

A domain-knowledge-inspired mathematical framework for the description and classification of H&E stained histopathology images

Melody L. Massar^a, Ramamurthy Bhagavatula^b, John A. Ozolek^c,
Carlos A. Castro^d, Matthew Fickus^a and Jelena Kovačević^{b,e}

^aDepartment of Mathematics and Statistics

Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, USA

^bDepartment of Electrical and Computer Engineering, ^eDepartment of Biomedical Engineering
Carnegie Mellon University, Pittsburgh, PA, USA

^cDepartment of Pathology, Children’s Hospital of Pittsburgh

^dDepartment of Obstetrics and Gynecology, Magee-Womens Research Institute and Foundation
University of Pittsburgh, PA, USA

ABSTRACT

We present the current state of our work on a mathematical framework for identification and delineation of histopathology images—local histograms and occlusion models. Local histograms are histograms computed over defined spatial neighborhoods whose purpose is to characterize an image locally. This unit of description is augmented by our occlusion models that describe a methodology for image formation. In the context of this image formation model, the power of local histograms with respect to appropriate families of images will be shown through various proved statements about expected performance. We conclude by presenting a preliminary study to demonstrate the power of the framework in the context of histopathology image classification tasks that, while differing greatly in application, both originate from what is considered an appropriate class of images for this framework.

Keywords: local histogram, occlusion, texture, classification, segmentation

1. INTRODUCTION

This paper highlights the main results given in a journal article⁶ about local histograms, and introduces some new concepts currently being pursued in follow-up research. *Local histograms* of an image are histograms of the values of the pixels that lie in a neighborhood of a given pixel’s location. They indicate the combination of pixel intensities or colors that appear in that neighborhood. We model our images as functions from a finite abelian group \mathcal{X} of pixel locations into a second finite abelian group \mathcal{Y} of pixel values. That is, our images f are members of the set $\ell(\mathcal{X}, \mathcal{Y}) := \{f : \mathcal{X} \rightarrow \mathcal{Y}\}$. The local histograms of an image f are defined in terms of a nonnegatively-valued *weighting function* $w \in \ell(\mathcal{X}, \mathbb{R})$ whose values sum to one. Specifically, the *local histogram transform* of f with respect to w is the function $\text{LH}_w f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$,

$$(\text{LH}_w f)(x, y) := \sum_{x' \in \mathcal{X}} w(x') \delta_y(f(x + x')), \quad (1)$$

where $\delta_y(f(x + x')) = 1$ if $f(x + x') = y$ and is otherwise zero.

Because local histograms mimic certain visual cues that people use when hand-segmenting an image, they have great potential as features in image classification schemes. We use them to automate the process of segmenting and classifying digital microscope images of biological tissues, like that of Figure 1(a), which is purple-pink from hematoxylin and eosin (H&E) staining. Although one can distinguish the different textures in such an image, segmenting the tissues by hand (Figure 1(b)) can be a tedious and time-consuming task, even for medical experts. Therefore, it is our goal to automate as much of this process as is possible:

Send correspondence to Matthew Fickus: E-mail: Matthew.Fickus@afit.edu

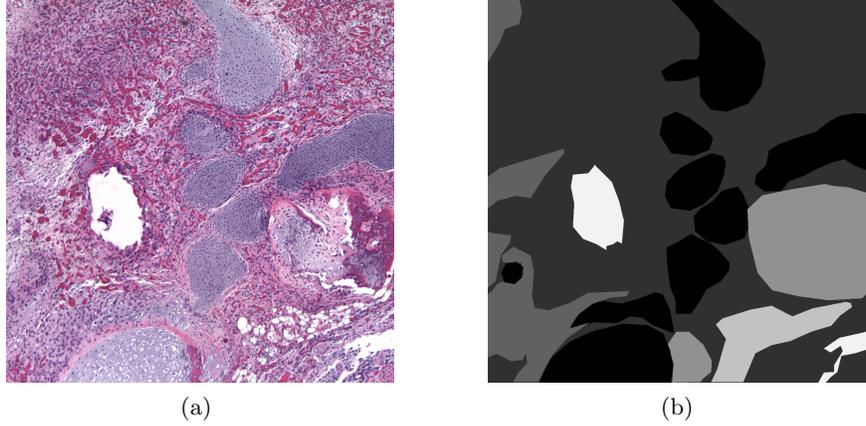


Figure 1. (a) A digital microscope image of a H&E-stained tissue section. (b) The histology image has been manually segmented and classified by a medical expert, resulting in the per-pixel labels. From darkest to lightest, the labels indicate cartilage, pseudovascular tissue, connective tissue, bone, fatty tissue, and background pixels, respectively. Our goal is to automate this segmentation-and-classification process.

Problem 1. Develop mathematical features for an algorithm which assigns a tissue label to each pixel in a histology image, thereby segmenting (delineating) and classifying (identifying) the image at the same time.

We use local features, specifically local histograms, in designing our algorithm for Problem 1 because considering pixels in isolation will often provide insufficient information to determine a label, and considering global features will often destroy spatial context. We also expect local histograms to have great potential in our classification scheme because textures such as cartilage, connective tissue and pseudovascular are distinguishable by their distinct color distributions.

The *occlusion* of a set of N images $\{f_n\}_{n=0}^{N-1} \in \ell(\mathcal{X}, \mathcal{Y})$ with respect to a given *label function* $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)$ is:

$$(\text{occ}_\varphi \{f_n\}_{n=0}^{N-1})(x) := f_{\varphi(x)}(x), \quad (2)$$

where \mathbb{Z}_N denotes the cyclic group of integers modulo N . That is, at any pixel location x , the label $\varphi(x)$ determines which of the potential pixel values $\{f_n(x)\}_{n=0}^{N-1}$ actually appears in the composite image $\text{occ}_\varphi \{f_n\}_{n=0}^{N-1}$ at that point. The *occlusion model* Φ is a random variable version of φ , meaning there exists a probability density function $P_\Phi : \ell(\mathcal{X}, \mathbb{Z}_N) \rightarrow [0, 1]$ such that $\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) = 1$. The role of the occlusion model Φ is to assign a probability to each of these possible φ 's in a manner that emphasizes those textures one expects to appear in a given tissue while de-emphasizing the rest.

Our main results⁶ are concerned with when the local histograms (1) of a composite image (2) are related to the local histograms of the individual f_n 's. Specifically, we show that by averaging over all possible label functions φ , the local histograms of the composite image is exactly a convex combination of the local histograms of each image provided our occlusion model Φ is *flat*, meaning that on average, the probability that Φ chooses label n at a given pixel location x is equal to the probability of choosing n at any other x' ; formally, Φ is *flat* if there exists scalars $\{\lambda_n\}_{n=0}^{N-1}$ such that:

$$\sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N) \\ \varphi(x)=n}} P_\Phi(\varphi) = \lambda_n, \quad \forall x \in \mathcal{X}. \quad (3)$$

Note that for any fixed $x \in \mathcal{X}$, summing (3) over all n yields that $\sum_{n=0}^{N-1} \lambda_n = 1$. Indeed, at any given pixel location x , the value λ_n represents the probability that the random label function Φ will have label n at that x . These concepts in hand, we present one of our main results, which formally claims that, on average, the local histograms of composite images produced from flat occlusion models are but mixtures of the local histograms of the source images:

Theorem 1. If Φ is flat as in (3), then the expected value of the local histogram transform (1) of a composite image (2) is a convex combination of the local histograms of each individual image:

$$\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_{\Phi}(\varphi)(\text{LH}_w \text{occ}_{\varphi} \{f_n\}_{n=0}^{N-1})(x, y) = \sum_{n=0}^{N-1} \lambda_n (\text{LH}_w f_n)(x, y). \quad (4)$$

By Theorem 1, it is reasonable to decompose local histograms into convex combinations of more basic distributions and use these distributions to segment and classify: given a new image, we assign a label at any given point by determining which particular set of learned distributions its local histogram is most consistent with.

