

MULTI-MODAL IMAGING OF HISTOLOGICAL TISSUE SECTIONS

Ali Can, Musodiq Bello, Harvey E. Cline, Xiaodong Tao, Fiona Ginty, Anup Sood, Michael Gerdes, Michael Montalto

{can, bello, cline, taox, ginty, sood, gerdes, montalto}@research.ge.com
GE Global Research Center, Niskayuna, NY, 12309.

ABSTRACT

Two common imaging modalities for histological sections are brightfield and fluorescence microscopy imaging. Hematoxylin-Eosin (H&E) based brightfield microscopy has been the traditional imaging technique for imaging morphology, while an epi-fluorescent microscope is used for immunofluorescent staining of specific proteins or fluorescent in situ hybridization (FISH) for genetic based analysis of DNA. Simultaneous imaging of both microscopy modalities has been difficult due to optical and chemical effects of the H&E dyes. We present a novel sequential imaging and registration technique that enables brightfield and fluorescent imaging on the same tissue section, hence combining the traditional anatomic pathology with the newly emerging field of molecular pathology. First the tissue is labeled with fluorescent biomarkers, and imaged through a fluorescence microscope, and then the tissue is re-labeled with H&E dyes, and imaged again with traditional brightfield. Our robust registration algorithms achieve 99.8% registration success rate on tissue micro array (TMA) sections.

Index Terms— Image Registration, Brightfield Microscopy, Fluorescent Microscopy

1. INTRODUCTION

Hematoxylin-Eosin (H&E) staining has been used by pathologists for over a hundred years [1]. Hematoxylin stains cell nuclei blue, while Eosin, as a counter-stain, stains cytoplasm and connective tissue pink. Due to the long history of H&E, abundance of well-established methods for use, and the tremendous amount of data and publications, there is a strong belief among many pathologists that H&E will continue to be the common practice in the next 50 years [1]. However, new technologies, such as analysis of molecular markers that provide functional and biological information, have proven to be valuable for diagnosis, prognosis, and survival rate analysis [2]. A general overview of molecular labeling, high throughput imaging, and pattern recognition techniques is presented by Price et al. [3].

The H&E staining has been favored due to its low cost, fast preparation and permanence of stain on tissue, ease of image

acquisition, and the extensive knowledge and training behind it. On the other hand, molecular biomarkers are increasingly being used and can provide protein-associated information of the tissue that is not visible by H&E techniques. Recently, immunohistochemistry (IHC) based image analysis algorithms have been presented to quantify the localization of proteins in the tissue [2]. Due to the common use of H&E for diagnosis and staging, most image analysis and automated quantification techniques presented in the literature use H&E data as well [4-6].

With the current imaging techniques, it is not possible to simultaneously image H&E dyes and immunofluorescent molecular biomarkers due to fluorescent characteristics of the H&E dyes, and due to chemical interactions of H&E dyes with the fluorescently labeled antibodies. We present sequential imaging and registration techniques that enable different modalities presented digitally from the same histological tissue section. Additionally, sequential imaging and registration enables imaging of multiple immunofluorescent stains acquired in multiple steps rather than conventional simultaneous multiplexing techniques. This allows an order of magnitude increase in the number of molecular markers to be imaged for the same tissue section. Our techniques make it possible to explore un-examined relationships between morphology, subcellular spatial distribution of proteins and protein-protein interactions.

2. METHODS

The breast cancer tissue sections are first labeled with fluorescent markers, and imaged with a fluorescent microscope with the right excitation energy source tuned to fluorochromes, and proper filters to collect the emitted light. The non-uniform illumination variations are corrected by parametric surface fitting techniques [7]. Multiple fluorescent dyes can either be simultaneously imaged, or added in a sequential process where between the steps the dyes are bleached and re-added with different antibodies.

