

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at SciVerse ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha



Local histograms and image occlusion models

Melody L. Massar^a, Ramamurthy Bhagavatula^b, Matthew Fickus^{a,*}, Jelena Kovačević^{b,c}^a Department of Mathematics and Statistics, Air Force Institute of Technology, Wright–Patterson Air Force Base, OH 45433, USA^b Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA^c Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:

Received 20 May 2011

Revised 31 March 2012

Accepted 15 July 2012

Available online 24 July 2012

Communicated by Henrique Malvar

Keywords:

Local histogram

Occlusion

Texture

Classification

Segmentation

ABSTRACT

The local histogram transform of an image is a data cube that consists of the histograms of the pixel values that lie within a fixed neighborhood of any given pixel location. Such transforms are useful in image processing applications such as classification and segmentation, especially when dealing with textures that can be distinguished by the distributions of their pixel intensities and colors. We, in particular, use them to identify and delineate biological tissues found in histology images obtained via digital microscopy. In this paper, we introduce a mathematical formalism that rigorously justifies the use of local histograms for such purposes. We begin by discussing how local histograms can be computed as systems of convolutions. We then introduce probabilistic image models that can emulate textures one routinely encounters in histology images. These models are rooted in the concept of image occlusion. A simple model may, for example, generate textures by randomly speckling opaque blobs of one color on top of blobs of another. Under certain conditions, we show that, on average, the local histograms of such model-generated-textures are convex combinations of more basic distributions. We further provide several methods for creating models that meet these conditions; the textures generated by some of these models resemble those found in histology images. Taken together, these results suggest that histology textures can be analyzed by decomposing their local histograms into more basic components. We conclude with a proof-of-concept segmentation-and-classification algorithm based on these ideas, supported by numerical experimentation.

Published by Elsevier Inc.

1. Introduction

A *local histogram* of an image is a histogram of the values of the pixels that lie in a neighborhood of a given pixel's location. It indicates the particular combination of pixel intensities or colors that appear in that neighborhood. When used as features in an image classification scheme, such histograms can help distinguish one texture from another. We, in particular, use them in automated segmentation-and-classification algorithms for digital microscope images of biological tissues.

To be precise, the work presented here was motivated by the need to identify and delineate the various tissues exhibited in images of histological sections of teratoma tumors derived from embryonic stem cells, such as the one given in Fig. 1(a). This image was provided by Dr. Carlos Castro of the University of Pittsburgh and Dr. John A. Ozolek of the Children's Hospital of Pittsburgh, who grow and image such teratomas to gain greater insight into tissue development. In this image, which is purple-pink from hematoxylin and eosin (H&E) staining, even a layman can discern several distinct textures, each corresponding to a distinct tissue type. For each image under study, Drs. Castro and Ozolek make use of their years of

* Corresponding author.

E-mail address: Matthew.Fickus@afit.edu (M. Fickus).

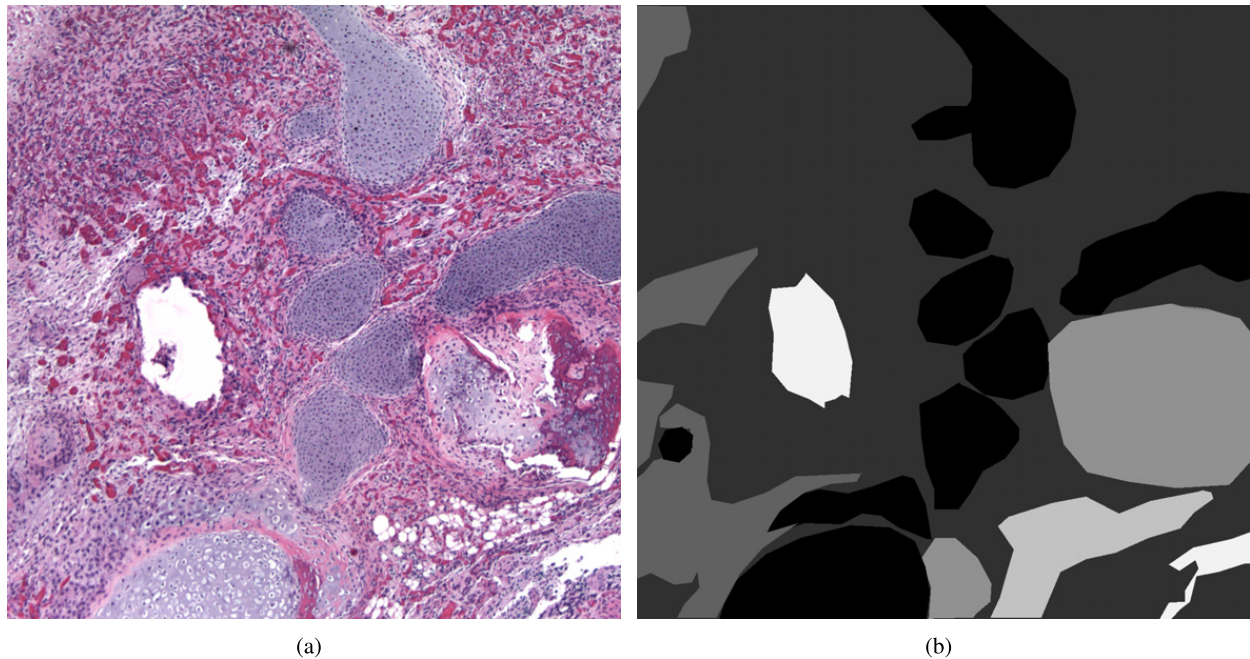


Fig. 1. (a) A digital microscope image of an H&E-stained tissue section. (b) The histology image has been manually segmented and classified by a medical expert, resulting in the per-pixel labels. From darkest to lightest, the labels indicate cartilage, pseudovascular tissue, connective tissue, bone, fatty tissue, and background pixels, respectively. Our goal is to automate this segmentation-and-classification process. The purpose of this paper is to provide a theoretical justification for using local histograms to achieve this goal. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

medical training and experience to identify what tissues are present, and to what degree. Moreover, when provided with a point-and-click interface, they can manually segment the image according to tissue type, resulting in per-pixel labels such as those given in Fig. 1(b). Though straightforward for medical experts, such tasks are nevertheless tedious and time-consuming, leading to inconsistencies when working with large data sets. It is therefore our goal to automate as much of this process as is possible. Our current algorithm is given in [1] and builds upon previous work given in [2,4,14].

Our use of local histograms was motivated by the unique image features found in histology images of teratomas derived from primate embryonic stem cells. In layman's terms, these tumors begin as masses of undifferentiated cells that are implanted in laboratory animals. Over time, these tumors grow and their cells differentiate into many various types – bone, cartilage, skin, etc. – until a point at which they are excised, sectioned, stained and viewed under a microscope, resulting in images such as the one in Fig. 1(a). As such, these images exhibit a wide variety of tissue types, arranged in a seemingly random fashion. Indeed, to a casual observer such images can appear as a jumbled mess. In truth however, the arrangement of these tissues is not completely random, and is rather the result of not yet well-understood biological mechanisms. Drs. Castro and Ozolek believe that by looking at many such images – many sections of many teratomas – they can gain greater insight into these mechanisms. Here, spatial context is crucial: one must identify which particular tissue is present at any given point in order to estimate the total amount of each type, as well as the degree to which any given type is adjacent to other types.

In light of these facts, we seek an algorithm which assigns a tissue label to each pixel, thereby segmenting (delineating) and classifying (identifying) the image at the same time. Indeed, such an algorithm would be useful in a broad class of digital pathology applications beyond the teratoma problem [1]. While designing such an algorithm, we must keep in mind that often no single pixel contains enough information to uniquely determine a label. Rather, the decisions will be made based on features computed over some fixed neighborhood of every given pixel location. To determine which specific features to use, it helps to have a closer look at each individual tissue. For example, for the 1200×1200 image given in Fig. 1(a) and thumbnailed in Fig. 2(a), we zoom in on three tissue types – cartilage, connective tissue and pseudovascular tissue – resulting in the 128×128 subimages given in Fig. 2(b), (c) and (d), respectively. Each of these three tissue types exhibits a unique aperiodic texture. For instance, the cartilage texture can be regarded as a light purple field speckled with darker reddish-purple blobs; each blob represents an individual cell's nucleus. Meanwhile, connective tissue appears as dark purple blobs over a light pink field; pseudovascular tissue is similar to connective tissue, but contains additional reddish-pink structures. In particular, these three textures exhibit distinct distributions of color, a fact which can quantitatively be confirmed by computing the two-dimensional histograms of their red-blue (RB) pixel value pairs, as depicted in Fig. 2(f), (g) and (h).

