Spectral Imaging Perspective on Cytomics

Richard M. Levenson*
CRI, Woburn, Massachusetts 01801
Received 15 February 2005; Revision Received 4 January 2006; Accepted 11 January 2006

Background: Cytomics involves the analysis of cellular morphology and molecular phenotypes, with reference to tissue architecture and to additional metadata. To this end, a variety of imaging and nonimaging technologies need to be integrated. Spectral imaging is proposed as a tool that can simplify and enrich the extraction of morphological and molecular information. Simple-to-use instrumentation is available that mounts on standard microscopes and can generate spectral image datasets with excellent spatial and spectral resolution; these can be exploited by sophisticated analysis tools.

Methods: This report focuses on brightfield microscopy-based approaches. Cytological and histological samples were stained using nonspecific standard stains (Giemsa; hematoxylin and eosin (H&E)) or immunohistochemical (IHC) techniques employing three chromogens plus a hematoxylin counterstain. The samples were imaged using the Nuance™ system, a commercially available, liquid-crystal tunable-filter-based multispectral imaging platform. The resulting data sets were analyzed using spectral unmixing algorithms and/or learn-by-example classification tools.

Results: Spectral unmixing of Giemsa-stained guinea-pig blood films readily classified the major blood elements. Machine-learning classifiers were also successful at the same task, as well in distinguishing normal from malignant regions in a colon-cancer example, and in delineating regions of inflammation in an H&E-stained kidney sample. In an example of a multiplexed ICH sample, brown, red, and blue chromogens were isolated into separate images without crosstalk or interference from the (also blue) hematoxylin counterstain.

Conclusion: Cytomics requires both accurate architectural segmentation as well as multiplexed molecular imaging to associate molecular phenotypes with relevant cellular and tissue compartments. Multispectral imaging can assist in both these tasks, and conveys new utility to brightfield-based microscopy approaches. © 2006 International Society for Analytical Cytology

Key terms: multispectral; imaging; immunohistochemistry; multiplexing; machine-learning

“Cytomics”—a term and discipline so new, it still seems to call for the quotation marks—builds on cytometry, the study of cellular properties, but hopes to transcend it in its goal of integrating multiple sources and scales of information. Its essence is still not completely defined, and probably depends on the various technologies available to its practitioners. The purpose of this communication is to illustrate the potential of yet another new technology, namely spectral imaging, along with parametric as well as machine-learning-based analysis tools, to contribute to the evolution of this field, as well as to established areas such as routine histopathology, immunohistochemistry, and the like. Not yet realized is the hope of creating tools that will in fact consolidate data from cells, tissues, whole-patient information, and statistical distillations from populations into new and more valid evaluation engines. Bayesian approaches, data fusion, systems biology, and data mining are some generic concepts that may soon prove valuable in the effective integration of cellular insights with other information sources. However, these mostly live in the (near-) future—what is hoped to be shown here are techniques for deriving high-quality and feature-rich data from cellular images, demonstrating new approaches to acquisition, segmentation, classification, evaluation, and quantitation.

Although anatomic pathology is a visual discipline, it is surprising that digital and quantitative imaging continues to play a minimal role in clinical practice. This is true despite remarkable improvements in the capabilities, cost, and availability of cameras (sensors) and computers. A number of factors are responsible for this. The human eye–brain combination is skilled at perceiving and interpreting visual scenes and is hard to improve upon. Second, pathology is difficult. Third, current imaging hardware–software configurations often require trained operator assistance, although gradually the demands on the

*Correspondence to: Richard Levenson. E-mail: rlevenson@cri-inc.com
Contract grant sponsor: National Cancer Institute SBIR Award; Contract grant numbers: 1 R44 CA88684-01; Contract grant sponsor: National Cancer Institute Bioengineering Research Partnership Award; Contract grant numbers: 1 R01 CA108468-01.
Published online 5 May 2006 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/cyto.a.20292
user are lessening. To move quantitative cytometry (or cytomics) forward, some advances over conventional (semi-)automated tools are required.

