR image, with the idea that the deformations between these intermediate stages are less drastic and therefore easier to estimate. Each step of the registration is completed using mutual information as the metric and thin plate splines to deform the image with control points initialized by hand. In [28], histology images are registered to MR slices to create a histology volume, then the MR volume is registered in 3-D to the histology volume and resliced. The process is iterated until convergence.

In each of these scenarios, comparison between methods is difficult because no ground truth exists. This problem is intensified in the first two scenarios because qualitatively good registrations may actually remove true differences between the adjacent tissue slices being registered.

DISPLAY AND ANNOTATION

A fundamental problem in the display of histology images is that they can be huge, easily several gigabytes for uncompressed WSIs. As a result, commercial slide scanners often save images in a proprietary format for which the manufacturers provides free viewing software, e.g., Aperio ImageScope. For the researcher interested in reading the images themselves, one solution is Openslide [30], a C library that aims to allow slides from any vendor to be opened, manipulated, or converted to other formats. One format that OpenSlide can convert to and is a natural fit for large histology images is the Deep Zoom (DZI) format, which creates from a large image a tiled image pyramid, allowing real-time viewing of images of arbitrary size, even streaming over the Internet. One example of this approach can be found in [31], which involves displaying WSIs from the Cancer Genome Atlas data set [32] on the Internet using the similar Zoomify format. The system in [33] is designed

specifically for viewing histology images and avoids the time-consuming calculation of image pyramids by creating the current view in real time from image tiles.

Beyond simply viewing large histology images, researchers may want to collect expert annotations for them for the purpose of training and testing their analysis algorithms. In the simplest case, the expert pathologist may label an entire image, e.g., as *normal* or *cancer*. For more fine-grained labeling, Aperio ImageScope, Sedeen Viewer, and Cytomine [34] provide freehand annotation tools for large images. The next step for histology image display and annotation is systems allowing the pathologist to interact with automated analyses, teaching and correcting the system in real time. Some efforts have been made in this area: after the automated segmentation in [35], experts can click on nuclei to indicate that they either need to be split or merged, and Cytomine is beginning to include collaborative proofreading of automated cancer segmentation.

ROI IDENTIFICATION

A single biopsy can generate dozens of high-resolution WSIs, however, often, only a small region of this vast quantity of tissue is diagnostically useful. There is, therefore, a need for fast computational methods that can identify these ROIs in whole-slide histology images. Once identified, these regions can be passed to a pathologist or to subsequent steps of a larger automated histology pipeline.

One approach is to simply downsample the input; e.g., [36] uses a low-resolution input to extract features based on color and sparse coding of subpatches. These features are classified via a support vector machine (SVM) to detect ROIs. Another approach is to process the image at multiple scales, such as in [37], where WSIs of breast cancer tissue are recursively partitioned via color clustering at increasingly fine resolutions to efficiently identify tissue versus nontissue and lesion versus normal regions. Yet another approach is to detect objects such as glands and identify ROIs based on these [38]. There are generally fewer objects on a slide than pixels, so processing objects can be more efficient than processing pixels.

NUCLEUS DETECTION

Nuclei are prevalent in histology images and their size, shape, distribution, and texture are relevant for many analysis tasks such as identifying inflammation, identifying and grading cancer, and

determining tissue type. Therefore, locating nuclei is a critical step in many histology analysis systems. Nucleus detection in histology images is challenging because nuclei can be tightly clustered (Figure 11) and vary in size, shape, and color depending on their cell type. Even within a single cell type, the stain may not penetrate all nuclei equally, leaving some darkly stained and others lightly stained. Finally, human nuclei are around the same scale as the thickness of the tissue slice (5 μm), meaning that, for some nuclei, only a portion appears on the slide.

Basic approaches to nucleus detection involve color clustering of the pixels; e.g., the work in [39] uses clustering in the lab color space to identify four subcellular components—nuclei, cytoplasm, neuropil, and background—and [40] uses expectation-maximization clustering to identify regions of lymphocyte nuclei, stroma, cancer nuclei, and background.

Using only local information omits the strong prior knowledge we have about nucleus size and shape. Approaches that move beyond local information include [36] and [40], which use active contours to refine their initial color segmentation, and [14], which uses a graph cut based on color and Laplacian of Gaussian features. Similarly, the authors in [35] use Laplacian of Gaussian filtering with clever scale selection to detect nucleus seed points followed by local maximum clustering to form a rough segmentation. Another approach is to frame nucleus detection as a classification problem; e.g., in [41], an SVM is trained to detect rectangular windows containing nuclei; the features used are pixel intensities and Laplacian of Gaussian edge intensities.

