Appearance Normalization of Histology Slides

Marc Niethammer, David Borland, J. S. Marron,
John Woosley, and Nancy E. Thomas

UNC Chapel Hill

Abstract. This paper presents a method for automatic color and intensity normalization of digitized histology slides stained with two different agents. In comparison to previous approaches, prior information on the stain vectors is used in the estimation process, resulting in improved stability of the estimates. Due to the prevalence of hematoxylin and eosin staining for histology slides, the proposed method has significant practical utility. In particular, it can be used as a first step to standardize appearances across slides, that is very effective at countering effects due to differing stain amounts and protocols, and to slide fading. The approach is validated using synthetic experiments and 13 real datasets.

1 Introduction

To highlight distinct structures in microscopy images of tissue samples, tissue staining is commonly used. Frequently two stains, such as hematoxylin and eosin (H&E), are applied for purposes such as discriminating cell nuclei and cytoplasm. Variations in staining results can be minimized by using fully standardized staining protocols. However, in practice precise control over stain color and staining intensity is typically not possible: stains may fade over time, stain colors may differ slightly, slides may have been imaged on different microscopes, or data that has already been digitized may need to be analyzed.

Standard (non-fluorescent) stains absorb light. Local stain concentrations and stain colors determine the appearance of an illuminated slide sample under the microscope. If no stain is present and the underlying tissue does not absorb a significant amount of light, the corresponding pixel will appear bright. Areas where the stains accumulate will appear darker. Absorption is wavelength dependent, and a particular stain can be characterized by its absorption coefficients, forming a vector (the *stain vector*) of dimension equal to the number of wavelengths in the sensor used for imaging (three for a standard RGB color camera, as in this work). Given the stain vectors, an image can be decomposed into individual stain components via color deconvolution [1], and stains can subsequently be intensity adjusted. This paper proposes a method for automatic stain vector estimation and slide appearance normalization (color and intensity), which can in turn improve quantitative, computerized analysis.

Previous approaches to extract stain vectors include (1) manual region of interest definition, (2) methods relying on non-negative matrix factorizations [2], and (3) plane fitting in the optical density domain [3]. The approach presented in this paper is most closely related to [3]. Novel contributions include: (1) a

rigorous theory for the color model used, (2) the introduction of prior information for the stain vector taking into account varying amounts of stain (such as that encountered in the case of sparsely distributed nuclei on large amounts of stained background tissue), (3) an alternating optimization method and its connection to a sub-problem from trust region optimization, (4) a novel twist on Otsu thresholding [4] which also includes prior information, (5) and quantitative validation on synthetic and real datasets.

Sec. 2 introduces the stain model and formalizes the planar assumption for the stain vectors of [3]. Sec. 3 discusses the plane fitting method with prior information. The clustering approach is presented in Sec. 4. Sec. 5 presents validation results, and Sec. 6 concludes.

2   Stain Vector Model

According to the Beer-Lambert law, the transmission of light through a material can be modeled as $I = I_0 e^{-\alpha c x}$, where I_0 is the intensity of the incident light and I the intensity of the light after passing through the medium; α is the absorption coefficient, c the concentration of the absorbing substance, and x the distance traveled through the medium. The absorbance, or optical density (OD), is: $OD = \alpha c x = -\log\left(\frac{I}{I_0}\right)$. The proposed method assumes that α and x are constant for a specimen and a given stain, but that a stain's concentration c may change. For a multi-spectral image the relation is:

$$\boldsymbol{I} = \boldsymbol{I_0} \odot e^{-\boldsymbol{\alpha} c x}, \quad \boldsymbol{OD} = -\log(\boldsymbol{I} \oslash \boldsymbol{I}_0) = \boldsymbol{\alpha} c x,$$

where the absorption coefficient is color dependent (α_i), $\odot$ denotes the Hadamard product (the element-wise vector product), and $\oslash$ the Hadamard division. Note that low intensities will correspond to large optical densities and high intensities (e.g. white areas) will correspond to low optical densities. Each stain has a characteristic vector $\boldsymbol{\alpha}$ of absorption coefficients. Given a specific distance x, the optical density vector $\boldsymbol{OD}$ is linearly related to the absorption coefficient vector, where the proportionality constant is given by the stain concentration: $\boldsymbol{OD} = \boldsymbol{\alpha} x c$. Applying the Beer-Lambert law to the two-stain color-image case (e.g., eosin and hematoxylin) yields $\boldsymbol{I} = \boldsymbol{I_0} \odot e^{-(\boldsymbol{\alpha}^1 c^1 x^1 + \boldsymbol{\alpha}^2 c^2 x^2)}$, where superscripts denote values for the two distinct stains. Converting to optical density results in