Although discussions on cytometry often focus on the analysis following image acquisition—since the techniques used for capturing images are well-known and relatively stable—it is useful to consider the role that the image acquisition process itself plays in the obtainable information content of a sample. Images of slides are typically captured using monochrome or color CCD or CMOS sensors, configured as either line-scanners or area detectors. Additional acquisition approaches can use scanned lasers, either in confocal mode (1) or in slide-based cytometry instruments (2). These can generate images with information in a few wavelength regions, and the latter type can simulate color images in brightfield using red, green, and blue laser illumination. More complex, confocal instruments, now available from all major microscope manufacturers, can be used to look at many wavelengths at once, and these can be useful in cytomics applications (3). However, these are too complex and expensive for routine use, and are not suitable for brightfield (standard transmission) microscopy, which is how the vast bulk of clinically relevant specimens are currently examined. Consequently, this report will focus on what can be accomplished in brightfield view, using multispectral imaging and analysis tools.

MULTISPECTRAL VS. RED-GREEN-BLUE

Information sources within static samples that are accessible to imaging approaches fall into at least 4 major categories: spatial, textural, chromatic (or spectral), and molecular, clearly with cross-talk. Some of these elements have been more exploited than others, and using mutual information from each remains under-explored.

Currently, pathologists evaluate stained tissue sections and determine disorders of cellular size, shape, and organization through a process of pattern recognition and, in histology, structural evaluation of overall tissue architecture, paying particular attention to the possible breaching of usually inviolate tissue boundaries. This examination occurs at a variety of magnifications. A diagnosis can be rendered based on a low-power survey if, for example, an easily recognizable tumor is present. Architectural clues usually are paramount, and specific color information is rarely necessary. In fact, until recently, most pathology texts were illustrated almost exclusively with grayscale (black and white) photomicrographs. However, no pathologist would be content with viewing specimens in black and white, so color definitely contributes to the overall assessment.

Cancer and other derangements are manifested by perturbations in cells’ macromolecular components (structural phenotype), which may cause alterations in the strength and pattern of binding of stains. The color and appearance of stains is altered depending on what they bind to and whether other stains form part of the coordination complex. The phenomenon known as metachromasia is the most visible of these alterations being exhibited, for example, when stains that normally stain bluish appear purple.

While metachromasia is detectable by eye, more subtle changes in the absorbance spectrum of a stain (or stain combination such as hematoxylin and eosin, H&E) can only be appreciated by more sensitive techniques.

Combining standard color-based imaging with automated analysis has proven to be difficult. Color is not an intrinsic physical property, but is rather a neuropsychological construct. The mapping of images acquired with instruments into visual color-space is not fully reproducible or quantitative—different illuminants, sensors, algorithms, and other variables make it difficult to rely on color-based metrics, especially when trying to compare results from one instrument platform to another. Also, color information is degenerate, and only roughly maps the true spectral content of an image, and is incapable of distinguishing signals that can be spectrally completely different. A simple example: the color, yellow. Yellow can be generated by photons of a limited energy (wavelength) range, as can be seen emerging from a prism, grating or rainbow, or it can be created by mixing roughly equal quantities of red light and green light. The resulting color is indistinguishable (by eye or camera) from the "pure" yellow, yet is completely unlike spectrally. In other words, colorimetry and spectroscopy are different disciplines. Nevertheless, important commercial systems dedicated to quantitative pathology and based on standard color sensors have been developed and are in use. These are well-suited to certain applications, but cannot perform many others because of the intrinsic limitations of standard color sensors and analysis algorithms.

An orthogonal problem, namely, the variability in sample preparation and staining protocols (such as those involving the ubiquitous H&E dyes, for example) also discourages the development of universally applicable quantitative imaging approaches relying on color information. Recent trends mitigate against this difficulty due in part to the increasing standardization and automation of both sample preparation and staining. Thin-prep specimen preparation and even standard H&E slide staining are now being accomplished using automated instruments and carefully quality-controlled reagents, ensuring better uniformity than was achievable in the past.