Finally, several methods use heuristics to separate clustered nuclei, including the curvature-based reasoning in [14] and concavity detection in [40]. The method in [35] separates nuclei via graph cuts (Figure 11). These approaches achieve impressive-looking results even when nuclei are clustered, but large, hand-annotated data sets will be necessarily to compare them and understand their strengths and weaknesses.

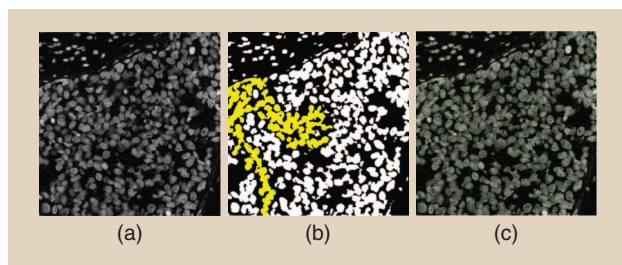
CELL CLASSIFICATION

Most histology images contain cells of several types; moving beyond nucleus detection to cell classification can provide valuable diagnostic information. For example, the method in [42] classifies cells as centroblasts or normal, which is useful for cancer grading. It uses color and Fourier-based texture features with quadratic discriminant analysis as the classifier. The nuclei found in [40] are classified as belonging to lymphocytes or other based on their color.

Other work focuses on detecting mitotic cells. Notably, [43] presents a contest data set for this task which comprises 50 images collected on each of two different slide scanners as well as a multispectral scanner. The highest-scoring approach at the time of the contest was based on a deep convolutional neural network [44].

MULTICELLULAR STRUCTURE DETECTION

Cells in histology images are not solitary, rather they are part of organized structures (e.g., glands, acini). Detecting and analyzing these structures is a unique challenge in histology image analysis.



[FIG11] The clustering of nuclei makes nucleus segmentation challenging. (a) In this grayscale, H-only image, the closely packed and overlapping nuclei cannot be separated with threshold-based clustering, resulting in (b) large connected components (yellow cluster). The method discussed in [35] breaks apart these clusters via graph cuts, resulting in (c) a good separation of nuclei. (Figure adapted with permission from [35].)

Because there are many types of multicellular structures, these tasks tend to be more varied and application-driven than the ones we have discussed before, hence we give only a few examples here.

Glands are a multicellular structure common to many tissue types (e.g., salivary, breast, prostate, pancreatic, sinonasal, gastrointestinal tissues), and changes in their morphology can be an important indicator of disease. In histology images, most glands appear as clear areas surrounded by cells. The method in [38] leverages the clear areas to find seed pixels and then uses region growing to segment the glands. Going further, [38] classifies glands as malignant or normal based on their size with a Markov random field to impose spatial smoothness. In [45], a graph is built on top of a colon tissue image with nodes corresponding to either nucleus or nonnucleus objects. The colon glands can then be described by subgraphs around a user-selected point at their center. Matching these subgraphs to reference ones from healthy or diseased glands allows classification of the input image.

Some approaches to segmenting multicellular structures treat them like nuclei, e.g., [46] segments lymphoid follicles, which are organized groups of lymphoid cells, using active contours. Detected regions are split based on curvature and false detections are further trimmed based on color.

TISSUE SEGMENTATION

Tissues are organized groups of cells. Identifying them in an image is important for diagnosis or giving context to subsequent analyses. The wide variety of tissues and the complexity of their appearances makes this a challenging problem. The method in [47] uses local pixel intensities as features and is able to segment bone, cartilage, and fat tissue in teratoma tumor images. We presented a segmentation method [48] inspired by the lack of edges in histology images. It uses local color histograms rather than edge-based features and outperforms generic methods for tissue segmentation. The authors in [49] showed that segmentations seeking homogeneity of objects such as cells and crypts, rather than simply pixel homogeneity, perform well on colon tissue.

Some tissue-level analyses amount to diagnoses; e.g., in [50], small subregions of a prostate tissue image are classified as *normal*, *stroma*, or *prostatic adenocarcinoma*. Normal regions are those around glands, which are simple to segment because of their white centers. Stroma and cancer regions are distinguished using Haralick texture features. Another group [51] used a graph-based methodology to segment regions of cancer in colon images, where features were based on the frequency of co-occurrence of nodes.