Next, the tissue section is stained using traditional H&E dyes, and placed back under the microscope and the images are acquired in the brightfield mode of the same region of interest. The lighting pattern and color balance of the images are then corrected similarly. Due to slide misplacement and

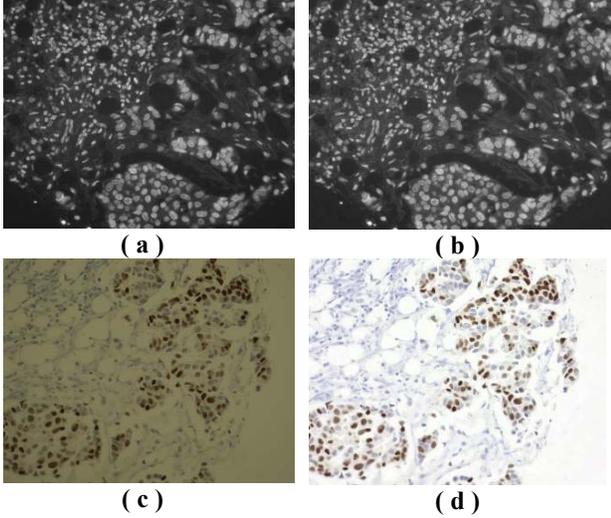


Figure 1. Fluorescent images showing DAPI staining of breast carcinoma tissue; **(a)** before correction, **(b)** after illumination correction. Illumination correction can also be used with brightfield DAB stained images, **(c)** before correction, **(d)** after illumination correction.

stage errors between each sequential step there is a variable degree of misalignment between images. We use mutual information based registration techniques to align all the images and define all the modalities in the same coordinate system. The registration failures are automatically detected and reinitialized. The quality of the overall process is controlled to detect tissue folding and loss.

2.1 Illumination Correction

We first correct the illumination pattern of the microscope by estimating it from a series of images. The observed image, $I(x, y)$, can be modeled as a product of the excitation pattern, $E(x, y)$, and the emission pattern, $M(x, y)$. While the emission pattern captures the tissue dependent staining, the excitation pattern captures the illumination. In the logarithm domain, the observed image can be modeled as a linear process;

$$\log(I(x, y)) = \log(E(x, y)) + \log(M(x, y)). \quad (1)$$

From a set of N images, let $I_n(x, y)$ denote an ordered set of pixels. In other words, the pixels are sorted for any given (x, y) location such that

$$I_1(x, y) \leq I_2(x, y) \leq \dots \leq I_n(x, y) \leq \dots \leq I_N(x, y) \quad (2)$$

Assuming that a certain percentage, p , of the image is formed from stained tissue (non-zero background), then a trimmed average of the brightest pixels can be used to estimate the excitation pattern:

$$E'_{AVE}(x, y) = \frac{1}{N - K + 1} \sum_{n=K}^N \log(I_n(x, y)), \quad (3)$$

where K is set to an integer closest to $N(1 - p) + 1$. In our experiments, we set p to 0.1 (10%). In the above equation,

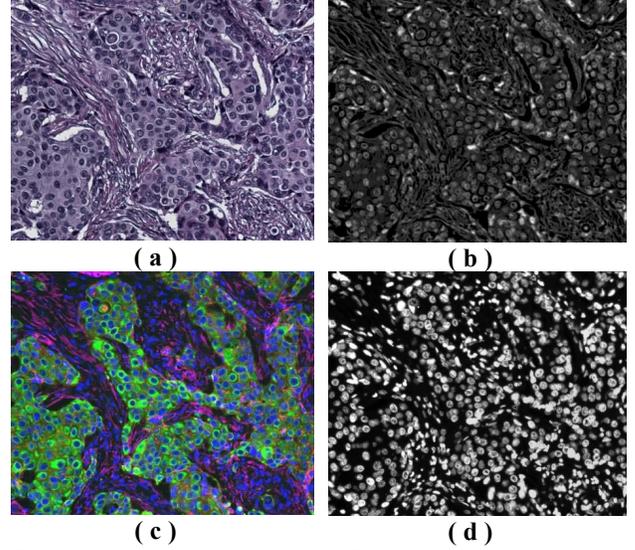


Figure 2. **(a)** A brightfield H&E image of breast cancer tissue section. **(b)** Computed nuclei component of the H&E image. **(c)** Multi-channel fluorescent microscopy image. **(d)** The nuclei component of the fluorescent image registered to the nuclei component of the H&E image.