$$-\log(\boldsymbol{I} \oslash \boldsymbol{I}_0) = \boldsymbol{\alpha}^1 c^1 x^1 + \boldsymbol{\alpha}^2 c^2 x^2,$$

which shows that the obtainable intensity vectors $\boldsymbol{I}$ for a given illumination $\boldsymbol{I}_0$ lie in the plane spanned by the absorption coefficient vectors $\boldsymbol{\alpha}^i$ (*the stain vectors*) in the optical density domain. Since $c^i \geq 0$ and $x^i \geq 0$ and the $\boldsymbol{\alpha}^i$ are linearly independent, any color (which can fully be explained by the imaging model) needs to lie within the convex cone $\mathcal{C} = \{\boldsymbol{x} | \boldsymbol{x} = q_1 \boldsymbol{\alpha}^1 + q_2 \boldsymbol{\alpha}^2, \ q_1, q_2 \geq 0\}$. Further, normalizing all possible optical density vectors, the resulting points must all lie within $\mathcal{C}_N = \{\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{x}} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \ \boldsymbol{x} \in \overset{\circ}{\mathcal{C}}\}$, where $\overset{\circ}{\mathcal{C}}$ denotes $\mathcal{C} \setminus \boldsymbol{0}$ and geometrically, $\mathcal{C}_N = \mathcal{S}^{n-1} \cap \mathcal{C}$, the intersection between $\mathcal{S}^{n-1}$ (the n-dimensional unit sphere – $n = 3$ for an RGB camera) and $\mathcal{C}$, which is a sector of a great circle.

3 Plane fitting with a plane prior

The optical density cone of Sec. 2 is a subset of a plane $\mathcal{P}$ passing through the origin $\mathcal{P} = \{\boldsymbol{x} : \boldsymbol{n}^T\boldsymbol{x} = 0\}$, where $\boldsymbol{n}$ is the plane's unit normal. The signed distance of any point to the plane can be computed as $d(\boldsymbol{x}, \mathcal{P}) = \boldsymbol{n}^T\boldsymbol{x}$. The plane closest to all points (wrt. their squared distances) minimizes

$$E(\boldsymbol{n}) = \boldsymbol{n}^T \left(\sum_i \boldsymbol{x}_i \boldsymbol{x}_i^T \right) \boldsymbol{n} = \boldsymbol{n}^T S \boldsymbol{n}, \quad \text{s.t.} \quad \|\boldsymbol{n}\| = 1.$$

Since S is by construction symmetric positive-semi-definite, $\boldsymbol{n}$ is the eigenvector of the smallest eigenvalue of S. Such an unconstrained estimation was proposed in [3]. Estimation results are reliable when a sufficient amount of both stains is present in a given slide sample. But when this assumption does not hold, which happens for example when stained nuclei are sparse on a given sample, or if artifactual stains (e.g., brown from melanin) is present, adding prior information on the color direction is very important to assure good performance of the estimator. The corresponding maximum-a-posteriori energy (with prior) is

$$E(\boldsymbol{n}) = \left(\frac{1}{2\sigma^2} \sum_i d^2(\boldsymbol{x}_i, \mathcal{P}) \right) + \frac{1}{2(\sigma^0)^2} \|\boldsymbol{n} - \boldsymbol{n}_p\|^2, \quad \text{s.t.} \quad \|\boldsymbol{n}\| = 1, \quad (1)$$

where σ and σ^0 are the standard deviations for the measured points (assumed to be independent) and the prior, respectively (assumed to be Gaussian).