As certain tissues can be distinguished from others based solely on the distributions of their pixel values, we propose to use histograms as image features in a segmentation-and-classification scheme. These histograms must be computed locally – over a fixed neighborhood of every pixel location – since global histograms, such as the one depicted in Fig. 2(e) derived

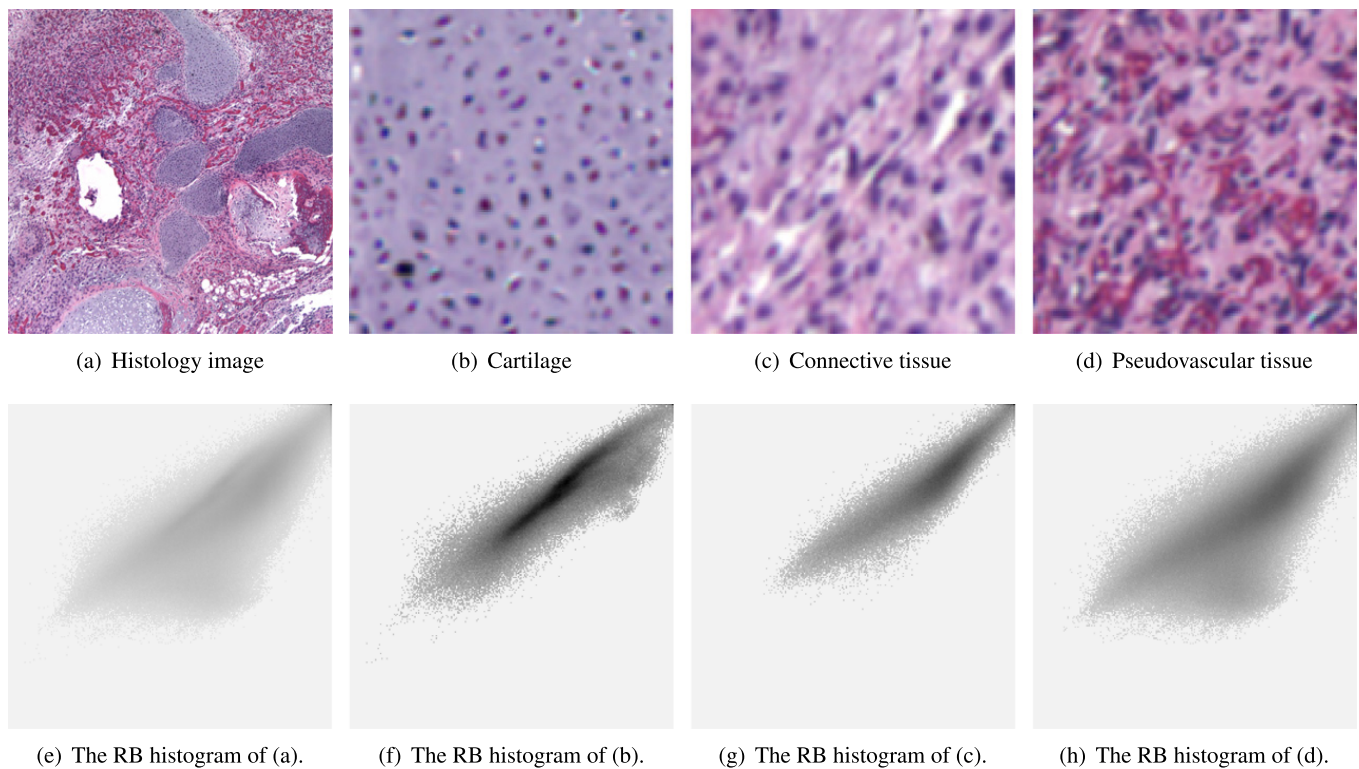


Fig. 2. A 1200×1200 histology image exhibiting multiple tissue types (a) and the 256×256 histogram of its red-blue (RB) pixel values (e). As is common with H&E staining, the tissues in (a) are purple-pink and so we ignore the green component of these red-green-blue (RGB) images when computing (e). This histogram is viewed from above, with red and blue ranging from 0 to 255 on the horizontal and vertical axes, respectively; here the height of the histogram is proportional to darkness for the sake of readability. In (b), (c) and (d) we zoom in on three 128×128 patches extracted from (a), each of which exhibit a single tissue type, namely cartilage, connective tissue and pseudovascular tissue, respectively. Each of these three tissue types has a distinct distribution of pixel values, as evidenced by their corresponding RB histograms (f), (g) and (h). These 256×256 histograms are similar to (e), but are only computed over those points of a given type according to the ground truth labels in Fig. 1(b). In particular, the histogram (f) of cartilage (b) is computed over all points labeled in black in Fig. 1(b). We see that cartilage is darker, on average, than connective: (f) is distributed more towards the lower left-hand side than (g) is. Moreover, pseudovascular is similar to connective, but possesses additional reddish-pink structures, as evidenced by the subdiagonal blob found in (h), but not (g). As such, it is plausible that local histograms can serve as discriminating features in segmentation-and-classification algorithms. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

from Fig. 2(a), destroy spatial context by mixing all of the individual distributions together. A similar issue arises in time-frequency analysis: spectrograms preserve spatial context while Fourier transforms do not. Indeed, local histograms are philosophically similar to spectrograms: in a neighborhood of a given point, the local histogram transform estimates the *frequency of occurrence* of a given value while the spectrogram estimates frequency in the traditional sense.

The purpose of this paper is to provide a mathematically rigorous justification for the use of local histograms in this fashion. To be precise, we regard 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 $\ell(\mathcal{X}, \mathcal{Y}) := \{f : \mathcal{X} \rightarrow \mathcal{Y}\}$. For example, the 1200×1200 , 8-bit red-green-blue (RGB) image given in Fig. 1(a) has $\mathcal{X} = \mathbb{Z}_{1200}^2 = \mathbb{Z}_{1200} \times \mathbb{Z}_{1200}$ and $\mathcal{Y} = \mathbb{Z}_{256}^3$, where \mathbb{Z}_N denotes the cyclic group of integers modulo N . For purple-pink H&E-stained images, we often omit the green channel for the sake of computational efficiency, at which point \mathcal{Y} becomes \mathbb{Z}_{256}^2 . The local histograms of an image f are defined in terms of a *weighting function*, that is, a nonnegatively-valued $w \in \ell(\mathcal{X}, \mathbb{R})$ 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. For any fixed $x \in \mathcal{X}$, the corresponding cross-section of this function, namely $(\text{LH}_w f)(x, \cdot) : \mathcal{Y} \rightarrow \mathbb{R}$, counts the number of instances at which f obtains a given value y in a w -neighborhood of x .

In this paper, we show that local histogram transforms (1) are well-suited to the analysis of a particular class of textures. In short, we want a rigorous explanation of the following hypothesis: say for the sake of argument that 80% of the cartilage texture in Fig. 2(b) consists of “background” light purple pixels while the remaining 20% of pixels lie in a “foreground” of darker-reddish purple blobs; we then expect a local histogram computed over a portion of cartilage to be a mixture – convex combination – of 0.8 of the background pixels’ distribution with 0.2 of the foreground pixels’ distribution. Other tissues arise from other distinct decompositions. For example, looking at the pseudovascular tissue of Fig. 2(d), we might guess it to be 0.5 light pink, 0.25 dark purple and 0.25 reddish-pink. We rigorously show that such decompositions of local

histograms indeed exist for textures arising from a certain class of probabilistic image models; our long-term goal is to exploit this fact in a segmentation-and-classification algorithm.

To see how to formalize these ideas, it helps to consider a toy example: imagine that at any given pixel location, a coin is flipped, with “heads” resulting in a pink pixel value, and “tails” resulting in a purple one. One expects that, on average, the local histogram at any point will consist of two peaks: one in the pink portion of \mathcal{Y} , and one in the purple. Such an image can be regarded as the result of *occluding* a solid purple image f_0 with a solid pink one f_1 : at each pixel, the flip of a coin determines whether f_1 lies on top of f_0 at that point, or vice versa. More generally, 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:

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

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 main results of this paper 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. Though it is unrealistic to expect a clean relation for any fixed φ , we can show that these quantities are indeed closely related, provided one averages over all possible label functions φ . Indeed, denoting the probability of getting “heads” in the above toy example as $\rho \in [0, 1]$, we would expect the volumes of the pink and purple peaks of the composite image's local histograms to be ρ and $1 - \rho$, respectively. That is, $\text{LH}_w \text{occ}_\varphi\{f_0, f_1\}$ should be $(1 - \rho)\text{LH}_w f_0 + \rho\text{LH}_w f_1$, on average. We generalize this idea so as to permit more realistic textures with more colors and with spatially-correlated pixels.

To be precise, fix a set of source images $\{f_n\}_{n=0}^{N-1}$ and consider the set $\{\text{occ}_\varphi\{f_n\}_{n=0}^{N-1}\}_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)}$ of all possible composite images (2) obtained by letting φ be any one of the $N^{|\mathcal{X}|}$ elements of $\ell(\mathcal{X}, \mathbb{Z}_N)$, where $|\mathcal{X}|$ denotes the cardinality of \mathcal{X} . We refer to a random method for choosing one of these composites as an *occlusion model* Φ . Formally speaking, 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 $\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) = 1$. For example, imagine three 128×128 images f_0, f_1 and f_2 which exhibit a nearly constant shade of pink, purple and red, respectively. Given any label function $\varphi : \mathbb{Z}_{128}^2 \rightarrow \mathbb{Z}_3$ we can produce a corresponding 128×128 composite image $\text{occ}_\varphi\{f_0, f_1, f_2\}$ whose pixels are some mixture of pink, purple and red. For some choices of φ the resulting composites will look like the pseudovascular tissue texture given in Fig. 2(d). However, even in this small example, there are an enormous number of such possible composites – one for each of the 3^{128^2} possibilities for φ – and only a few of these will look like pseudovascular tissue; most will appear as pink-purple-red static. 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. That is, one particular choice of Φ will generate textures that look like pseudovascular tissue (Fig. 2(d)). Another choice of Φ will lead to cartilage-like textures (Fig. 2(b)). Most Φ 's will generate textures that look nothing like biological tissues.