SPECTRAL IMAGING AND PATHOLOGY

With the advent of spectral imaging techniques that can capture images with accurate spectral content (intensity as a function of wavelength), new possibilities emerge. Spectral imaging generates what is often referred to as a "data cube," a three-dimensional data-set that contains spectral information about the sample at every spatial position or picture-element (pixel) of the image, data that can reveal not only what a scene "looks like" but also what chemical or anatomic features it may contain. It is information-rich, encouraging the application of numerous, powerful analytical techniques (4–6). In biomedicine, it appears that spectral imaging will be valuable in surgical pathology, multicolor fluorescence, and immunohistochemistry (7–12). The spectral information can be used in pathology specimens, for example, to allow the differen-
tiation of each pixel in an image according to its spectral "signature," with each signature having the advantage of correlation with spatial information.

One hypothesis underlying this approach is that samples stained with standard histological dyes contain useful spectral information, some of which is not retrievable using standard monochrome or RGB-based sensors. Preliminary evidence suggests that this is the case (13–19). These staining phenomena, reflecting poorly understood molecular interactions between tissue and dyes, may be as informative as conventional spatial morphology in certain cases (20). The challenge has been how to capture that information in a quantitative manner.

Supplementing the use of the standard, relatively unselective histology dyes in recent decades, specific molecular probes have been added to the imaging armamentarium available to cell biology and pathology. These probes are typically coupled to fluorescent or chromogenic indicators. To be able to examine more than one or two of these targets at a time (multiplexing), imaging techniques have to be able to separate all the signals into their own channel, without cross-talk. Spectral imaging is one of the most promising techniques for providing high levels of multiplexing. Of course, adding spectral data to RGB-based information promises to exacerbate the challenges posed by the increasingly popular automated high-resolution whole-slide scanning approach, namely, acquisition time and already huge file sizes. Nevertheless, it can be anticipated that ever-faster scanning technologies will be implemented, which along with sophisticated data compression techniques, should in part alleviate these problems.

**METHODS**

Many techniques are now available for acquiring spectral images in microscopy (12). A pioneer in this area is Applied Spectral Imaging, (Migdal Ha-Emek, Israel), which developed a Fourier-transform interferometer for this purpose. Band-sequential devices such as acousto-optical tunable filters (AOTFs, (7)) liquid crystal tunable filters (LCTFs), circular or linear variable interference filters (21) can be used, as can line-scanners coupled to gratings or to prisms. Other approaches use tunable light sources (22), and some single-shot techniques in which spectral and spatial information are captured with a single acquisition event have also been developed (11,23). As can be expected, all of these approaches have advantages and disadvantages, and any of them would be appropriate for acquiring spectral data for cytomics purposes.

**Liquid Crystal Tunable Filter-based Imaging**

The examples shown in this communication were all acquired using an LCTF-based system (24) that can be rapidly tuned to transmit any wavelength within their design range with about 1-nm precision, and can be inserted into the microscope optical train without discernible effect on image resolution. Such tunable filters are well-suited to fluorescence-based analyses, having proved useful for multicolor fluorescence in situ hybridization (FISH), for resolving multiple species of green fluorescent proteins with overlapping emission spectra, and for the identification and elimination of interfering autofluorescence, and have recently shown great utility in small-animal imaging applications (25).

The principle is simple: using the Nuance™ spectral imaging system (CRI, Woburn, MA), images can be automatically captured at a number of wavelengths. While a typical acquisition series for brightfield might extend from 440 nm to 680 nm, stepping every 10 or 20 nm, more or fewer wavelengths can be imaged at arbitrarily spaced wavelength regions. In brightfield, an autoexposure function is used to ensure that all white (clear) regions of a specimen fill the wells on the CCD, giving a completely flat “white” spectrum and maximum achievable signal-to-noise performance.