DIAGNOSIS

Automated diagnosis is essentially image classification: given a histology image, what disease does it represent? As such, most approaches follow the paradigm of feature extraction followed by classification. The authors in [52] address the problem of classifying subtypes of renal tumor in expert-selected ROIs. They use Fourier shape descriptors extracted from binary masks of nuclei, cytoplasm, and unstained regions as features and a series of SVM classifiers arranged in a directed acyclic graph to distinguish between three types of renal cell carcinoma and one benign tumor. Based on

which shape descriptors are most distinguishing during classification, the authors can identify which shapes are indicative of each tissue type. In [53], hand-selected ROIs of breast tissue are classified as *normal*, *in situ cancer*, or *invasive cancer*. They use generic features including local binary patterns, co-occurrence matrix statistics, and curvelet coefficient statistics. Classification proceeds in two stages, first a random subspace ensemble of SVMs, then a random subspace ensemble of neural networks. Each stage may either classify an image or reject it as too difficult. Images rejected from the first stage move on to the second, while images rejected from the second stage are viewed by a human expert. The method achieves accuracy of over 99% with a 1.94% rejection rate. The recent conference paper [54] moves away from hand-designed features by learning features from the data using sparse representation.

Grading is diagnosing the severity of a disease. Established grading scales are useful guides for algorithm development, and automated grading promises increased repeatability over human grading. The method in [36] grades breast cancer by analyzing the size, shape, and texture of the nuclei inside an ROI using a Bayesian classifier. Another approach is to include some notion of cellular components without specifically segmenting nuclei. In, [55] grayscale thresholding and morphology are used to find blobs that are then classified based on intensity and size into three different nucleus types and stroma. The image is then segmented into areas of high and low nucleus density. Features including nucleus count, nucleus spacing, and tubule count are extracted for the high density areas only. The images are finally classified into grade one, two, or three with a quadratic classifier. In [39], Haralick features are extracted from cytoplasm and neuropil regions and used to determine the differentiation level of neuroblastoma tumors. The classification step is an ensemble of seven different classifiers and dimensionality reduction methods combined via weighted voting. The method handles WSIs by splitting them into small tiles that are processed in parallel as well as first classifying a downsampled version of the tile and using higher resolution data only if the classification certainty is low.

Finally, generic features can be used. In [56], prostate cancer is detected and graded based on color and color co-occurrence features and a random forest classifier. In [57], two different lung cancer subtypes are differentiated in images of tissue microarrays and hand-selected ROIs from full slides. The red and blue channels of the image are histogram-stretched to enhance the H&E contrast, and then Haralick and densitometric (e.g., mean pixel value, pixel center of mass, etc.) features are used in a boosting decision tree, achieving accuracy over 90%.

CONCLUSIONS

Histology is a critical tool in medicine and therefore automated histology analysis could have a profound effect on health-care quality, availability, and cost. At the same time, the field provides a host of fascinating signal processing challenges. How can tissue folds be detected and corrected? How can images of deformable tissue sections be registered to recover the 3-D shape of a tumor? What features are needed to explain the complex architecture of tissues such as skin? How can we detect rare cell types

with a scarcity of labeled data? And how can we run any of these methods efficiently on huge WSIs?

We hope this discussion has served both as a useful primer and as a call to action. Much has been accomplished in this field, but these systems are far from clinical acceptance. To get there, we need algorithms that demonstrate robustness and that solve relevant problems in medicine. Robustness will come from sharing data sets and algorithms so that they can be truly validated and compared. Relevance will depend on the continued efforts of pathologists and engineers to collaborate on defining and refining algorithms. Given the impact histology image analysis can make on the future of health care, it is well worth the effort.

AUTHORS

Michael T. McCann (mtmccann@cmu.edu) received the B.S. degree in biomedical engineering in 2010 from the University of Michigan. He is currently a Ph.D. student in the Center for Bioimage Informatics and the Department of Biomedical Engineering at Carnegie Mellon University, where he received a National Science Foundation Graduate Research Fellowship. His work focuses on developing signal processing tools for histology images, including methods for normalization, registration, and tissue segmentation/identification. He is a Student Member of the IEEE.