In this paper, we provide a sufficient hypothesis on the occlusion model Φ so as to ensure that the local histograms (1) of a composite image (2) can, on average with respect to P_Φ , be decomposed in terms of the local histograms of the individual images. In particular, we focus on the special case where the 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:

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

That is, Φ is flat if the marginal distributions obtained by fixing any given $x \in \mathcal{X}$ are identical. 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 . In our toy example where the values of φ are determined by $|\mathcal{X}|$ spatially-independent coin flips, the probability of getting any particular $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)$ is $P_\Phi(\varphi) = \rho^{|\varphi^{-1}\{1\}|} (1 - \rho)^{|\mathcal{X}| - |\varphi^{-1}\{1\}|}$; substituting this expression into (3), the binomial theorem implies that this model is indeed flat with $\lambda_0 = 1 - \rho$ and $\lambda_1 = \rho$. Note that, if $\rho > \frac{1}{2}$, the resulting random image $\text{occ}_\Phi\{f_0, f_1\}$ will be more pink than purple; flatness does not mean that each label is equally likely, but rather that the chance of being pink at any given pixel location is the same as at any other location. 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:*

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

From the point of view of our motivating application, the significance of Theorem 1 is that it gives credence to a certain type of segmentation-and-classification algorithm. To be precise, given a set of training images which are manually segmented and labeled by medical experts, we, for any given tissue type, can compute local histograms over regions which are labeled as that type. In light of Theorem 1, it is reasonable to *demix* – decompose into convex combinations – the local histograms of that type into a set of more basic distributions. For example, we expect that the local histograms of pseudovascular tissue (Fig. 2(d)) can be demixed into three simpler distributions – one pink, another purple and a third reddish-pink – while those of connective tissue (Fig. 2(c)) are mixtures of only the first two. Once sparse demixings of each tissue type have been found, we then use them 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 Theorem 1. Specifically, the next section contains several basic results on local histograms. In Section 3, we prove Theorem 1 and also a generalization of it – Theorem 4 – to the non-flat case. In Section 4, we provide various methods – Theorems 5, 6 and 7 – for constructing flat Φ 's, and some of these produce textures that resemble those found in digital microscope images of histological tissues. The final section discusses a preliminary segmentation-and-classification algorithm inspired by Theorem 1 in which local histograms are decomposed using principal component analysis (PCA).

Both local histograms and probabilistic image occlusion models have long been subjects of interest. Theorem 2 below details how local histograms can be computed as systems of convolutions; a similar result is given in [8], and both [8] and [18] discuss how such a computation can be implemented in optical hardware. Recently, local histograms have been used in an active contour-based segmentation scheme [16]; this algorithm partitions an image into two smoothly bounded regions whose pixel values are maximally separated with respect to the Wasserstein (earth mover's) distance. Local histograms have also recently been used as smoothing filters [7]. Though the work we present here focuses exclusively on local histograms of the pixel values themselves, an alternative approach is to first pass the image through a filter bank and then compute histograms of the resulting values [6,12]. Local histograms, like time–frequency transforms, preserve global spatial context while obscuring all local spatial context, and as such they are well-suited to the processing of *locally orderless* images [5,9,10]. We use local histograms to analyze a class of textures generated by a certain probabilistic occlusion model; this model, like the *dead leaves model* [3,11,15], generates these textures via a sequential superposition of random sets. Our contribution to this body of literature is a formalism that unifies the theory of local histograms with that of occlusion models and permits us to rigorously prove that local histograms are indeed a useful transform for the analysis of a particular class of textures.

2. Local histograms

In this section, we discuss an efficient means of computing local histograms (1) and discuss several of their basic properties. 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 is given in Theorem 2 below: (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 $\tilde{w}(x) := w(-x)$. This method for computing local histograms is illustrated in Fig. 3. Alternatively, (1) can be computed as a single convolution over $\mathcal{X} \times \mathcal{Y}$; here, the tensor product of $w \in \ell(\mathcal{X}, \mathbb{R})$ with $\omega \in \ell(\mathcal{Y}, \mathbb{R})$ is defined as $w \otimes \omega \in \ell(\mathcal{X} \times \mathcal{Y}, \mathbb{R})$, $(w \otimes \omega)(x, y) := w(x)\omega(y)$.

Theorem 2. For any $w \in \ell(\mathcal{X}, \mathbb{R})$, $\omega \in \ell(\mathcal{Y}, \mathbb{R})$, $f \in \ell(\mathcal{X}, \mathcal{Y})$, $x \in \mathcal{X}$, and $y \in \mathcal{Y}$:

- (a) Local histograms (1) can be evaluated as a system of $|\mathcal{Y}|$ convolutions over \mathcal{X} : $(\text{LH}_w f)(x, y) = (\tilde{w} * 1_{f^{-1}\{y\}})(x)$.
- (b) Alternatively, (1) may be computed as a single convolution over $\mathcal{X} \times \mathcal{Y}$: $(\delta_0 \otimes \omega) * \text{LH}_w f = (\tilde{w} \otimes \omega) * 1_f$.
In particular, taking $\omega = \delta_0$ gives $\text{LH}_w f = (\tilde{w} \otimes \delta_0) * 1_f$.

Proof. For (a), replacing x' with $-x'$, and substituting the relation $\delta_y(f(x - x')) = 1_{f^{-1}\{y\}}(x - x')$ into (1) yields:

$$\begin{aligned} (\text{LH}_w f)(x, y) &= \sum_{x' \in \mathcal{X}} w(x') \delta_y(f(x + x')) = \sum_{x' \in \mathcal{X}} w(-x') 1_{f^{-1}\{y\}}(x - x') = \sum_{x' \in \mathcal{X}} \tilde{w}(x') 1_{f^{-1}\{y\}}(x - x') \\ &= (\tilde{w} * 1_{f^{-1}\{y\}})(x). \end{aligned} \quad (6)$$

For (b), the definition of δ_0 gives:

$$[(\delta_0 \otimes \omega) * \text{LH}_w f](x, y) = \sum_{(x', y') \in \mathcal{X} \times \mathcal{Y}} (\delta_0 \otimes \omega)(x', y') (\text{LH}_w f)(x - x', y - y') = \sum_{y' \in \mathcal{Y}} \omega(y') (\text{LH}_w f)(x, y - y'). \quad (7)$$