**Samples**

Histopathology and cytology samples representing a variety of applications will be shown. Guinea-pig bronchoalveolar lavage (BAL) smears (for differentials), H&E-stained sections of inflamed kidney and colon carcinoma, and multicolor immunohistochemistry are examples.

**RESULTS**

**Segmentation**

The major task in image analysis is segmentation, that is, the separation of the image data (a set of pixels) into meaningful regions, such as nucleus vs. cytoplasm. This is still surprisingly hard in many situations, as is the task of further splitting the resulting segmented objects into biologically apt units (separating merged blobs into the appropriate number of individual nuclei, for example (26). It is after segmentation that classification and evaluation typically occurs. Spectral cues prove to be very useful for the segmentation process and superior to monochrome-only approaches, since they are less susceptible to variations in local intensities, and thresholding difficulties. Simple metrics for capturing spectral differences, such as minimum square error (Euclidean distance) (27) or Mahalanobis distances (28) can often be used.

**Guinea-Pig White-Blood-Cell Differential**

While automated blood cell analyzers have been developed for humans, there is little automation available for species without health insurance. Spectral imaging analysis techniques, including minimum square error classification, and sequential component removal (features of the commercial Nuance software package) were applied to spectral images of Giemsa-stained guinea-pig BAL taken at 20× magnification (after conversion to optical density). Using spectral features alone, polymorphonuclear cells, eosinophils, and monocytes were robustly separated, as shown in Figure 1. Although this image is just an RGB representation, it would be simple to complete the analysis and identify and count the blood elements, since the differences between cell types is greatly enhanced after the
spectral processing. Note another useful feature of this approach: red blood cells (RBCs) can be spectrally “lysed” and caused to disappear from the output image.

**Evaluation**

After segmentation may come evaluation in which spectral and/or spatial properties of the segmented objects are used for determining important qualities, for example, identification, diagnosis, prognosis, and prediction of response to specific therapies (“theragnosis”). Examples of methods for the automated analysis of pathology specimens include Weyn et al. (29) who used morphometry (measurement of form, shape and nuclear dimensions), texture and syntactic structure analysis (quantifying the spatial arrangements of the structures in the tissue) for the detection of malignant mesothelioma; Mutter et al. (30) who used computerized morphometry in the diagnosis of endometrial precancer; and Van de Wouwer et al. (31) who used wavelets as chromatin texture descriptors for the identification of invasive breast cancer. Successful correlation of chromatin texture (karyometry) with malignant potential has been demonstrated in bladder and skin lesions (32).

Researchers exploring such image analysis tools have employed up to 400 different (but not necessarily independent) morphological features. The existing techniques for automated detection and classification generally use some predetermined image processing algorithms to extract a small subset of these features, which are then used as input to some classification scheme. These techniques are usually targeted at specific abnormalities, each with its own unique information requirements and those features of use in classification have been determined only after considerable time and effort. The problem with many of these approaches, and the reason why almost none of them have been transferred to practice, is that they are both problem-specific and center-specific, that is, the details of how specimens are prepared and images are acquired can affect transferability of the algorithms from one research group to another. In addition, the techniques may require a great deal of manual preprocessing.

**Statistical Machine Learning**

A useful approach is to apply machine-learning techniques in which the desired outcomes are “taught” by example. Using neural nets, genetic algorithms or support vector machines (33,34), for example, sets of features can be randomly combined and their classification fitness automatically evaluated and optimized. A particularly useful aspect of correctly designed machine-learning tools is that spectral and spatial information can be exploited more or less simultaneously. The same guinea-pig BAL specimen was shown analyzed using a genetic-algorithm-based tool that uses a two-stage process: (i) the feature selection part, consisting of a graph- (or tree)-like algorithm constructed from simple, fundamental image-processing steps; and (ii) the final-stage classification. An example of this approach combining spectral imaging and machine learning software originally developed for remote sensing applications (Genie) was recently applied to cytometry data (35).