the average emission pattern of the tissue is assumed to be uniform. Since the images are recovered upto a scale factor, we can drop the constant term introduced by the uniform emission pattern. To overcome the limited sampling size at each pixel location, we approximate the log of the excitation pattern for the entire image with polynomials;

$$E'_{AVE}(x, y) = \sum_{0 \leq i, j \leq p, i+j \leq p} a_{ij} x^i y^j. \quad (4)$$

The parameters a_{ij} are solved by minimizing the mean squared error [7]. The surface generated by the polynomial coefficients are then used to correct the images. Figures 1a&b show a fluorescent image before and after correction, respectively. Each color channel of the brightfield images are corrected similarly using Equations 1-4, with the only difference that the pixel ordering is reversed in Equation 2. Processing each color channel independently corrects the color temperature of the light source as well (See Figures 1c&d).

2.2 Registration of Brightfield and Fluorescent Images

We design the experiments such that at each step of the sequential staining across modalities, we acquire images of the nuclei. For brightfield images hematoxylin stains the nuclei blue (Figure 2a), and for fluorescent images we use DAPI to stain nuclei (blue channel in Figure 2c). The first nuclei image is set as the reference image and each of the subsequent nuclei images are registered to the reference. Once the transformation parameters are estimated, then all the channels at a sequential step are mapped onto the reference coordinate system (Figure 3).

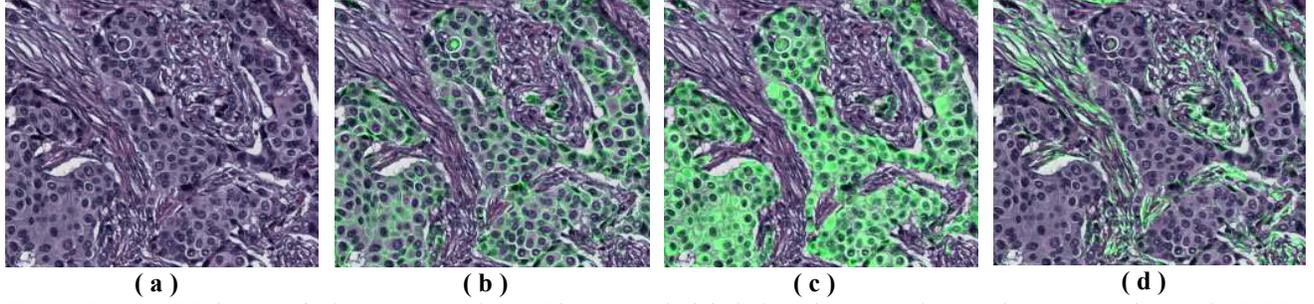


Figure 3: (a) H&E image of a breast tumor tissue. Fluorescently labeled markers superimposed as green color on the H&E image, (b) β -catenin, (c) pan-keratin, and (d) smooth muscle α -actin, markers.

Given two set of nuclei images, one being the reference image from the first step, $I_i^{(1)}(x, y)$, and the second being from the subsequent step (either the nuclei channel from an H&E step or another fluorescent acquisition step for multiplexing), $I_i^{(k)}(x, y)$, we find a transformation $\mathbf{T}_i^{k,1}$, such that the image similarity measure, S , between $I_i^{(1)}(x, y)$ and $I_i^{(k)}(\mathbf{T}_i^{k,1}(x, y))$ is maximized,

$$\arg \max_{\mathbf{T}_i^{k,1}} S(I_i^{(1)}(x, y), I_i^{(k)}(\mathbf{T}_i^{k,1}(x, y))). \quad (5)$$

There is a great number of cost functions, such as mean-square-error, cross correlation, Kullback-Liebler distance, gradient difference metric, mutual information, etc [8]. Due to its robustness in registering multi modality images in this work, we use a rigid transformation and a mutual information based image similarity measure [9]:

$$S(\mathbf{T}_i^{k,1}) = -\sum_t \sum_\kappa p(t, \kappa | \mathbf{T}_i^{k,1}) \log \frac{p(t, \kappa | \mathbf{T}_i^{k,1})}{p_M(t | \mathbf{T}_i^{k,1}) p_F(\kappa)} \quad (6)$$

where p , p_M , and p_F are the joint, marginal moving, and marginal fixed probability distribution of the image intensities; $\mathbf{T}_i^{k,1}$ is the parameter vector of the rigid transform; t and κ are the intensity values in the respective images. In order to improve the robustness of the algorithm, we use a multi-resolution strategy to find the transform that aligns the two images. The details of the minimization of this cost function is described in [10, 11].