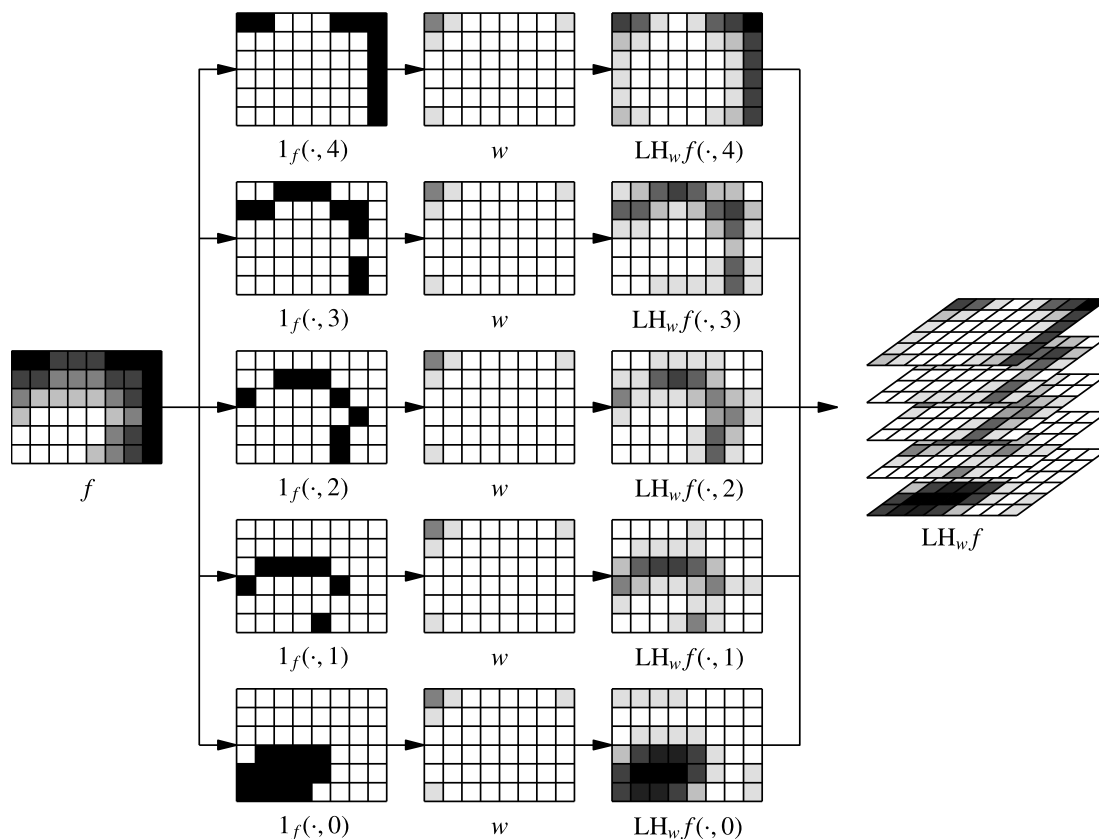


Fig. 3. An example of how to compute local histograms using Theorem 2(a). For the sake of readability, larger numerical values are represented by darker shades throughout. The source image (far left) f is 6×8 and has grayscale values ranging from 0 to 4. That is $f \in \ell(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X} = \mathbb{Z}_6 \times \mathbb{Z}_8$ and $\mathcal{Y} = \mathbb{Z}_5$. Its characteristic function (5) is a $\{0, 1\}$ -valued $6 \times 8 \times 5$ data cube whose cross-sections (left column) indicate those locations at which f attains any given value. By Theorem 2(a), the $6 \times 8 \times 5$ data cube that contains the local histograms of f (far right) may be computed one level at a time (right column) by filtering these binary-valued cross-sections with a real-scalar-valued weighting function (middle column). In this simple example, the weighting function is $w = \frac{1}{2}\delta_{0,0} + \frac{1}{8}(\delta_{-1,0} + \delta_{1,0} + \delta_{0,-1} + \delta_{0,1})$, where the origin lies in the upper left-hand corner of the grid.

Substituting (6) into (7) and using (5), gives our result:

$$\begin{aligned}
 [(\delta_0 \otimes \omega) * \text{LH}_w f](x, y) &= \sum_{y' \in \mathcal{Y}} \omega(y') (\tilde{w} * 1_{f^{-1}\{y-y'\}})(x) \\
 &= \sum_{y' \in \mathcal{Y}} \omega(y') \sum_{x' \in \mathcal{X}} \tilde{w}(x') 1_{f^{-1}\{y-y'\}}(x - x') \\
 &= \sum_{(x', y') \in \mathcal{X} \times \mathcal{Y}} (\tilde{w} \otimes \omega)(x', y') 1_f(x - x', y - y') \\
 &= [(\tilde{w} \otimes \omega) * 1_f](x, y). \quad \square
 \end{aligned}$$

The next result summarizes several other basic properties of local histograms, the proofs of which are given in [13].

Proposition 3. For any $w \in \ell(\mathcal{X}, \mathbb{R})$ and $f \in \ell(\mathcal{X}, \mathcal{Y})$:

- (a) The levels of a local histogram transform sum to 1: for any $x \in \mathcal{X}$, $\sum_{y \in \mathcal{Y}} (\text{LH}_w f)(x, y) = 1$.
- (b) Local histograms commute with spatial translation T^x : for any $x \in \mathcal{X}$, $\text{LH}_w T^x = T^{(x,0)} \text{LH}_w$.
- (c) Adding constants to images shifts their local histograms along \mathcal{Y} : for any $y \in \mathcal{Y}$, $\text{LH}_w(f + y) = T^{(0,y)} \text{LH}_w f$.
- (d) Quantizing an image will bin its local histograms: for any $q \in \ell(\mathcal{Y}, \mathcal{Y}')$,

$$[\text{LH}_w(q \circ f)](x, y') = \sum_{\substack{y \in \mathcal{Y} \\ q(y) = y'}} (\text{LH}_w f)(x, y).$$

This basic understanding of local histograms in hand, we turn to the theory of applying them to textures generated by the probabilistic image occlusion models discussed in the introduction.

3. Local histograms of randomly-generated textures

In this section, we rigorously confirm our intuition regarding local histograms of textures generated via random occlusions: if a texture, such as that found in the pseudovascular tissue of Fig. 2(d), is some sufficiently-spatially-random combination of 50% pink pixels, 25% purple pixels and 25% red pixels, then its local histograms should, on average, be a mixture of three simpler distributions, namely a convex combination of 0.5 of a purely pink distribution with 0.25 purely purple and red ones.

To do this, 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):

$$\bar{1}_\Phi(x, n) := \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) 1_\varphi(x, n) = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N) \\ \varphi(x)=n}} P_\Phi(\varphi). \quad (8)$$

Essentially, $\bar{1}_\Phi(x, n)$ is the probability that a random label function φ generated by the occlusion model Φ will assign label n to pixel location x . 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 4. 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:

$$E_\Phi(\text{LH}_w \text{occ}_\Phi \{f_n\}_{n=0}^{N-1})(x, y) = \sum_{n=0}^{N-1} \bar{1}_\Phi(x, n) (\text{LH}_w f_n)(x, y) + \varepsilon, \quad (9)$$

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,

$$\sum_{n=0}^{N-1} \bar{1}_\Phi(x, n) = 1, \quad (10)$$

and so (9) states that, on average, the local histograms of the composite image $\text{occ}_\Phi \{f_n\}_{n=0}^{N-1}$ can be approximated by convex combinations of local histograms of each individual image f_n .

Proof. The expected value of the local histogram (1) of a composite image (2) is:

$$E_\Phi(\text{LH}_w \text{occ}_\Phi \{f_n\}_{n=0}^{N-1})(x, y) = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) \sum_{x' \in \mathcal{X}} w(x') \delta_y((\text{occ}_\varphi \{f_n\}_{n=0}^{N-1})(x+x')). \quad (11)$$

For any fixed φ , x , and x' , we have $\varphi(x+x') = n$ for exactly one n . For any fixed x , x' and y , we can therefore split a sum of $1_\varphi(x+x', n) \delta_y(f_n(x+x'))$ over all n into one summand where $n = \varphi(x+x')$ and the remaining $N - 1$ summands for which $n \neq \varphi(x+x')$:

$$\begin{aligned} \sum_{n=0}^{N-1} 1_\varphi(x+x', n) \delta_y(f_n(x+x')) &= (1) \delta_y(f_{\varphi(x+x')}(x+x')) + \sum_{n \neq \varphi(x+x')} (0) \delta_y(f_n(x+x')) \\ &= \delta_y((\text{occ}_\varphi \{f_n\}_{n=0}^{N-1})(x+x')), \end{aligned} \quad (12)$$

where the final equality follows immediately from (2). Substituting (12) into (11) and using (8) yields:

$$\begin{aligned} E_\Phi(\text{LH}_w \text{occ}_\Phi \{f_n\}_{n=0}^{N-1})(x, y) &= \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) \sum_{x' \in \mathcal{X}} w(x') \left(\sum_{n=0}^{N-1} 1_\varphi(x+x', n) \delta_y(f_n(x+x')) \right) \\ &= \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x') \delta_y(f_n(x+x')) \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) 1_\varphi(x+x', n) \\ &= \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x') \delta_y(f_n(x+x')) \bar{1}_\Phi(x+x', n). \end{aligned} \quad (13)$$

Rewriting (13) in terms of $\varepsilon := \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x') \delta_y(f_n(x+x')) [\bar{1}_\Phi(x+x', n) - \bar{1}_\Phi(x, n)]$ gives our first claim (9):

$$\begin{aligned} E_\Phi(\text{LH}_w \text{occ}_\Phi \{f_n\}_{n=0}^{N-1})(x, y) &= \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x') \delta_y(f_n(x+x')) \bar{1}_\Phi(x, n) + \varepsilon \\ &= \sum_{n=0}^{N-1} \bar{1}_\Phi(x, n) \sum_{x' \in \mathcal{X}} w(x') \delta_y(f_n(x+x')) + \varepsilon \\ &= \sum_{n=0}^{N-1} \bar{1}_\Phi(x, n) (\text{LH}_w f_n)(x, y) + \varepsilon. \end{aligned}$$

For the second claim, we bound ε using the triangle inequality and the fact that $|\delta_y(f_n(x+x'))| \leq 1$:

$$|\varepsilon| = \left| \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x') \delta_y(f_n(x+x')) [\bar{1}_\Phi(x+x', n) - \bar{1}_\Phi(x, n)] \right| \leq \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x') |\bar{1}_\Phi(x+x', n) - \bar{1}_\Phi(x, n)|.$$

Finally, to prove our third claim (10), note that for any fixed $x \in \mathcal{X}$, (8) gives:

$$\sum_{n=0}^{N-1} \bar{1}_\Phi(x, n) = \sum_{n=0}^{N-1} \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) 1_\varphi(x, n) = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) \sum_{n=0}^{N-1} \begin{cases} 1, & \varphi(x) = n, \\ 0, & \varphi(x) \neq n. \end{cases} \quad (14)$$

Since as previously noted we have $\varphi(x) = n$ for exactly one n , (14) becomes: $\sum_{n=0}^{N-1} \bar{1}_\Phi(x, n) = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) = 1$. \square

An example illustrating the direct computation of the left-hand side of (9) is given in Fig. 4. Note that Theorem 4 implies that the error term ε in (9) will be small provided the probability $\bar{1}_\Phi(x, n)$ of assigning label n to x changes little as x varies over regions smaller than the support of w . The extreme case of this is when the occlusion model Φ is flat, meaning $\bar{1}_\Phi(x, n)$ is constant with respect to x . In this case, ε vanishes entirely, leading to Theorem 1 as given in the introduction:

Proof of Theorem 1. If Φ is flat, $\bar{1}_\Phi(x+x', n) = \bar{1}_\Phi(x, n)$ for all $x, x' \in \mathcal{X}$. The error bound in Theorem 4 then gives $\varepsilon = 0$. Denoting $\bar{1}_\Phi(x, n)$ as λ_n in (9) thus yields our claim. \square

That is, when Φ is flat, (9) simplifies to (4), and so the in-depth computation of Fig. 4 can be replaced by the much simpler one depicted in Fig. 5. 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; this is the topic of the next section.