John A. Ozolek (ozolja@upmc.edu) graduated from Case Western Reserve University in 1985 with a B.S. degree in biology and obtained his M.D. degree from the University of Pittsburgh School of Medicine in 1989. He completed postgraduate residency training in pediatrics and anatomic pathology and fellowships in neonatal-perinatal medicine and pediatric pathology at the University of Pittsburgh. He is currently an assistant professor of pathology at the University of Pittsburgh and staff pathologist at the Children's Hospital of Pittsburgh. His research interests focus on the role of stem and neural crest cells, particularly in the pathogenesis of branchial cleft cysts, developmental abnormalities, and tumors of the head and neck in children. He has extensive collaborations with the Center for Bioimage Informatics as well as the Department of Biomedical Engineering and the Department of Mathematics at Carnegie Mellon University to develop algorithms for automating tissue recognition and correlation of histology and high-resolution magnetic resonance images.

Carlos A. Castro (ccastro@mwri.magee.edu) received the M.D. degree from Rosario University Medical School, Bogota, Colombia, in 1982 and the D.M.D. degree from Javeriana University Dental School in Bogota, Colombia, in 1988. He went on to do his maxillofacial surgery fellowship at Rosario University Medical School and subsequently did oral, orthognathic surgery, and gnathology specialties at the same universities as well as the Military University School of Medicine. He has worked in general pathology and subsequently surgical and perinatal pathology at the University of Pittsburgh and as a pathologist assistant at Magee-Womens Hospital in Pittsburgh, Pennsylvania. He has also worked as a senior research associate, visiting research instructor, senior faculty research instructor, and codirector of the Magee-Womens Research Institute and Foundation Histology Core. He is currently the director of the same histology core.

Bahram Parvin (b_parvin@lbl.gov) is a principal scientist at the Lawrence Berkeley National Laboratory and a professor of electrical and biomedical engineering at the University of Nevada, Reno. His laboratory focuses on technology development for realization of pathway pathology, elucidating molecular signatures of aberrant morphogenesis in normal and engineered matrices, and screening for probes for labeling and cargo delivery. He has published over 100 papers and was the general chair of the 2013 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. He is a member of the IEEE Bioimaging and Signal Processing Technical Committee and is an associate editor of *IEEE Transactions on Medical Imaging*.

Jelena Kovačević (j.kovacevic@ieee.org) received a Ph.D. degree from Columbia University. She then joined Bell Labs, followed by Carnegie Mellon University in 2003, where she is currently the Edward David Schramm Professor and head of the Department of Electrical and Computer Engineering, professor of biomedical engineering, and the director of the Center for Bioimage Informatics. She received the Belgrade October Prize and the E.I. Jury Award at Columbia University. She is a coauthor on an IEEE Signal Processing Society award-winning paper and is a coauthor of *Wavelets and Subband Coding* and *Foundations of Signal Processing*. She was editor-in-chief of *IEEE Transactions on Image Processing* (2002–2006). She was a keynote speaker at a number of meetings and has been involved in organizing numerous conferences. Her research interests include multiresolution techniques and biomedical applications. She is a Fellow of the IEEE.