The remainder of our main results are in support of this interpretation of using Theorem 1 to solve Problem 1. Specifically, the next section discusses local histograms: their basic properties and how to efficiently compute them. Additionally, we give a generalization of Theorem 1 to the non-flat case, namely Theorem 2, discuss a few methods used to construct flat occlusion models, and present a special version of flatness used to compute the variance of the local histograms of a composite image. The final section discusses the results of a preliminary segmentation-and-classification algorithm inspired by Theorem 1 in which local histograms are decomposed using principal component analysis (PCA).

2. LOCAL HISTOGRAMS OF RANDOMLY-GENERATED TEXTURES

Computing local histograms can be time consuming, especially as \mathcal{X} and \mathcal{Y} become large. In particular, for a general window w , a direct computation of (1) requires $\mathcal{O}(|\mathcal{X}|^2|\mathcal{Y}|)$ operations: $\mathcal{O}(|\mathcal{X}|)$ operations for each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. A more efficient method, which is proven in earlier work,^{4,6} gives that (1) can be computed as a system of $|\mathcal{Y}|$ convolutions over \mathcal{X} , which only requires $\mathcal{O}(|\mathcal{X}||\mathcal{Y}|\log|\mathcal{X}|)$ operations if discrete Fourier transforms are used. In particular, we filter the *characteristic function* of the graph of f , namely $1_f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$,

$$1_f(x, y) := 1_{f^{-1}\{y\}}(x) = \delta_y(f(x)) = \begin{cases} 1, & f(x) = y, \\ 0, & f(x) \neq y, \end{cases} \quad (5)$$

with the *reversal* of $w \in \ell(\mathcal{X}, \mathbb{R})$, namely $\tilde{w}(x) := w(-x)$. Therefore, local histograms (1) are computed as

$$(\text{LH}_w f)(x, y) = (\tilde{w} * 1_{f^{-1}\{y\}})(x).$$

Local histograms possess several other basic properties: the levels of their transform sum to 1, they commute with spatial translation, they are shifted along the group of pixel values when a constant is added to the image, and they are binned when the image is quantized; the proofs of these basic properties are given in earlier work.⁴

Now fix any set of N source images $\{f_n\}_{n=0}^{N-1}$ and let Φ be any occlusion model as defined in the introduction. That is, let Φ be a random variable version of a label function $\varphi : \mathcal{X} \rightarrow \mathbb{Z}_N$, as defined by a probability density function $P_{\Phi} : \ell(\mathcal{X}, \mathbb{Z}_N) \rightarrow [0, 1]$ where $\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_{\Phi}(\varphi) = 1$. In the results that follow, a useful quantity to consider is the expected value—with respect to P_{Φ} —of the characteristic function 1_{φ} obtained by letting $f = \varphi$ in (5):

$$\bar{1}_{\Phi}(x, n) := \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_{\Phi}(\varphi) 1_{\varphi}(x, n) = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N) \\ \varphi(x) = n}} P_{\Phi}(\varphi).$$

When compared with the definition of flatness (3), we see that Φ is flat if and only if there exist scalars $\{\lambda_n\}_{n=0}^{N-1}$ such that $\bar{1}_{\Phi}(x, n) = \lambda_n$ for all $x \in \mathcal{X}$ and $n \in \mathbb{Z}_N$. That is, Φ is flat if and only if $\bar{1}_{\Phi}(x, n)$ is constant with respect to pixel location x . Having this concept, we present one of our main results:

Theorem 2. For any sequence of images $\{f_n\}_{n=0}^{N-1} \in \ell(\mathcal{X}, \mathcal{Y})$, weighting function w and any N -image occlusion model Φ , the expected value of the local histogram (1) of the composite image (2) with respect to w is:

$$E_{\Phi}(\text{LH}_w \text{occ}_{\Phi} \{f_n\}_{n=0}^{N-1})(x, y) = \sum_{n=0}^{N-1} \bar{1}_{\Phi}(x, n) (\text{LH}_w f_n)(x, y) + \varepsilon, \quad (6)$$

where the error term ε is bounded by $|\varepsilon| \leq \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x') |\bar{1}_\Phi(x + x', n) - \bar{1}_\Phi(x, n)|$. Moreover,

$$\sum_{n=0}^{N-1} \bar{1}_\Phi(x, n) = 1, \quad (7)$$

and so (6) states that, on average, the local histograms of the composite image $\text{occ}_\varphi \{f_n\}_{n=0}^{N-1}$ can be approximated by convex combinations of local histograms of each individual image f_n .