4. Flat occlusion models

Theorem 1 gives some insight into the behavior of the local histograms of images generated via random occlusions. However, this result only holds when Φ is flat (3), namely when its average characteristic function $\bar{1}_\Phi(x, n)$, as defined in (8), is constant with respect to pixel location x , but is still permitted to vary with label value n . In this section, we demonstrate that flatness is a reasonable assumption. In particular, we provide a variety of methods for constructing flat occlusion models. Some of these models produce textures similar to those encountered in digital microscope images of histological tissues. Our first method involves the translation operator $T^x : \mathcal{X} \rightarrow \mathcal{X}$, $T^x \varphi(x') := \varphi(x' - x)$. To be precise, we show that an occlusion model Φ is flat if it is *translation-invariant*, meaning that its probability density function P_Φ satisfies:

$$P_\Phi(T^x \varphi) = P_\Phi(\varphi), \quad \forall \varphi \in \ell(\mathcal{X}, \mathbb{Z}_N), x \in \mathcal{X}. \quad (15)$$

Theorem 5. If Φ is translation-invariant (15), then Φ is flat (3).

Proof. Fix any $x, x' \in \mathcal{X}$ and any $n \in \mathbb{Z}_N$. Note that the set $\{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N) : \varphi(x) = n\}$ of all functions φ that assign label n at pixel location x is merely a translate of the set $\{\psi \in \ell(\mathcal{X}, \mathbb{Z}_N) : \psi(x') = n\}$ of all functions ψ that assign n at x' . In particular, any such φ can be written uniquely in terms of such a ψ as $\varphi = T^{x-x'} \psi$. Making this change of variables, our assumption of translation-invariance gives:

$$\sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N) \\ \varphi(x)=n}} P_\Phi(\varphi) = \sum_{\substack{T^{x-x'} \psi \in \ell(\mathcal{X}, \mathbb{Z}_N) \\ (T^{x-x'} \psi)(x)=n}} P_\Phi(T^{x-x'} \psi) = \sum_{\substack{\psi \in \ell(\mathcal{X}, \mathbb{Z}_N) \\ \psi(x')=n}} P_\Phi(\psi).$$

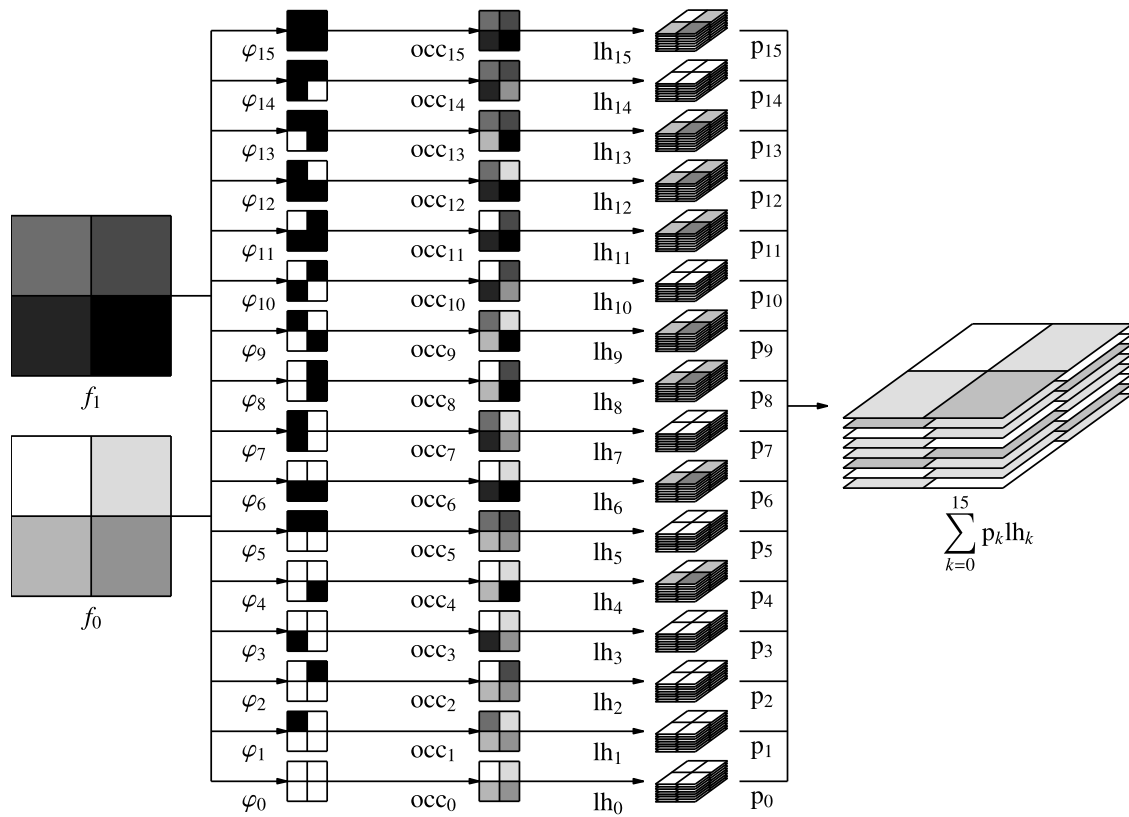


Fig. 4. An example of how to compute the left-hand side of (9) explicitly as a probability-weighted sum. For the sake of readability, larger numerical values are represented by darker shades throughout. We consider two 2×2 , 3-bit images, namely $\{f_n\}_{n=0}^{N-1}$ in $\ell(\mathcal{X}, \mathcal{Y})$ where $N = 2$, $\mathcal{X} = \mathbb{Z}_2 \times \mathbb{Z}_2$ and $\mathcal{Y} = \mathbb{Z}_8$. In this particular example, the values of the f_n 's are all distinct, with f_0 assuming values $\{0, 1, 2, 3\}$ and f_1 assuming values $\{4, 5, 6, 7\}$ (far left). There are $N^{|\mathcal{X}|} = 2^{2^2} = 16$ distinct label functions $\varphi: \mathbb{Z}_2 \times \mathbb{Z}_2 \rightarrow \mathbb{Z}_2$ (left column) each yielding a composite image $\text{occ}_\varphi\{f_0, f_1\}$ (center column); in accordance with (2), we take values from f_0 in places where φ is white and values from f_1 where φ is black. Each of these composites has a $2 \times 2 \times 8$ local histogram transform (1) (right column). Since occlusion (2) is nonlinear, there is no clean relationship between the local histograms of any single composite and the local histograms of the source images f_0 and f_1 . Nevertheless, under certain hypotheses, we can say something about the average of these local histograms (far right) with respect to some probability density function P_Φ on the set $\ell(\mathbb{Z}_2 \times \mathbb{Z}_2, \mathbb{Z}_2)$ of all possible φ 's. In particular, if the occlusion model Φ is flat (3), meaning in this case that the probability-weighted-sum of all φ 's is a constant function of x , then Theorem 1 states that this average is a convex combination of the local histograms of f_0 and f_1 as depicted in Fig. 5.

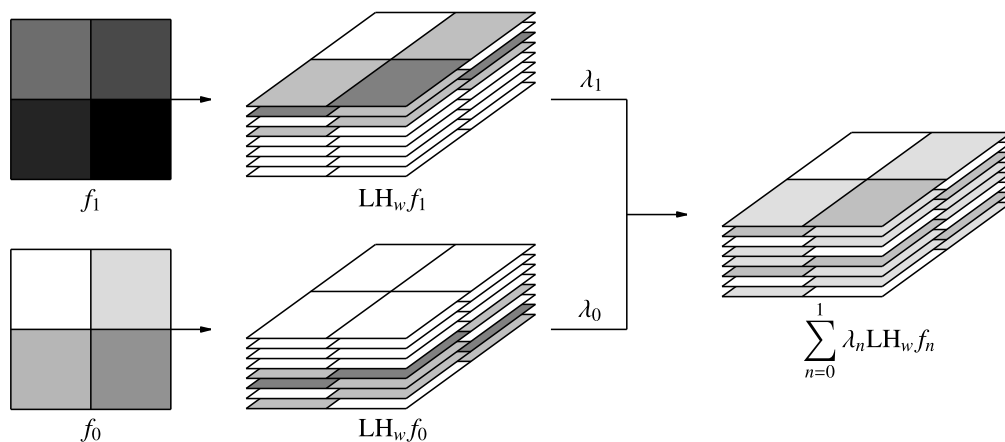


Fig. 5. A continuation of the example of Fig. 4. When the occlusion model Φ is flat, Theorem 4 becomes Theorem 1, with (9) simplifying to (4). Though each of the 16 distinct composite images shown in Fig. 4 has a distinct local histogram transform, the average of these 16 local histogram transforms with respect to P_Φ is but a convex combination (right) of the local histograms (center) of the two source images (left). That is, when Φ is flat, the average-over-all-composites local histogram computed in Fig. 4 is equal to the average-over-all-sources local histogram computed above.