REFERENCES

- [1] K. S. Suvarna, C. Layton, and J. D. Bancroft, *Bancroft's Theory and Practice of Histological Techniques*, 7th ed. London, UK: Churchill Livingstone, Nov. 2012.
- [2] D. J. Cook, *Cellular Pathology: An Introduction to Techniques and Applications*, 2nd ed. Oxfordshire, UK: Scion Publishing Ltd., June 2006.
- [3] T. C. Cornish, R. E. Swapp, and K. J. Kaplan, "Whole-slide imaging: Routine pathologic diagnosis," *Adv. Anat. Pathol.*, vol. 19, no. 3, pp. 152–159, May 2012.
- [4] S. Kothari, J. H. Phan, T. H. Stokes, and M. D. Wang, "Pathology imaging informatics for quantitative analysis of whole-slide images," *J. Am. Med. Inform. Assoc.*, vol. 20, no. 6, pp. 1099–1108, Aug. 2013.
- [5] T. W. Bauer, L. Schoenfield, R. J. Slaw, L. Yerian, Z. Sun, and W. H. Henricks, "Validation of whole slide imaging for primary diagnosis in surgical pathology," *Arch. Pathol. Lab. Med.*, vol. 137, no. 4, pp. 518–524, Apr. 2013.
- [6] S. Krishnamurthy, K. Mathews, S. McClure, M. Murray, M. Gilcrease, C. Albarra-cin, J. Spinoso, B. Chang, J. Ho, J. Holt, A. Cohen, D. Giri, K. Garg, R. L. Bassett, and K. Liang, "Multi-institutional comparison of whole slide digital imaging and optical microscopy for interpretation of hematoxylin-eosin-stained breast tissue sections," *Arch. Pathol. Lab. Med.*, vol. 137, no. 12, pp. 1733–1739, Dec. 2013.
- [7] L. Pantanowitz, J. H. Sinard, W. H. Henricks, L. A. Fatheree, A. B. Carter, L. Contis, B. A. Beckwith, A. J. Evans, A. Lal, and A. V. Parwani, "Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the college of American pathologists pathology and laboratory quality center," *Arch. Pathol. Lab. Med.*, vol. 137, no. 12, pp. 1710–1722, Dec. 2013.
- [8] S. Al-Janabi, A. Huisman, P. G. J. Nikkels, F. J. W. ten Kate, and P. J. van Diest, "Whole slide images for primary diagnostics of paediatric pathology specimens: A feasibility study," *J. Clin. Pathol.*, vol. 66, no. 3, pp. 218–223, Mar. 2013.
- [9] M. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, Apr. 2009.
- [10] T. J. Fuchs and J. M. Buhmann, "Computational pathology: Challenges and promises for tissue analysis," *Comput. Med. Imaging Graph.*, vol. 35, nos. 7–8, pp. 515–530, Oct. 2011.
- [11] M. W. Davidson. (2014, Feb.). Introduction to optical microscopy, digital imaging, and photomicrography. [Online]. Available: <http://micro.magnet.fsu.edu/primer/index.html>
- [12] S. Kothari, J. H. Phan, R. A. Moffitt, T. H. Stokes, S. E. Hassberger, Q. Chaudry, A. N. Young, and M. D. Wang, "Automatic batch-invariant color segmentation of

histological cancer images," in *Proc. IEEE Int. Symp. Biomedical Imaging*, Chicago, IL, Mar. 2011, pp. 657–660.

[13] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A non-linear mapping approach to stain normalisation in digital histopathology images using image-specific colour deconvolution," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729–1738, June 2014.

[14] H. Chang, J. Han, A. Borowsky, L. Loss, J. W. Gray, P. T. Spellman, and B. Parvin, "Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association," *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 670–682, Apr. 2013.

[15] U. D. Braumann, J. P. Kuska, J. Einkenkel, L. C. Horn, M. Löffler, and M. Hockel, "Three-dimensional reconstruction and quantification of cervical carcinoma invasion fronts from histological serial sections," *IEEE Trans. Med. Imag.*, vol. 24, no. 10, pp. 1286–1307, Oct. 2005.

[16] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, Aug. 2001.

[17] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomedical Imaging*, Boston, MA, June 2009, pp. 1107–1110.

[18] M. Gavrilovic, J. C. Azar, J. Lindblad, C. Wahlby, E. Bengtsson, C. Busch, and I. B. Carlsson, "Blind color decomposition of histological images," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 983–994, June 2013.

[19] P. Tadrous, "Digital stain separation for histological images," *J. Microsc.*, vol. 240, no. 2, pp. 164–172, Nov. 2010.

[20] M. T. McCann, J. Majumdar, C. Peng, C. A. Castro, and J. Kovacevic, "Algorithm and benchmark dataset for stain separation in histology images," in *Proc. IEEE Int. Conf. Image Processing*, Paris, Oct. 2014.

[21] P. A. Bautista and Y. Yagi, "Improving the visualization and detection of tissue folds in whole slide images through color enhancement," *J. Pathol. Inform.*, vol. 1, p. 25, Nov. 2010.

[22] S. Kothari, J. H. Phan, and M. D. Wang, "Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade," *J. Pathol. Inform.*, vol. 4, p. 22, Aug. 2013.

[23] J. Chappelow, J. E. Tomaszewski, M. Feldman, N. Shih, and A. Madabhushi, "HistoStitcher(copyright): An interactive program for accurate and rapid reconstruction of digitized whole histological sections from tissue fragments," *Comput. Med. Imaging Graph.*, vol. 35, nos. 7–8, pp. 557–567, Mar. 2011.