While the fluorescent images include dedicated nuclei channels (such as DAPI), the nuclei images from the H&E images are computed using the red, green and blue channels as follows;

$$I_N^{HE}(x, y) = c \cdot \left(I_{BLUE}^{HE}(x, y) / \sqrt{I_{RED}^{HE}(x, y) \cdot I_{GREEN}^{HE}(x, y)} \right)^\gamma \quad (7)$$

where c and γ are tuning parameters for contrast and gamma correction, respectively. Since the hematoxylin stains the nuclei blue, the highest contrast is achieved when the blue channel is normalized by the geometric mean of the red and green channels. Figures 2a and 2b show an H&E image and its estimated nuclei component computed by the above equation, respectively. A multi-channel fluorescent image of the same tissue stained with molecular biomarkers

with red representing a membrane related marker (β -catenin), blue representing a nuclei related stain (DAPI), green representing keratin, and magenta representing smooth muscle α -actin is shown in Figure 2c. The registered DAPI image in the H&E coordinate system is shown in Figure 2d. The registration parameters estimated from the DAPI and H&E images are then used to map all the other fluorescent biomarkers at that sequential step into the H&E coordinate system (Figure 3). The nuclear stain DAPI which is used for registration is not shown in Figure 3.

2.3. Registration Failure Detection and Initialization

When multi-modality images obtained for a given tissue sample are co-registered, the registration needs to be validated prior to further analysis. For tissue micro array (TMA) applications, we developed an automated system to detect registration failures accurately, which then automatically re-initializes the registration process for such cases. The registration described earlier is performed on individual images in local image coordinate space to obtain rotation and translation parameters of a rigid transformation. The images can be defined in a metric coordinate system, using an image to microscopy stage calibration transformation, \mathbf{T}_i^Δ which incorporates scaling transformations and stage coordinates of each image recorded by the microscope. In the metric coordinate system the slide has only one global transformation \mathbf{T}_G , which can be estimated using the individual transformation of all the images,

$$\arg \min_{\mathbf{T}_G} \sum_i \rho \left(\sum_{x,y} (\mathbf{T}_i^\Delta(x, y) - \mathbf{T}_G(\mathbf{T}_i^\Delta(\mathbf{T}_i^{k,1}(x, y))))^2 \right) \quad (8)$$

where ρ is a non-quadratic error function that is robust to our registration failures. The quadratic term measures the warping of the transformation in the metric coordinate system. The global transformation is solved by first initializing by random sampling and Least Median of Squares estimation [12] followed by M-estimators [13]. The equation above is approximate because it does not incorporate stage errors; nevertheless it is useful to detect failures by checking high residual values, and reinitialize these failures.

2.4. Quality Assurance

We developed automated quality assurance tools to detect tissue folding, tissue loss, and registration failures for TMA sections. Although the success rate of the registration algorithm is 99.8%, possible failures are detected automatically and presented to the user for confirmation. At each step of sequential staining, the following metrics are computed and any possible defect is presented to the user for confirmation;

- i) Different image-to-image similarity metrics to detect tissue folding, or tissue loss,
- ii) Nuclei count to differentiate tissue loss from other types of defects,
- iii) Registration accuracy of each tissue spot in a TMA computed using the methods described in Section 2.3.

The results are stored in a spreadsheet including details about the registration status, tissue quality, and tissue viability. The spreadsheet can be modified easily if a user makes an assessment different from the tool. The quality assurance results are also illustrated graphically with annotated thumbnail images of each tissue core arranged according to the location of the cores in the actual TMA slide. The tool also allows the user to visualize a summarized representation of the tissue quality of all tissue cores from multiple sequential steps all at once.