Thus, we see that the quantity on the left-hand side of (3) is indeed independent of x , as claimed. We thank one of the anonymous reviewers of this manuscript for suggesting this proof. \square

Theorem 5 indicates that flatness is not too strong of an assumption. Indeed, one method for producing a flat model Φ is to generalize the coin-flipping example given in the introduction: given any random method for picking a number from \mathbb{Z}_N – a probability spinner – produce φ by conducting $|\mathcal{X}|$ independent spins. The resulting model Φ is translation-invariant, and therefore flat, since $P_\Phi(\varphi)$ is solely determined by the number of times that φ achieves each given value n . Other translation-invariant examples abound. For instance, for any fixed φ_0 , we can assign equal probability $\frac{1}{|\mathcal{X}|}$ to φ_0 and each of its translates, and assign probability 0 to all others; if the source images $\{f_n\}_{n=0}^{N-1}$ are constant, the composite images (2) produced by such a model are all translates of a single image. More generally, we can always partition the $N^{|\mathcal{X}|}$ elements of $\ell(\mathcal{X}, \mathbb{Z}_N)$ into translation-invariant equivalence classes and assign any fixed probability to the members of each class, provided we ensure that in the end they all sum to one. For example, for the case $N = 2$ and $\mathcal{X} = \mathbb{Z}_2 \times \mathbb{Z}_2$ depicted in Fig. 4, we may partition the 16 possible φ 's into 7 such classes, and pick any probabilities $\{p_k\}_{k=0}^{15}$ such that $p_1 = p_2 = p_3 = p_4$, $p_5 = p_6$, $p_7 = p_8$, $p_9 = p_{10}$, $p_{11} = p_{12} = p_{13} = p_{14}$. Armed with one method – translation-invariance – for producing flat models Φ , we now turn to ways of combining known models to produce more complicated and realistic ones.

4.1. Expansion

Digital microscope images of histological tissues often contain randomly distributed blobs. These blobs correspond to biological structures: cells, nuclei, etc. The nature of these processes guarantees that the distribution of such structures is roughly uniform, both spatially and in terms of color: two cells cannot occupy the same space; cells will usually grow and reproduce so as to occupy any empty space; cells in a given tissue all have approximately the same size and color patterns. We want to construct flat occlusion models that emulate such textures, since in light of Theorem 1, doing so would formally justify the demixing of local histograms as part of a segmentation-and-classification algorithm. Note that there is a natural method for randomly generating a set of roughly uniformly-distributed points: flip a coin at each point x . Here, we explore the idea of *expanding* each of these randomly generated points into a given blob.

To be precise, let $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)$ indicate a set of randomly generated points. For each of the points $x \in \mathcal{X}$ for which $\varphi(x) = 1$, we will replace it with a blob whose shape is indicated by some $\psi_x \in \ell(\mathcal{X}, \mathbb{Z}_2)$. The new texture will be the union of all such blobs. Formally, given any $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)$ and $\{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}$, we define the *expansion* of φ by $\{\psi_x\}_{x \in \mathcal{X}}$ to be $\varphi \star \{\psi_x\}_{x \in \mathcal{X}} \in \ell(\mathcal{X}, \mathbb{Z}_2)$,

$$(\varphi \star \{\psi_{x'}\}_{x' \in \mathcal{X}})(x) := \begin{cases} 1, & x = x' + x'', \varphi(x') = 1, \psi_{x''}(x'') = 1, \\ 0, & \text{else.} \end{cases} \tag{16}$$

Two examples of this expansion operation are given in Fig. 6. Note that expansion itself (16) is not an occlusion model. Indeed, (16) is but a way of combining functions in $\ell(\mathcal{X}, \mathbb{Z}_2)$ to produce other ones, whereas an occlusion model is a random variable Φ defined by a probability density function P_Φ over $\ell(\mathcal{X}, \mathbb{Z}_2)$. This fact notwithstanding, the expansion operation (16) on label functions φ and $\{\psi_x\}_{x \in \mathcal{X}}$ does in fact induce a parallel operation on their random variable cousins Φ and Ψ . To be precise, given two occlusion models Φ and Ψ from \mathcal{X} into \mathbb{Z}_2 , we define the expansion of Φ by Ψ to be the occlusion model $\Phi \star \Psi$ whose probability density function is $P_{\Phi \star \Psi} : \ell(\mathcal{X}, \mathbb{Z}_2) \rightarrow [0, 1]$,

$$P_{\Phi \star \Psi}(\sigma) := \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}} \\ \varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sigma}} P_\Phi(\varphi) \prod_{x \in \mathcal{X}} P_\Psi(\psi_x). \tag{17}$$

Note that the probability that $\Phi \star \Psi$ will produce a given label function σ depends on the ways in which σ can be written as $\varphi \star \{\psi_x\}_{x \in \mathcal{X}}$ and, moreover, the probability that Φ and Ψ will produce those particular φ 's and ψ_x 's, respectively. In the next result, we verify that (17) indeed defines a probability density function on $\ell(\mathcal{X}, \mathbb{Z}_2)$. We further show that if Φ is translation-invariant (15), then $\Phi \star \Psi$ is translation-invariant which implies that $\Phi \star \Psi$ is flat by Theorem 5. In particular, image models which produce collections of blobs similar to those found in biological tissues will indeed be flat provided the distribution that produces the “centers” of these blobs is translation-invariant. Moreover, if the flatness of $\Phi \star \Psi$ is all that is desired, we can weaken the requirement that Φ be translation-invariant so as to only require that Φ is itself flat, provided Φ and Ψ are *effectively disjoint*:

$$\text{If } P_\Phi(\varphi) > 0 \text{ and } P_\Psi(\psi_x) > 0 \text{ for all } x \in \mathcal{X}, \text{ then } \varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sum_{\substack{x \in \mathcal{X} \\ \varphi(x)=1}} T^x \psi_x. \tag{18}$$

Put another way, (18) means that there is only at most one way, with nontrivial probability, in which the x in (16) can be written as $x = x' + x''$ where both $\varphi(x') = 1$ and $\psi_{x''}(x'') = 1$.