[24] M. Schwieter, T. Böhler, H. K. Hahn, U. Dahmen, and O. Dirsch, "Registration of histological whole slide images guided by vessel structures," *J. Pathol. Inform.*, vol. 4 (Suppl.), no. 2, p. 10, 2013.

[25] D. Mueller, D. Vossen, and B. Hulsken, "Real-time deformable registration of multi-modal whole slides for digital pathology," *Comput. Med. Imaging Graph.*, vol. 35, nos. 7–8, pp. 542–556, June 2011.

[26] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 196–205, Jan. 2010.

[27] L. Cooper, O. Sertel, J. Kong, G. Lozanski, K. Huang, and M. Gurcan, "Feature-based registration of histopathology images with different stains: An application for computerized follicular lymphoma prognosis," *Comput. Methods Programs Biomed.*, vol. 96, no. 3, pp. 182–192, Dec. 2009.

[28] G. Xiao, B. N. Bloch, J. Chappelow, E. M. Genega, N. M. Rofsky, R. E. Lenkinski, J. Tomaszewski, M. D. Feldman, M. Rosen, and A. Madabhushi, "Determining histology-MRI slice correspondences for defining MRI-based disease signatures of prostate cancer," *Comput. Med. Imaging Graph.*, vol. 35, nos. 7–8, pp. 568–578, Jan. 2011.

[29] H. Park, M. R. Pierr, A. Khan, R. Shah, H. Hussain, J. Siddiqui, T. L. Chen-evert, and C. R. Meyer, "Registration methodology for histological sections and in vivo imaging of human prostate," *Acad. Radiol.*, vol. 15, no. 8, pp. 1027–1039, Aug. 2008.

[30] A. Goode, B. Gilbert, J. Harkes, D. Jukic, and M. Satyanarayanan, "OpenSlide: A vendor-neutral software foundation for digital pathology," *J. Pathol. Inform.*, vol. 4, p. 27, Sept. 2013.

[31] H. Chang, G. Fontenay, J. Han, G. Cong, F. Baehner, J. Gray, P. Spellman, and B. Parvin, "Morphometric analysis of TCGA glioblastoma multiforme," *BMC Bioinform.*, vol. 12, no. 1, p. 484, Dec. 2011.

[32] TCGA Research Network. (2014, Aug.). The cancer genome atlas. [Online]. Available: <http://cancergenome.nih.gov/>

[33] W.-K. Jeong, J. Schneider, S. Turney, B. Faulkner-Jones, D. Meyer, R. Westermann, R. Reid, J. Lichtman, and H. Pfister, "Interactive histology of large-scale biomedical image stacks," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1386–1395, Nov. 2010.

[34] R. Marée, B. Stévens, L. Rollus, N. Rocks, X. M. Lopez, I. Salmon, D. Cataldo, and L. Wehenkel. (2013, Sept.). A rich internet application for remote visualization and collaborative annotation of digital slides in histology and cytology. *Diagn. Pathol.*, 8(Suppl 1), p. S26. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3849538/>

[35] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 841–852, Apr. 2010.

[36] C.-H. H. Huang, A. Veillard, L. Roux, N. Loménie, and D. Racoceanu, "Time-efficient sparse analysis of histopathological whole slide images," *Comput. Med. Imaging Graph.*, vol. 35, nos. 7–8, pp. 579–591, Dec. 2011.

[37] V. Roullier, O. Lézoray, V.-T. Ta, and A. Elmoataz, "Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization," *Comput. Med. Imaging Graph.*, vol. 35, nos. 7–8, pp. 603–615, May 2011.

[38] J. P. Monaco, J. E. Tomaszewski, M. D. Feldman, I. Hagemann, M. Moradi, P. Mousavi, A. Boag, C. Davidson, P. Abolmaesumi, and A. Madabhushi, "High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models," *Med. Image Anal.*, vol. 14, no. 4, pp. 617–629, Aug. 2010.

[39] J. Kong, O. Sertel, H. Shimada, K. L. Boyer, J. H. Saltz, and M. N. Gurcan, "Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation," *Pattern Recogn.*, vol. 42, no. 6, pp. 1080–1092, June 2009.

[40] H. Fatakdawala, J. Xu, A. Basavanahally, G. Bhanot, S. Ganesan, M. Feldman, J. E. Tomaszewski, and A. Madabhushi, "Expectation-maximization-driven geodesic active contour with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1676–1689, July 2010.