An approximately planar distribution of points will only be observed if large numbers of measurement points *for both* stains are contained in the set of data points. This is often not the case. Therefore a weighting of the data points assigned to either one of the stain directions can be very useful. Assume the partitioning into two classes is given by a clustering method (as described in Sec. 4). Then the energy (up to constants) can be decomposed as

$$E(\boldsymbol{n}) = \left[\sum_{j=1}^{2} \left(\sum_{i \in P_j}^{n_j} \frac{d^2(\boldsymbol{x}_i, \mathcal{P})}{2\sigma^2} + \frac{\alpha}{2(\sigma^0)^2} \|\boldsymbol{n} - \boldsymbol{n}_p\|^2 \right) w_j \right],$$

where w_i are appropriately chosen weights, P_i indicate the partitions, and n_i are the number of points in the respective partitions. A reweighting should take into account the presence or absence of a sufficient number of datapoints in either of the clusters. Further, a simplification to the form of Eq. 1 is desirable in case both clusters are of equal size. For these properties to be fulfilled the following conditions should hold:

$$\alpha w_1 + \alpha w_2 = 1, \quad w_1 \frac{N}{2} + w_2 \frac{N}{2} = N, \quad \frac{\gamma}{n_1} = w_1, \quad \frac{\gamma}{n_2} = w_2.$$

These conditions are fulfilled for $\alpha = \frac{1}{2}$, $w_1 = \frac{2n_2}{n}$, and $w_2 = \frac{2n_1}{n}$. After some algebra and rescaling (by $2\sigma^2 N/(4n_1 n_2)$) the energy becomes

$$E(\boldsymbol{n}) = \boldsymbol{n}^T \left(\frac{1}{2n_1} \sum_{i=1}^{n_1} \boldsymbol{x}_i \boldsymbol{x}_i^T + \frac{1}{2n_2} \sum_{i=1}^{n_2} \boldsymbol{x}_i \boldsymbol{x}_i^T \right) \boldsymbol{n} + \frac{N}{4n_1 n_2} \frac{\sigma^2}{(\sigma^0)^2} \|\boldsymbol{n} - \boldsymbol{n}_p\|^2,$$

which is of the same form as Eq. 1: a weighted covariance matrix and a cluster-dependent weighting of the prior term. The optimization problem is closely related to the sub-problem of finding a minimum over a boundary in trust-region optimization [5]. The overall solution alternates between a solution of the optimization problem and a reclustering (see Sec. 4) of the data points, to convergence.

Theorem 1 (Optimal plane fit). *Given the optimization problem*

$$\min_{\boldsymbol{n}} \ \boldsymbol{n}^T S \boldsymbol{n} + \frac{1}{\sigma^2} \|\boldsymbol{n} - \boldsymbol{n}_p\|^2, \quad \|\boldsymbol{n}\| = 1,$$

where S is a real positive-semi-definite matrix, at optimality

$$(\overline{S} + \lambda I)\boldsymbol{n} = \frac{1}{\sigma^2}\boldsymbol{n}_p, \quad \|\boldsymbol{n}\| = 1, \quad \overline{S} = S + \frac{1}{\sigma^2}I, \quad (\overline{S} + \lambda I) \succeq 0.$$

Proof. Follows from the proof of the trust-region optimality conditions [5].

If $\overline{S} + \lambda I$ is invertible, the Lagrangian multiplier can be obtained by solving

$$\sigma^4 \prod_{i=1}^{3}(\lambda_i + \lambda)^2 = \sum_{i=1}^{3}(\tilde{\boldsymbol{n}}_p)_i^2 \prod_{j=1, j\neq i}^{3}(\lambda_j + \lambda)^2, \quad \lambda \in (-\lambda_1, \infty),$$

where λ_1 is the smallest eigenvalue of $\overline{S}$, and $\tilde{\boldsymbol{n}}_p = Q^T \boldsymbol{n}_p$, with Q the matrix of eigenvectors (as columns) of $\overline{S}$. If the matrix is not invertible, the point distribution exhibits symmetries which makes the solution non-unique.

Typically, the prior plane normal $\boldsymbol{n}_p$ will not be directly available, but will be specified through a set of two given stain vectors $\{\boldsymbol{s}^1, \boldsymbol{s}^2\}$. The associated normal vector prior is then $\boldsymbol{n}_p = \frac{\boldsymbol{s}^1 \times \boldsymbol{s}^2}{\|\boldsymbol{s}^1 \times \boldsymbol{s}^2\|}$.