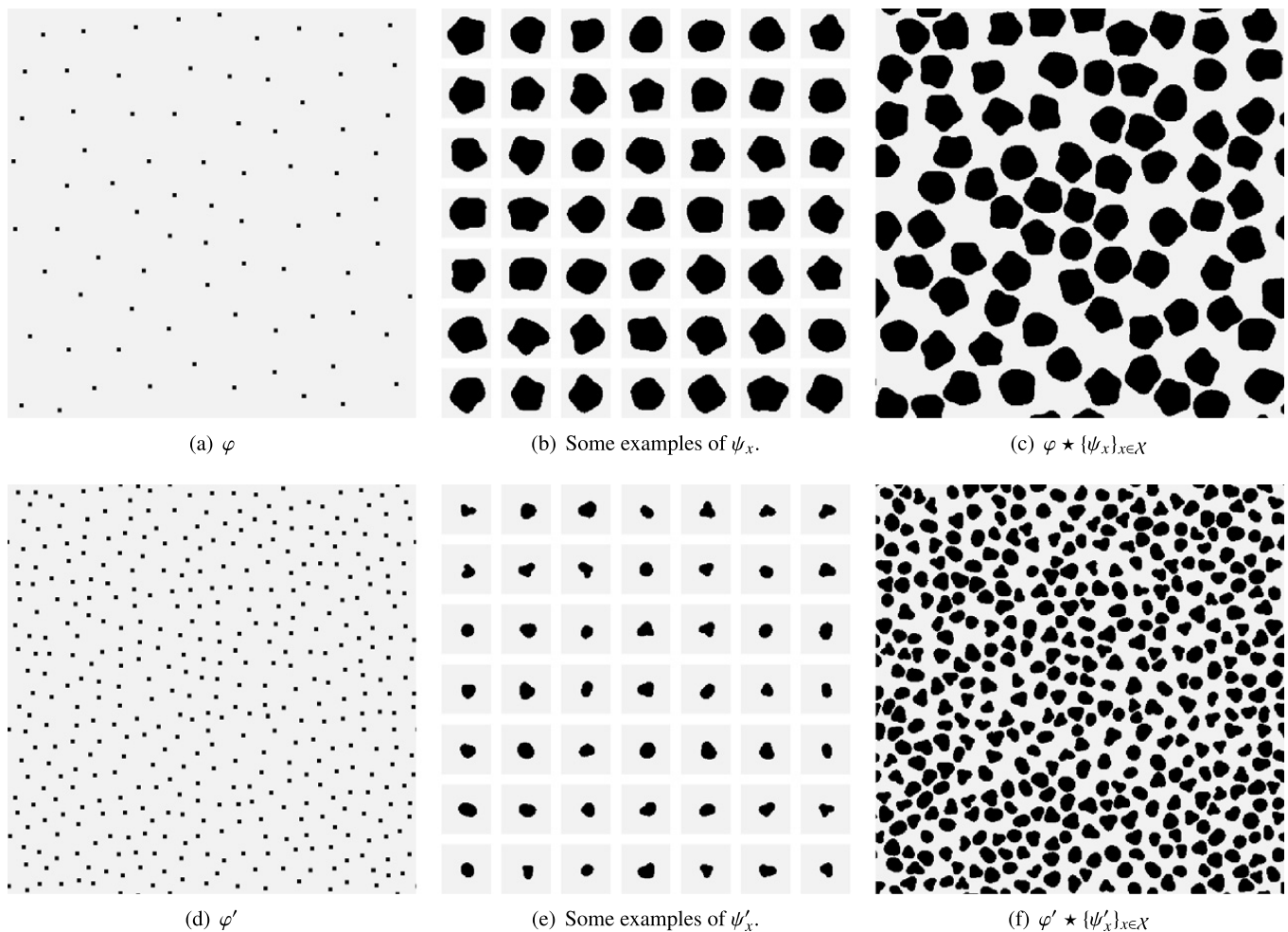


Fig. 6. Examples of the expansion operation (16), where black denotes the value of 1, and the lighter shade denotes the value of 0. A function $\varphi : \mathcal{X} \rightarrow \{0, 1\}$ is given in (a), and can be chosen, for example, via a sequence of $|\mathcal{X}|$ independent coin flips. Meanwhile, for each $x \in \mathcal{X}$, we pick a corresponding function $\psi_x : \mathcal{X} \rightarrow \{0, 1\}$. Cropped versions of a few examples of such ψ_x 's are given in (b). The expansion $\varphi \star \{\psi_x\}_{x \in \mathcal{X}}$ of φ by $\{\psi_x\}_{x \in \mathcal{X}}$ is given in (c). Essentially, each point x for which $\varphi(x) = 1$ is replaced with the corresponding blob ψ_x , with the origin of the ψ_x coordinates being translated to x . In the second row, (f) shows the expansion of a second set of points φ' by a second set of blobs $\{\psi'_x\}_{x \in \mathcal{X}}$. These examples notwithstanding, note that (16) does not require these blobs to be disjoint. We could have, for instance, produced a texture by expanding the points in (d) by the blobs in (b). Nevertheless, stronger conclusions can be made if such disjointness is enforced; see Theorem 6.

Theorem 6. *If Φ and Ψ are occlusion models from \mathcal{X} into \mathbb{Z}_2 , then their expansion $\Phi \star \Psi$, with probability density function (17), is as well. Moreover, if Φ is translation-invariant (15), then $\Phi \star \Psi$ is translation-invariant. Furthermore, if Φ and Ψ are effectively disjoint (18) and either Φ or Ψ is flat (3), then $\Phi \star \Psi$ is flat.*

Proof. We first show that (17) defines a probability density function, namely that values of $P_{\Phi \star \Psi}(\sigma)$ over all σ in $\ell(\mathcal{X}, \mathbb{Z}_2)$ sum to 1. Since P_Φ is a probability density function by assumption, we have:

$$1 = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_\Phi(\varphi). \tag{19}$$

Similarly, for any fixed $x \in \mathcal{X}$, we have:

$$1 = \sum_{\psi_x \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_\Psi(\psi_x), \tag{20}$$

where the subscript “ x ” on ψ indicates that this particular ψ is intended to expand φ at the particular point x as opposed to at some other point. Taking the product of (19) with the product of (20) over all x yields:

$$1 = 1(1)^{|\mathcal{X}|} = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_\Phi(\varphi) \prod_{x \in \mathcal{X}} \sum_{\psi_x \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_\Psi(\psi_x) = \sum_{\{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^\mathcal{X}} P_\Phi(\varphi) \prod_{x \in \mathcal{X}} P_\Psi(\psi_x), \tag{21}$$

where the final quantity in (21) contains all of the cross terms resulting from distributing the product over all sums of the form (20). Now, since for each choice of φ and $\{\psi_x\}_{x \in \mathcal{X}}$ there is exactly one resulting $\sigma = \varphi \star \{\psi_x\}_{x \in \mathcal{X}}$, we can rewrite (21) in terms of the definition (17) of $P_{\Phi \star \Psi}$, obtaining our claim:

$$1 = \sum_{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)} \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}} \\ \varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sigma}} P_{\Phi}(\varphi) \prod_{x \in \mathcal{X}} P_{\Psi}(\psi_x) = \sum_{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi \star \Psi}(\sigma).$$

Thus, (17) indeed defines a probability density function, as claimed.

We next show that the occlusion model $\Phi \star \Psi$ is translation-invariant, if Φ is translation-invariant. To do this, we claim that if $T^{\tilde{x}}\sigma = \varphi \star \{\psi_x\}_{x \in \mathcal{X}}$ then $\sigma = (T^{-\tilde{x}}\varphi) \star \{\psi_{x+\tilde{x}}\}_{x \in \mathcal{X}}$. To see this claim, note that

$$\sigma(x - \tilde{x}) = (T^{\tilde{x}}\sigma)(x) = (\varphi \star \{\psi_{x'}\}_{x' \in \mathcal{X}})(x) = 1$$

if and only if there exists some x', x'' in \mathcal{X} such that $x = x' + x''$, $\varphi(x') = 1$, and $\psi_{x'}(x'') = 1$. Letting $\hat{x} = x - \tilde{x}$, we thus have that $\sigma(\hat{x}) = 1$ if and only if $\hat{x} = (x' - \tilde{x}) + x''$, where $(T^{-\tilde{x}}\varphi)(x' - \tilde{x}) = \varphi(x' - \tilde{x} + \tilde{x}) = \varphi(x') = 1$ and $\psi_{(x' - \tilde{x}) + \tilde{x}}(x'') = \psi_{x'}(x'') = 1$, implying $\sigma = (T^{-\tilde{x}}\varphi) \star \{\psi_{x+\tilde{x}}\}_{x \in \mathcal{X}}$, as claimed. Having the claim, (17) implies:

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

To continue, we make the change of variables $\varphi' := T^{-\tilde{x}}\varphi$ and $\psi'_x := \psi_{x+\tilde{x}}$:

$$P_{\Phi \star \Psi}(T^{\tilde{x}}\sigma) = \sum_{\substack{\varphi' \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi'_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}} \\ (\varphi') \star \{\psi'_x\}_{x \in \mathcal{X}} = \sigma}} P_{\Phi}(T^{\tilde{x}}\varphi') \prod_{x \in \mathcal{X}} P_{\Psi}(\psi'_{x-\tilde{x}}).$$

Since Φ is translation-invariant and $\prod_{x \in \mathcal{X}} P_{\Psi}(\psi'_{x-\tilde{x}}) = \prod_{x \in \mathcal{X}} P_{\Psi}(\psi'_x)$, we have:

$$P_{\Phi \star \Psi}(T^{\tilde{x}}\sigma) = \sum_{\substack{\varphi' \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi'_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}} \\ (\varphi') \star \{\psi'_x\}_{x \in \mathcal{X}} = \sigma}} P_{\Phi}(\varphi') \prod_{x \in \mathcal{X}} P_{\Psi}(\psi'_x) = P_{\Phi \star \Psi}(\sigma),$$

and so $\Phi \star \Psi$ is indeed translation-invariant (15), as claimed.

For our final claim, we assume that Φ and Ψ are effectively disjoint (18) and that either Φ or Ψ is flat. To do so, it is helpful to characterize the flatness of an arbitrary occlusion model Φ from \mathcal{X} to \mathbb{Z}_2 in terms of the corresponding function $\bar{\Phi} := \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi}(\varphi)\varphi$. Indeed, for any $\varphi : \mathcal{X} \rightarrow \mathbb{Z}_2$, (5) may be rewritten as $1_{\varphi}(x, 1) = \varphi(x)$ and so:

$$\bar{1}_{\Phi}(x, 1) = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi}(\varphi)1_{\varphi}(x, 1) = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi}(\varphi)\varphi(x) = \bar{\Phi}(x). \tag{22}$$

In light of (22), we claim that Φ is flat if and only if $\bar{\Phi}$ is constant. Indeed, if Φ is flat, then there exists λ_1 such that $\bar{\Phi}(x) = \bar{1}_{\Phi}(x, 1) = \lambda_1$ for all $x \in \mathcal{X}$. Conversely, if $\bar{\Phi}(x)$ is constant, then there exists λ_1 such that $\bar{1}_{\Phi}(x, 1) = \bar{\Phi}(x) = \lambda_1$ for all $x \in \mathcal{X}$; by (10), this further implies that $\bar{1}_{\Phi}(x, 0) = 1 - \bar{1}_{\Phi}(x, 1) = 1 - \lambda_1$ for all $x \in \mathcal{X}$ and so Φ is flat.