[41] J. Han, T. Breckon, D. Randell, and G. Landini, "The application of support vector machine classification to detect cell nuclei for automated microscopy," *Mach. Vis. Appl.*, vol. 23, no. 1, pp. 15–24, June 2012.

[42] K. Belkacem-Boussaid, M. Pennell, G. Lozanski, A. Shana'ah, and M. Gurcan. (2010, Oct.). Computer-aided classification of centroblast cells in follicular lymphoma. *Anal. Quant. Cytol. Histol.*, 32(5), pp. 254–260. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3078581/>

[43] L. Roux, D. Racoceanu, N. Loménie, M. Kulikova, H. Irshad, J. Klossa, F. Capron, C. Genestie, G. Le Naour, and M. N. Gurcan, "Mitosis detection in breast cancer histological images An ICPR 2012 contest," *J. Pathol. Inform.*, vol. 4, p. 8, May 2013.

[44] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, Nagoya, Japan, 2013, pp. 411–418.

[45] E. Ozdemir and C. Gunduz-Demir, "A hybrid classification model for digital pathology using structural and statistical pattern recognition," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 474–483, Feb. 2013.

[46] K. Belkacem-Boussaid, S. Samsi, G. Lozanski, and M. N. Gurcan, "Automatic detection of follicular regions in H&E images using iterative shape index," *Comput. Med. Imaging Graph.*, vol. 35, nos. 7–8, pp. 592–602, Oct. 2011.

[47] C. Chen, J. A. Ozolek, W. Wang, and G. K. Rohde, "A general system for automatic biomedical image segmentation using intensity neighborhoods," *Int. J. Biomed. Imag.*, vol. 2011, Mar. 2011.

[48] M. T. McCann, D. G. Mixon, M. C. Fickus, C. A. Castro, J. A. Ozolek, and J. Kovacevic, "Images as occlusions of textures: A framework for segmentation," *IEEE Trans. Image Processing*, vol. 23, no. 5, pp. 2033–2046, May 2014.

[49] A. B. Tosun, M. Kandemir, C. Sokmensuer, and C. Gunduz-Demir, "Object-oriented texture analysis for the unsupervised segmentation of biopsy images for cancer detection," *Pattern Recogn.*, vol. 42, no. 6, pp. 1104–1112, June 2009.

[50] J. Diamond, N. H. Anderson, P. H. Bartels, R. Montironi, and P. W. Hamilton, "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Hum. Pathol.*, vol. 35, no. 9, pp. 1121–1131, Sept. 2004.

[51] A. Cagri, A. Burak, C. Aykanat, C. Sokmensuer, and C. Gunduz-Demir. (2012, June). Multilevel segmentation of histopathological images using co-occurrence of tissue objects. *IEEE Trans. Biomed. Eng.*, 59(6), pp. 1681–1690. [Online]. Available: <http://dx.doi.org/10.1109/tbme.2012.2191784>

[52] S. Kothari, J. Phan, A. Young, and M. Wang, "Histological image classification using biologically interpretable shape-based features," *BMC Med. Imag.*, vol. 13, no. 9, Mar. 2013.

[53] Y. Zhang, B. Zhang, F. Coenen, and W. Lu, "Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles," *Mach. Vis. Appl.*, vol. 24, no. 7, pp. 1405–1420, Oct. 2013.

[54] H. Chang, N. Nayak, P. Spellman, and B. Parvin, "Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching," in *Proc. Int. Conf. Medical Image Computing Computer-Assisted Intervention*, Nagoya, Japan, Sept. 2013, vol. 8150, pp. 91–98.

[55] S. Petushi, F. Garcia, M. Haber, C. Katsinis, and A. Tozeren. (2006, Oct.). Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC Med. Imag.*, 6(1), p. 14. [Online]. Available: <http://www.biomedcentral.com/1471-2342/6/14>

[56] M. D. DiFranco, G. O'Hurley, E. W. Kay, R. W. Watson, and P. Cunningham, "Ensemble based system for whole-slide prostate cancer probability mapping using color texture features," *Comput. Med. Imaging Graph.*, vol. 35, nos. 7–8, pp. 629–645, Jan. 2011.

[57] C.-W. Wang and C.-P. Yu, "Automated morphological classification of lung cancer subtypes using H&E tissue images," *Mach. Vis. Appl.*, vol. 24, no. 7, pp. 1383–1391, Oct. 2013.