To perform the analysis shown in Figure 2, a few monocytes, eosinophils and PMNs were highlighted using a drawing tool and identified as separate classes (with pseudo-colors green, red, and blue, respectively). At the same time, the RBCs and white background were also identified as a class (colored white). After a brief training period, the tool “learned” to separate the blood elements with a high degree of accuracy (although close viewing

![Image](https://example.com/image.png)
will reveal a few misidentified cells). This task was purely segmentation—no information was generated about the properties of the different cell classes. Such segmented images could either be analyzed further using standard image analysis techniques or, alternatively, additional machine-learning could be applied to a series of blood samples to identify characteristics of different diseases or metabolic states, for example.

Colon Cancer

A slightly more challenging example is the training of such a tool to distinguish colon cancer from normal colonic tissue starting with a spectral dataset, which was acquired at 10×-magnification with images taken from 450 to 690 nm at 10-nm intervals. The left panel (Fig. 3) shows the RGB representation of the spectral image containing “normal” colon and colonic cancer, separated more or less along the top-left to bottom-right diagonal. The red and green areas are the training regions used for machine learning. The right panel shows the classification results. This is a relatively trivial example in that the training set and the test set are similar (regions from the same image), although work not shown here suggests that results from out-of-training set data can be as robust. One interesting observation is that one of the ostensibly “normal” glands (to the right of the asterisk) classifies as part normal and part cancer. This may reflect the “field-effect” phenomenon in that...
normal structures bordering cancers can have abnormal morphological or molecular phenotypes (36).

**Kidney Inflammation**

A common problem in pathology is the detection and quantitation of inflammatory processes—useful for the assessment of the activity of renal or hepatic disease, for example, or the presence of rejection in transplanted organs. While this determination is almost always performed by visual examination, this can be far from quantitative or reproducible. Automating this evaluation for application to samples stained with standard H&E dyes (rather than expensive immunostains) could be helpful. In kidney, especially, this can be a difficult task, since, at low power, small, dark blue inflammatory cells closely resemble apoptotic tubular cells or glomerular constituents. An example shown here indicates how machine-learning can readily generate algorithms that discriminate between these similar elements. The training sets consisted only of regions of inflammation (blue) vs. tubules (green) and glomeruli, connective tissue and blood vessels (red). In Figure 4, the top panel shows classification results using training sets present within that image; the lower panel shows how the same classifier segmented another field from the same slide. The top panel, bottom-left corner, contains degenerating tubules with necrotic, hyperchromatic nuclei, which, although small and dark, are nevertheless not confused with areas of inflammation.

The suggestion from these and other preliminary investigations is that many image segmentation problems can be readily approached using spectral image datasets and machine-learning-based spectral and spatial tools.

**Chromogenic Staining and Spectral Unmixing**

Although fluorescence is still the procedure of choice for detecting multiple molecular entities in histologic and
cytologic preparations, it is not used widely in pathology or cytology except for certain very specific applications, such as FISH for Her2-neu amplification. Instead, immunohistochemistry, a brightfield, chromogen-based detection approach, is used. The problem is that the chromogens are difficult to resolve if they overlap spatially, either by eye or by using conventional color sensors. Spectral imaging allows users to employ three or more chromogens as molecularly specific contrast agents—even in the presence of a counterstain and even when they are targeted to similar or identical anatomic structures. Figure 5 shows a triple-immunostained tonsil section, with granulocytes visualized with DAB (brown), and two lymphocyte populations identified with anti-CD43 (BCIP-NBT, blue) and anti-CD20 (Fast Red). All cellular nuclei are also counterstained with hematoxylin, which here stains them with a light robin’s-egg blue. The linear unmixing algorithms (37) that were used to perform the spectral unmixing calculation are only valid in situations in which pure spectra combine linearly, as is the case with fluorescence images. Transmission images, such as what was originally acquired for this immunohistochemical (IHC) brightfield sample, must be converted to optical density first, in accordance with the Beer-Lambert law (38). Thus, the unmixed images shown in the bottom half of this figure reveal the separated signals as bright pixels against a black background. The signals corresponding to the presence of the three immunostains are combined in the bottom right figure. The nuclear counterstain was also unmixed into its own channel (not shown) and was omitted from the composite image for clarity.