4 Clustering of the data points

The plane fitting algorithm of Sec. 3 requires a clustering method to partition the set of given normalized stain vectors with respect to the two stains applied to a tissue specimen. K-means is arguably one of the most popular clustering methods, and simplifies to Otsu-thresholding when two cluster centers are sought for one-dimensional features. A globally optimal clustering can be computed by Otsu thresholding [4] by discretizing the space of possible thresholds. The computations are efficiently performed using the feature histogram. A suitable one-dimensional feature for the clustering of the plane-fitting algorithm is angle in the fitting plane with respect to a given reference direction (the midpoint of the projection of the stain vector priors). The Otsu threshold minimizes within-cluster variance weighted by cluster size. Prior stain vector information should be used for the thresholding as well to avoid gross mis-clusterings when the number of data points for one stain direction clearly dominates the other.

Prior information can be incorporated into Otsu thresholding by minimizing

$$E(I \leq I_\theta, \mu_1, \mu_2) = \left(\frac{1}{\sigma_1^2} \sum_{i \in j: I_j \leq I_\theta} (I_i - \mu_2)^2 \right) + \left(\frac{1}{\sigma_2^2} \sum_{i \in j: I_j > I_\theta} (I_i - \mu_1)^2 \right)$$
$$+ \frac{1}{(\sigma_2^\mu)^2}(\mu_2 - \overline{\mu}_2)^2 + \frac{1}{(\sigma_1^\mu)^2}(\mu_1 - \overline{\mu}_1)^2,$$

with respect to the unknown threshold I_θ (which completely specifies the segmentation) and the unknown central elements for the two stain angles μ_1 and μ_2. For a given partitioning, the optimal values are

$$\mu_i = \frac{\sigma_i^2}{\sigma_i^2 + (\sigma_i^\mu)^2 n_i} \overline{\mu}_i + \frac{(\sigma_i^\mu)^2 n_i}{\sigma_i^1 + (\sigma_i^\mu)^2 n_i} \overline{I}_i,$$

where n_i denotes the numbers of points and $\overline{I}_i$ the mean angles in the partitions. Note that μ_1 and μ_2 *are not* the foreground and the background mean respectively, but are a weighted average of the means and the priors.

Computing the angle histogram for a set of points allows for direct, efficient computation of $\overline{I}_i$, and consequentially of μ_i. Searching over all discretized threshold values results in the globally optimal threshold I_θ. Neglecting the prior terms or specifying uniform priors and Gaussian distributions for the image likelihoods recovers the standard Otsu thresholding technique.

To avoid specifying separate priors for $\boldsymbol{n}_p$ and for the angle center priors $\overline{\mu}_i$, they are computed based on the stain vector priors. Since the stain vectors themselves are (according to the stain vector model) extreme directions specifying the boundary of the optical density cone, they are not directly useful as priors for Otsu thresholding, as pure colors are the exception, and a distribution of mixed colors is observed. Therefore, given two stain vectors s_1 and s_2, the priors are chosen as the angles with respect to the reference direction of the projections $\Pi\{\boldsymbol{q}_i\} = \boldsymbol{q}_i - \boldsymbol{q}_i^T \boldsymbol{n}\boldsymbol{n}$ onto the current estimate of the plane, where $q_1 = (1-\alpha)s_1 + \alpha s_2$ and $q_2 = \alpha s_1 + (1-\alpha)s_2$, $\alpha \in [0, 0.5)$ ($\alpha = 0.15$ in experiments done here) are directions moved slightly inward of the cone boundaries.

Once the plane has been fit, estimates of stain intensities are computed as the medians of the clusters. Estimates of the stain vectors are obtained by computing the robust minima (γ-percentile) and maxima ($(1-\gamma)$ percentile) within the cluster centers; $\gamma = 1$ for all experiments performed here. These estimates enable the transformation of images into any chosen color-space, and therefore normalization of appearance across multiple slides. While related to the approach by [3] the proposed method alleviates problems with uneven cluster sizes by defining statistics *within* the clusters, by weighting the plane fit with respect to cluster size, and by incorporating prior information. The benefits of this new approach are presented in Sec. 5. Fig. 1 shows an overview of the plane fitting algorithm with prior.