When Φ is flat, (6) simplifies to (4); a comparison of these computations is demonstrated in Figures 4 and 5 of the accompanying journal article.⁶ Thus, flatness is indeed an important theoretical assumption for the analysis of local histograms of textures generated via random occlusions. It nevertheless remains to be shown that flatness is also a realistic assumption from the point of view of our motivating application.

In previous work,^{4,6} we show that flat occlusion models can be built in several ways. In particular, we show that if an occlusion model Φ is *translation-invariant*, that is, if $P_\Phi(T^x \varphi) = P_\Phi(\varphi)$ for all $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)$ and for all $x \in \mathcal{X}$, then Φ is flat (3). We also present two operators that can be used to construct flat occlusion models: the expansion operator and the overlay operator.

Given any $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)$ and $\{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^\mathcal{X}$, the *expansion* of φ by $\{\psi_x\}_{x \in \mathcal{X}}$ is $\varphi \star \{\psi_x\}_{x \in \mathcal{X}} \in \ell(\mathcal{X}, \mathbb{Z}_2)$,

$$(\varphi \star \{\psi_{x'}\}_{x' \in \mathcal{X}})(x) := \begin{cases} 1, & x = x' + x'', \varphi(x') = 1, \psi_{x'}(x'') = 1, \\ 0, & \text{else,} \end{cases}$$

and given any occlusion models Φ and Ψ from \mathcal{X} into \mathbb{Z}_2 , the expansion of Φ by Ψ is the occlusion model $\Phi \star \Psi$ whose probability density function is $P_{\Phi \star \Psi} : \ell(\mathcal{X}, \mathbb{Z}_2) \rightarrow [0, 1]$,

$$P_{\Phi \star \Psi}(\sigma) := \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^\mathcal{X} \\ \varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sigma}} P_\Phi(\varphi) \prod_{x \in \mathcal{X}} P_\Psi(\psi_x).$$

We can show that^{4,6} $\Phi \star \Psi$ is flat (3), if either Φ or Ψ is flat and Φ and Ψ are *effectively disjoint*:

$$\text{If } P_\Phi(\varphi) > 0 \text{ and } P_\Psi(\psi_x) > 0 \text{ for all } x \in \mathcal{X}, \text{ then } \varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sum_{\substack{x \in \mathcal{X} \\ \varphi(x)=1}} T^x \psi_x. \quad (8)$$

Meanwhile, for any $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi})$, $\psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi})$ and $\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)$, the *overlay* of φ over ψ with respect to σ is $\varphi \#_\sigma \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi})$,

$$(\varphi \#_\sigma \psi)(x) := \begin{cases} \varphi(x), & \sigma(x) = 0, \\ \psi(x) + N_\varphi, & \sigma(x) = 1, \end{cases}$$

and given probability density functions P_Φ , P_Ψ and P_Σ on $\ell(\mathcal{X}, \mathbb{Z}_{N_\Phi})$, $\ell(\mathcal{X}, \mathbb{Z}_{N_\Psi})$ and $\ell(\mathcal{X}, \mathbb{Z}_2)$, respectively, the overlay of the occlusion model Φ over Ψ with respect to Σ is the new occlusion model $\Phi \#_\Sigma \Psi$ whose probability density function is $P_{\Phi \#_\Sigma \Psi} : \ell(\mathcal{X}, \mathbb{Z}_{N_\Phi + N_\Psi}) \rightarrow [0, 1]$,

$$P_{\Phi \#_\Sigma \Psi}(v) := \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Phi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \varphi \#_\sigma \psi = v}} P_\Phi(\varphi) P_\Psi(\psi) P_\Sigma(\sigma).$$

We can show that^{4,6} if Φ , Ψ , and Σ are flat (3), then $\Phi \#_\Sigma \Psi$ is also flat.