3. RESULTS AND CONCLUSION

We present techniques to digitally merge different histological imaging modalities. This is a step forward in integrating molecular pathology with classical anatomical pathology. Superimposing the molecular biomarker information on the H&E information provides a qualitative tool for the pathologist to view both modalities on the same tissue. This will provide great diagnostic value since the pathways can now be easily superimposed on a standard H&E slide. For example, all the molecular markers a pathologist is interested in quantifying such as the biomarkers shown in Figure 3b-d can be superimposed on the traditional H&E view in Figure 3a. The pathologists can look at the multi-channel image in the H&E mode, seeing the traditional H&E image, but can also superimpose the overlaying molecular information by clicking a button on the computer screen. In this paper we only show a set of sample markers to prove the concept. Any diagnostics markers such as ER, PR, Her2 that are standard for breast cancer diagnosis can be superimposed on the same tissue, where pathologist can do morphological analysis as well.

We demonstrate our method using H&E dyes for brightfield staining. As long as there are common information channels across the modalities, our methods can be applied for other staining protocols. We apply our techniques to TMAs, and quantify the registration success rate as over 99.8%. Our

local-global-local estimation framework detected 96.6% of errors, and corrected 86% of these errors. We observed that the automated tools for TMA applications increases the quality assurance productivity up to fifteen times, compared to manual validation techniques.

This application is not limited to the field of histopathology. Other medical and bioscience applications can benefit from the information provided with extended multi-channels. The current trend in biology is towards adding as many markers as possible to the specimen to visualize the pathways simultaneously. Our approach provides a flexible framework where the markers can be imaged sequentially without being limited by optical, chemical, and biological interactions. Incorporating functionality with morphology, our methods enable better diagnosis, and design of large clinical studies using TMAs.

REFERENCES

- [1] H. Fox, "Is H&E morphology coming to an end?" *Journal of Clinical Pathology*, vol. 53, pp. 38-40, 2000.
- [2] M. A. Rubin, et. al. "Quantitative Determination of Expression of the Prostate Cancer Protein α -Methylacetyl-CoA Racemase Using Automated Quantitative Analysis (AQUA)," *American Journal of Pathology*, vol. 164, pp. 831-840, 2004.
- [3] J. H. Price, et. al. "Advances in Molecular Labeling, High Throughput Imaging and Machine Intelligence Portend Powerful Functional Cellular Biochemistry Tools," *Journal of Cellular Biochemistry Supplement*, vol. 39, pp. 194-210, 2002.
- [4] P. H. Bartels, T. Gahm, and D. Thompson, "Automated Microscopy in Diagnostic Histopathology: From Image Processing to Automated Reasoning," *Int. J. IS&T*, vol. 8, pp. 214-223, 1997.
- [5] W. A. Christens-Barry and A. W. Partin, "Quantitative Grading of Tissue and Nuclei in Prostate Cancer for Prognosis Prediction," *Johns Hopkins Apl. Tech.Digest*, vol. 18, pp. 226-233, 1997.
- [6] C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," RPI, Department of Computer Science, Technical Report, TR-05-09, 2005, pp. 1-16.
- [7] H. Narasimha-Iyer, A. Can, B. Roysam, V. Stewart, H. L. Tanenbaum, A. Majerovics, and H. Singh, "Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy," *IEEE TBE*, vol. 53, pp. 1084-98, 2006.
- [8] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Comp.*, vol. 21, pp. 977-1000, 2003.
- [9] P. Viola and W. M. Wells Iii, "Alignment by Maximization of Mutual Information," *IJCV*, vol. 24, pp. 137-154, 1997.
- [10] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE TMI*, vol. 22, pp. 120-128, 2003.
- [11] T. Philippe and U. Michael, "Optimization of Mutual Information for Multiresolution Image Registration," *IEEE Trans. on Image Proc.*, vol. 9, pp. 2083-2099, 2000.
- [12] P. J. Rousseeuw, "Least Median of Squares Regression," *Journal of the American Statistical Association*, vol. 79, pp. 871-880, 1984.
- [13] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics-Theory and Methods*, vol. 6, pp. 813-827, 1977.