Data: σ/σ^0, $\boldsymbol{s}_1$, $\boldsymbol{s}_2$
Result: Normal vector for plane fit: $\boldsymbol{n}$
Compute prior normal vector $\boldsymbol{n}_p = \frac{\boldsymbol{s}_1 \times \boldsymbol{s}_2}{\|\boldsymbol{s}_1 \times \boldsymbol{s}_2\|}$;
Initialization $\boldsymbol{n} = \boldsymbol{n}_p$;
repeat

Project data points onto plane $\boldsymbol{n}_p^T \boldsymbol{x} = 0$;
Project priors $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ onto the plane (computed from s_1, s_2);
Express all points (including the priors) in angular coordinates ;
Perform globally optimal Otsu thresholding with priors in angular domain;
Compute new scatter matrix (based on clustering) ;
Compute new data variance (based on clustering) ;
Compute optimal normal vector $\boldsymbol{n}$;

until *convergence (i.e., cluster assignments no longer change)* ;

Algorithm 1: Optimal plane fit.

Fig. 1. Algorithmic description of the optimal plane-fit algorithm.

Fig. 2. Estimation consistency for the proposed method and the method not using prior information [3] comparing the mean deviation from the mean normal vector across a slide (with respect to the tiles) in degrees. Smaller values and a tighter distribution demonstrate the advantage of the proposed method. Results are statistically significantly different.
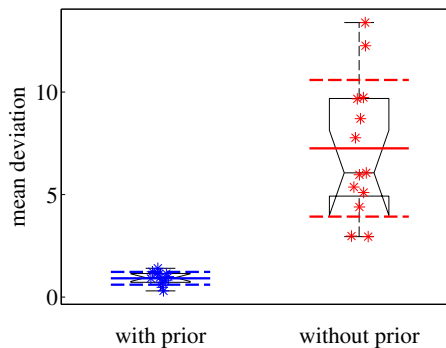
5 Validation and Experimental Results

Figs. 3, 2 and 4 show the performance of the plane fitting algorithm with prior. Fig. 3 shows the results of a synthetic experiment to estimate the plane normal. Three methods are compared: (1) estimation without a plane normal prior (corresponding to the method in [3]), (2) estimation with a plane normal prior, but without the clustering step, and (3) the full algorithm 1. Estimation results shown are deviations (in degrees) of the estimated normal vector $\hat{\boldsymbol{n}}$ with respect to the ground truth normal vector. To assess the influence of varying cluster sizes and normal priors, two stain vectors s_1, s_2 were chosen at a 15 degree angle. Varying numbers of stain vectors were generated for the two stains using an isotropic Gaussian distribution (standard deviation of 0.1 which is similar to that in our real data sets). Priors were generated from the stain vectors by tilting the normal vector by an angle θ with respect to the axis defined by the stain vectors (translated to the origin) and by rotating the priors within the plane by an angle Φ. Fig. 3 shows that all three methods to determine the plane normal have similar performance for clusters of equal size. Prior information improves the

results greatly for uneven point distributions (by almost 10 degrees on average for the most extreme point imbalance 5/1000 as expected to occur for example in regions with sparsely distributed nuclei). In cases where the effect of the prior is most pronounced, the clustering further improves estimation results.

Fig. 2 shows the performance of the method on 13 real datasets compared to direct plane fitting without a plane prior as in [3]. The histology images were subdivided into areas of 1000x1000 pixels and were independently adjusted for stain intensity and stain direction using the two methods. Fig. 2 shows the estimation consistency for the two methods by comparing the mean deviation from the mean normal vector across a slide (mean with respect to the tiles). Estimation consistency is statistically significantly better for the proposed method (with $p < 1e-4$ using a t-test or non-parametric permutation test). The mean deviation from the prior was around 11 degrees for the method using prior information

Fig. 3. Synthetic experiments: Angle difference (in degrees) between the estimated normal vector and the ground truth. Top row: proposed method. Middle row: plane fit with prior without clustering error minus with clustering error. Bottom row: plane fit without prior error minus plane fit with prior with clustering error. Estimates are results of 1000 random samples for different priors and varying numbers (5/1000, ...) of points in the two stain clusters. The proposed method performs best. The white line shows the zero level set of angle differences.

and 20 degrees for the method not using the prior information. The tight distribution for the consistency results for the proposed method demonstrates that the prior was not chosen to dominate the results.