The spectra used for the unmixing were derived directly from this spectral dataset. However, these spectra can be stored in “libraries” and applied to other datasets in automated fashion. Thus, this multiplexing approach is suitable for automated analysis of high-density tissue arrays, since the acquisition and analysis can take place with little or no operator attention once the imaging and analysis parameters are set.

DISCUSSION

Cytomics = Morphometry in a Molecular World?

Until recently, all diagnostic and prognostic evaluations depended on clinical staging and microscopic evaluation of parameters such as size, shape, color, and organization of stained biopsied tissues and cells—aspects detected and integrated, albeit subjectively and to some degree unconsciously, by the pathologist. This approach still represents the available “gold standard.” Advances in our understanding of cancer biology and the advent of new detection techniques have led to an emphasis on the molecular genotype and phenotype, somewhat to the exclusion of the morphological. While molecular techniques examining genes and gene products in isolation have provided tremendous insights, the intellectual pendulum is now swinging back to put more emphasis on the exploration of multi-state molecular systems and networks. A further integration of these molecular systems occurs in intact cells and is manifested by the structure and makeup of (potentially visible) cellular macromolecular components. With each step up the integration ladder, individual details can be obscured, but systems behavior may be better captured (39). Cytomics may thus be considered to be a combination of technologies and data sources whose sum is designed to exceed the individual contribution of the parts.

One of these parts can be spectral imaging, which may serve to advance this field in both central and supportive ways. In one role, it can capture an accurate measure of the tintorial properties of samples stained with conventional (or possibly unconventional) molecularly unspecific dyes. It has been shown, by work at CRI and elsewhere, that even H&E stains have enough spectral content as to enable the distinction between certain cancers.
and their normal counterparts, based on spectral signatures alone. Shown in this communication is the utility of spectral imaging to discriminate blood cell types by spectral analysis of a Giemsa stain, when by visual examination the pink-to-purple gamut appears too limited to support unambiguous discrimination. Down the road, spectral content—if substantial issues of staining variability and sample preparation can be addressed—may prove clinically useful.

Spectral imaging combined with more versatile tools that combine spectral and spatial information seems quite likely to be a better route than using either alone for extracting relevant information from histology and cytology specimens. The examples in this communication illustrate the flexibility of such approaches, which can be applied to high-magnification cytology images, low-power architectural views, or a combination of both. In this case, the spectral-spatial learning engine can either provide the final output or serve as a power segmentation tool to deliver relevant objects to more focused algorithms designed for specific purposes.

“High-Context” Imaging

Finally, spectral imaging provides a convenient connection to molecular sources of information, by enabling the use of standard IHC (and brightfield-based in situ hybridization) assays in a moderately multiplexed manner. Thus, 2–4 or more antigens or other molecular species can be monitored in cells whose morphological properties can also be assessed. For higher degrees of multiplexing, a switch to fluorescence and novel reagents such as quantum dots, in combination with spectral imaging techniques, can provide in excess of 5 (the upper limit is not known) simultaneous molecular signals (data not shown).

To return to the themes mentioned at the beginning, true cytomtics would then be able to integrate spectral, spatial, and molecular data from imaged samples, with individual clinical as well as generalized statistical metadata to provide the maximum possible yield of useful information.

ACKNOWLEDGMENTS

Samples were kindly provided by David Leaffer, Michael Nalesnik, and Thomas Grogan. The manuscript benefited from review by Alton Floyd and Clifford Hoyt.

LITERATURE CITED

32. Montironi R, ScarPELLI M, Mazzucchelli R, Hamilton PH, Thompson D, Ranger-Moore J, Bartels PH. Subvisual changes are detected...