In our current work, we compute the variance of the local histograms of a composite image. In order to do this, we make use of the definition of *2-flatness*: Φ is 2-flat if there exists scalars $\{\lambda_n\}_{n=0}^{N-1}$ such that:

$$\lambda_{n'}\lambda_{n''} = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_{\Phi}(\varphi) 1_{\varphi}(x', n') 1_{\varphi}(x'', n''), \quad \forall n', n'' \in \mathbb{Z}_N, \forall x', x'' \in \mathcal{X}, x' \neq x''. \quad (9)$$

In other words, Φ is 2-flat if, on average, the probability that Φ chooses labels n' and n'' at two separate pixel locations is equal to the probability of choosing n' and n'' at any other two separate pixel locations. Having this property of 2-flatness allows us to express the variance of the local histograms of a composite image as a combination of the local histograms of each image with respect to the square of the weighting function:

Theorem 3. Let $\{f_n\}_{n=0}^{N-1} \in \ell(\mathcal{X}, \mathcal{Y})$ be any sequence of images such that $f_{n'}(x) = f_{n''}(x)$ implies $n' = n''$, w be any weighting function, and Φ be any N -image occlusion model. If Φ is 2-flat as in (9), then the variance of the local histogram transform (1) of a composite image (2) is:

$$V_{\Phi}(\text{LH}_{w, \text{occ}_{\Phi}}\{f_n\}_{n=0}^{N-1})(x, y) = \sum_{n=0}^{N-1} \lambda_n(1 - \lambda_n)(\text{LH}_{w^2} f_n)(x, y).$$

Two-flatness is important in order to understand the variance of the local histograms of a composite image and as such, we must consider if 2-flatness is a realistic assumption in terms of our motivating application. Indeed, just as flat occlusion models were constructed from the expansion and overlay operators, we can also construct 2-flat occlusion models from such operators. In particular, for the expansion operator, we have that $\Phi \star \Psi$ is 2-flat (9) if Φ and Ψ are effectively disjoint (8) and 2-flat, while for the overlay operator, we have that $\Phi \#_{\Sigma} \Psi$ is 2-flat if Φ , Ψ , and Σ are 2-flat.

3. LOCAL HISTOGRAM-BASED SEGMENTATION-AND-CLASSIFICATION

In this section, we present the results of a proof-of-concept segmentation-and-classification scheme that is inspired by Theorem 1 which decomposes local histograms using PCA. In the algorithm,⁶ local histograms are the only image features that are computed. That is, the decision of which label to assign to a given pixel is based purely on the distribution of color in its surrounding neighborhood. We do this to demonstrate the validity of the concept embodied by Theorem 1 as an image processing tool.

Two runs of this classification algorithm are depicted in Figure 2. For our particular set of histology images, we experimentally found that we could still obtain good accuracies even if we discard the green channel of our purple-pink images, and moreover quantize the 8-bit red and blue channels down to 3-bits apiece. For computational advantages, in the first run, we train the classifier on the quantized version of Figure 1(a) which is given in Figure 2(a); for the sake of readability, a 3-bit quantized version of the unused green channel was included in this rendering. For the sake of simplicity, we restrict ourselves to classifying 3 tissue types: cartilage, connective tissue and pseudovascular tissue; all other tissue types are ignored in the confusion matrices given below. Using our PCA-based algorithm⁶ to segment and classify Figure 2(a) results in the per-pixel labels given in Figure 2(c) for 4 principal components. Comparing Figure 2(c) and the ground truth of Figure 2(b), we see both the power and limitations of local histograms: color is a big factor in determining tissue type, but by ignoring shape, we suffer from oversmoothing. The accuracy percentages for various choices of the number of principal components N are given by a confusion matrix:

	$N = 1$			$N = 2$			$N = 3$			$N = 4$		
	Ca	Co	Ps									
Ca	77	22	1	87	11	2	96	3	1	96	3	1
Co	0	95	5	0	91	9	3	94	3	3	92	5
Ps	2	11	87	2	8	90	6	7	87	5	5	90

Here each row of the matrix tells us the percentage a certain tissue was labeled as cartilage (Ca), connective tissue (Co), and pseudovascular tissue (Ps). In particular, the first three entries of the first row of this table

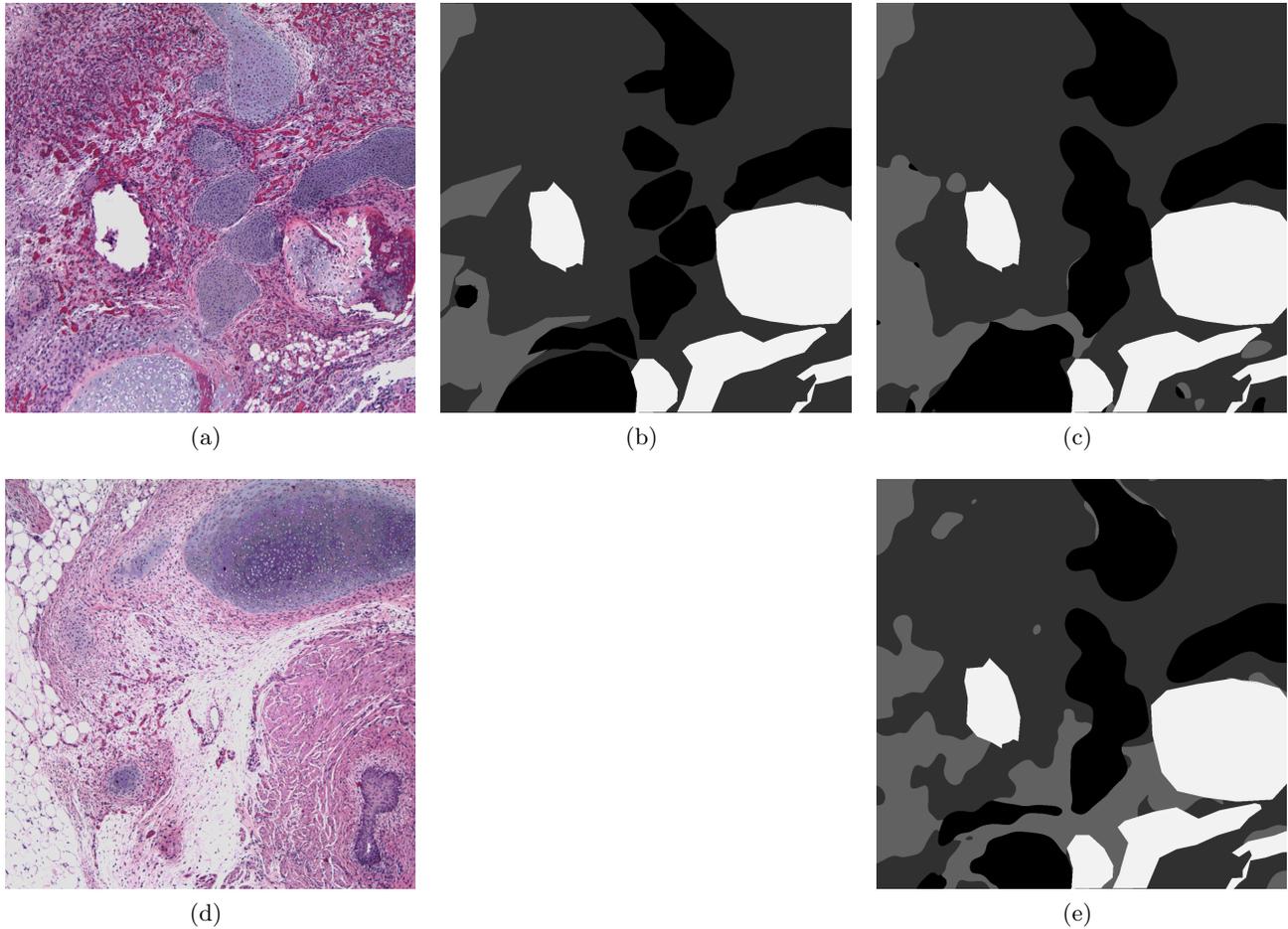


Figure 2. An example of using PCA of local histograms to perform segmentation and classification of the image given in (a), which is a 3-bit quantized version of Figure 1(a). A manually segmented and labeled version of (a) is shown in (b) where black represents cartilage, light gray represents connective tissue, dark gray represents pseudovascular tissue, and white represents other tissues that have been ignored in this proof-of-concept experimentation. Using (a) as both the training and testing image in our PCA-based classification scheme, we obtain the labels shown in (c). A similar, but less-accurate classification of (a) can still be obtained if we instead train on (d), resulting in the labels given in (e).

tell us that when using a single principal component, those points labeled as cartilage by a medical expert in Figure 2(b) are correctly labeled as such by our algorithm 77% of the time, while 22% of it is mislabeled as connective tissue and 1% of it is mislabeled as pseudovascular tissue. Note here that we have trained and tested on the same image; such experiments indicate the feasibility of our approach in a semi-automated classification scheme.

The second run of this algorithm is almost identical to the first, with the exception that we use a distinct image (Figure 2(d)) in the training phase. We then apply the principal components obtained from Figure 2(d) to generate labels (Figure 2(e)) for Figure 2(a) using our PCA-based algorithm. Compared to the first run, the algorithm's performance here is a better indication of its feasibility as a fully automated classification scheme,

and is summarized by the following confusion matrix:

	$N = 1$			$N = 2$			$N = 3$			$N = 4$		
	Ca	Co	Ps									
Ca	90	9	1	91	4	5	90	5	5	83	11	6
Co	25	61	14	10	62	28	7	79	14	8	70	22
Ps	30	12	58	4	10	86	4	50	46	2	17	81

Though the performance in the second run is understandably worse than that of the first, it nevertheless demonstrates the real-world potential of the idea exemplified by Theorem 1: the local histograms of certain types of textures can be decomposed into more basic distributions, and this decomposition can serve as an image processing tool.

We conclude by noting that for algorithms intended for real-world use, color information should be combined with morphological data—size, local and global shape, orientation and organization—in order to obtain better classification accuracies. An example of such an algorithm, accompanied by thorough testing and comparisons against other state-of-the-art methods, is given in earlier work;² these facts are not reprinted here.

Acknowledgments

The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

REFERENCES

- [1] R. Bhagavatula, M. Fickus, J. A. Ozolek, C. A. Castro, J. Kovačević, Automatic identification and delineation of germ layer components in H&E stained images of teratomas derived from human and nonhuman primate embryonic stem cells, Proc. IEEE Int. Symp. Biomed. Imag. (2010) 1041–1044.
- [2] R. Bhagavatula, M. L. Massar, M. Fickus, C. A. Castro, J. A. Ozolek, J. Kovačević, Automated identification of tissues for digital pathology, submitted.
- [3] A. Chebira *et al.*, Multiresolution identification of germ layer components in teratomas derived from human and nonhuman primate embryonic stem cells, Proc. IEEE Int. Symp. Biomed. Imag. (2008) 979–982.
- [4] M. L. Massar, Local histogram for per-pixel classification, Ph.D. Prospectus (2010).
- [5] M. L. Massar, R. Bhagavatula, M. Fickus, J. Kovačević, Local histograms for classifying H&E stained tissues, Proc. 26th Southern Biomed. Eng. Conf. (2010) 348–352.
- [6] M. L. Massar, R. Bhagavatula, M. Fickus and J. Kovačević, Local histograms and image occlusion models, submitted, arXiv:1105.4069.