To illustrate the behavior of the estimation method graphically Fig. 4 shows the results for a real dataset compared to direct plane fitting without a plane prior. Stain intensity scaling factors and the deviations from the mean of the estimates for the normal direction are shown. For a well working method, the results are expected to be approximately uniform. While the difference for the stain correction between the two methods (in both cases prior information was used for the clustering to obtain the intensity scalings) is not as drastic as for the normal direction, the plane fitting method with prior improves intensity scalings as can be witnessed by the reconstruction results, which are almost perfectly uniform for the method using prior information and inconsistent otherwise.
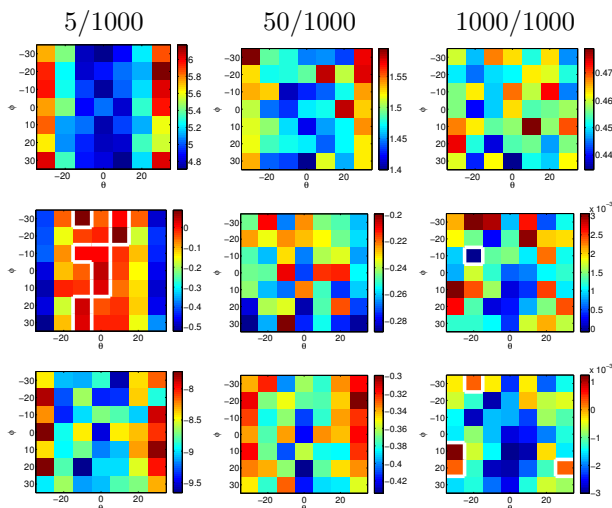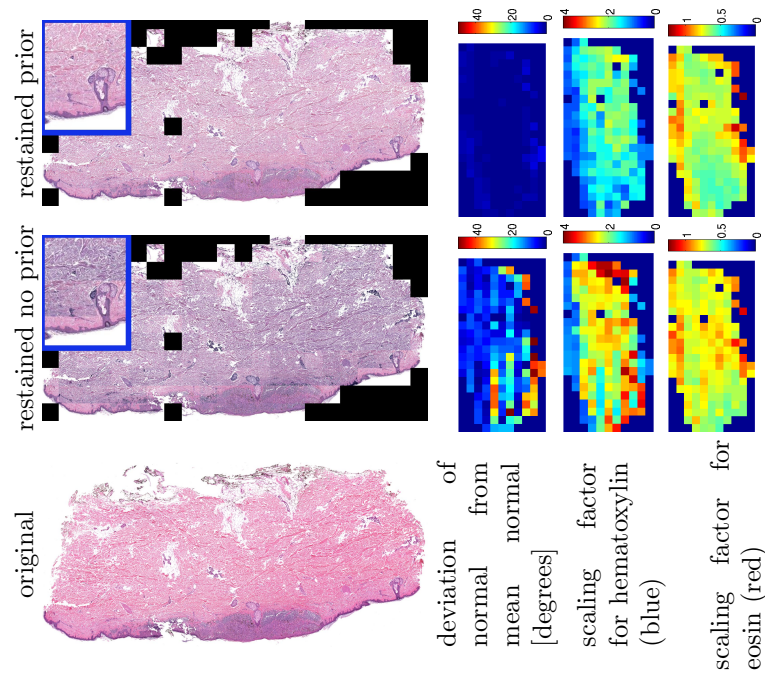
Fig. 4. Restaining of a real dataset using the proposed method (right) and the method not using prior information (middle). Zoom-ins (top-right).

6   Conclusions

This paper presented a method to automatically adjust the appearance of stained histology slides. It described a novel way of adding prior information for the stain vectors and how to deal with unequal stain distribution through a clustering process. The clustering is a novel adaptation of Otsu thresholding including prior information. The underlying optimization problem relates to trust-region optimization and is therefore well studied and easy to solve. Real and synthetic experiments showcase the superior performance of the method developed.

References

1. Ruifrok, A., Johnston, D.: Quantification of histochemical staining by color deconvolution. In: Analytical and Quantitative Cytology and Histology. Volume 23. (August 2001) 291–299
2. Rabinovich, A., Agarwal, S., Laris, C., Price, J., Belongie, S.: Unsupervised color decomposition of histologically stained tissue samples. In: Advances in Neural Information Processing Systems. (2003)
3. Macenko, M., Niethammer, M., Marron, J., Borland, D., Woosley, J., Guan, X., Schmitt, C., Thomas, N.: A method for normalizing histology slides for quantitative analysis. In: Proceedings of the Sixth IEEE International Symposium on Biomedical Imaging (ISBI). (2009) 1107–1110
4. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11** (1975) 285–296
5. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer (2006)