Having this claim, we show that $\Phi \star \Psi$ is flat by showing that $\overline{\Phi \star \Psi}$ is constant. To do this, we show that if Φ and Ψ are effectively disjoint then $\overline{\Phi \star \Psi} = \bar{\Phi} \star \bar{\Psi}$ where “ \star ” denotes standard convolution over \mathcal{X} . According to the definition of $\Phi \star \Psi$ (17) we have:

$$\overline{\Phi \star \Psi} = \sum_{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi \star \Psi}(\sigma)\sigma = \sum_{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)} \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}} \\ \varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sigma}} P_{\Phi}(\varphi) \left(\prod_{x \in \mathcal{X}} P_{\Psi}(\psi_x) \right) (\varphi \star \{\psi_x\}_{x \in \mathcal{X}}). \tag{23}$$

Since any particular choice of φ and $\{\psi_x\}_{x \in \mathcal{X}}$ produces a unique σ via \star we can simplify (23) to

$$\overline{\Phi \star \Psi} = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}}} P_{\Phi}(\varphi) \left(\prod_{x \in \mathcal{X}} P_{\Psi}(\psi_x) \right) (\varphi \star \{\psi_x\}_{x \in \mathcal{X}}). \tag{24}$$

Moreover, since Φ and Ψ are effectively disjoint (18) we have $\varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} T^{x'} \psi_{x'}$ meaning (24) becomes:

$$\begin{aligned} \overline{\Phi \star \Psi} &= \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}}} P_{\Phi}(\varphi) \left(\prod_{x \in \mathcal{X}} P_{\Psi}(\psi_x) \right) \left(\sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} T^{x'} \psi_{x'} \right) \\ &= \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi}(\varphi) \sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} T^{x'} \left[\sum_{\{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}} \left(\prod_{x \in \mathcal{X}} P_{\Psi}(\psi_x) \right) \psi_{x'} \right]. \end{aligned} \quad (25)$$

Now, for any fixed $x' \in \mathcal{X}$ such that $\varphi(x') = 1$, we factor the corresponding innermost sum in (25) into a product of $|\mathcal{X}'|$ distinct sums – one for each $x \in \mathcal{X}'$ – to obtain:

$$\sum_{\{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}} \left(\prod_{x \in \mathcal{X}} P_{\Psi}(\psi_x) \right) \psi_{x'} = \left[\prod_{x \neq x'} \left(\sum_{\psi_x \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Psi}(\psi_x) \right) \right] \sum_{\psi_{x'} \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Psi}(\psi_{x'}) \psi_{x'} = \left(\prod_{x \neq x'} 1 \right) \overline{\Psi} = \overline{\Psi}. \quad (26)$$

Substituting (26) into (25) then gives:

$$\overline{\Phi \star \Psi} = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi}(\varphi) \sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} T^{x'} \overline{\Psi} = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi}(\varphi) \left(\sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} \delta_{x'} \right) * \overline{\Psi} = \left(\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi}(\varphi) \varphi \right) * \overline{\Psi} = \overline{\Phi} * \overline{\Psi}.$$

Thus, the effective disjointness of Φ and Ψ indeed implies $\overline{\Phi \star \Psi} = \overline{\Phi} * \overline{\Psi}$. As such, if we further assume that either Φ or Ψ is flat, then either $\overline{\Phi}$ or $\overline{\Psi}$ is constant, implying in either case that $\overline{\Phi \star \Psi}$ is constant and so $\Phi \star \Psi$ is flat. \square

4.2. Overlay

Above, we discussed how the expansion (17) of a binary-valued occlusion model Φ with another such model Ψ is a new model $\Phi \star \Psi$ that randomly generates label functions of the form $\sigma = \varphi \star \{\psi_x\}_{x \in \mathcal{X}}$ as defined in (16). Under certain hypotheses, Theorem 6 gives that such models $\Phi \star \Psi$ are flat, meaning their local histograms can be understood in terms of Theorem 1. Moreover, some examples of these models produce textures that resemble those encountered in histological tissues: if f_0 and f_1 are roughly constant light purple and dark purple fields, respectively, then the composite image $\text{occ}_{\varphi}\{f_0, f_1\}$ obtained by picking φ as in Fig. 6(f) bears some similarity to an actual image of cartilage, such as the one given in Fig. 2(b). Taken together, these facts provide some theoretical justification for the use of local histograms for the analysis of such tissues.

There is however a deficit with this theory: due to the nature of the construction (16), models produced by expansion (17) can only be binary-valued, and as such are insufficient to emulate textures that exhibit three or more distinct color modes, such as the pseudovascular tissue depicted in Fig. 2(d). In this subsection, we discuss a method for *laying* one occlusion model over another which, amongst other things, permits us to build multivalued models out of binary-valued ones. To be precise, for any $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_{\varphi}})$, $\psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_{\psi}})$ and $\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)$, we define the *overlay* of φ over ψ with respect to σ to be $\varphi \#_{\sigma} \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_{\varphi} + N_{\psi}})$,

$$(\varphi \#_{\sigma} \psi)(x) := \begin{cases} \varphi(x), & \sigma(x) = 0, \\ \psi(x) + N_{\varphi}, & \sigma(x) = 1. \end{cases} \quad (27)$$

Essentially, an overlay (27) is the result of cutting holes out of an image of φ and laying it on top of an image of ψ ; the location of these holes is indicated by σ and the values of ψ are increased by a factor of N_{φ} so that they cannot be confused with those of φ . Examples of this overlay operation are given in Fig. 7.

In a manner similar to the relationship between (16) and (17), we have that (27) naturally induces a parallel operation on occlusion models: 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, we define the overlay of the occlusion model Φ over Ψ with respect to Σ to be the new occlusion model $\Phi \#_{\Sigma} \Psi$ whose probability density function is $P_{\Phi \#_{\Sigma} \Psi} : \ell(\mathcal{X}, \mathbb{Z}_{N_{\Phi} + N_{\Psi}}) \rightarrow [0, 1]$,

$$P_{\Phi \#_{\Sigma} \Psi}(v) := \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_{\Phi}}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_{\Psi}}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \varphi \#_{\sigma} \psi = v}} P_{\Phi}(\varphi) P_{\Psi}(\psi) P_{\Sigma}(\sigma). \quad (28)$$

In the next result, we verify that (28) indeed defines a probability density function, and moreover that the corresponding model $\Phi \#_{\Sigma} \Psi$ is flat provided Φ , Ψ and Σ are flat, meaning that the local histograms (1) of composite images (2) produced by such a model will behave according to Theorem 1.

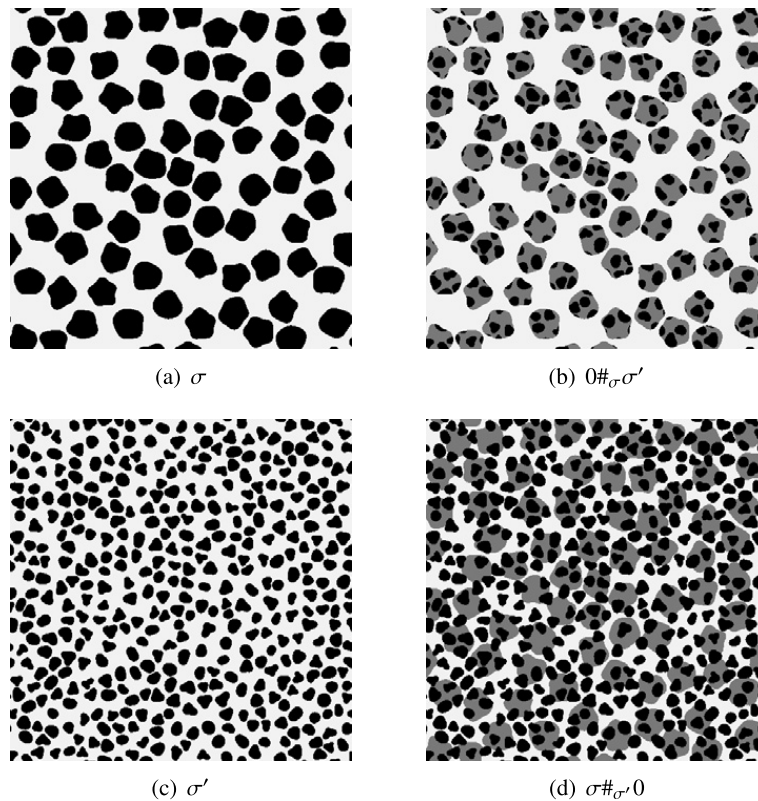


Fig. 7. Two examples of the overlay operation (27). Recall the two $\{0, 1\}$ -valued label functions $\sigma = \varphi \star \{\psi_x\}_{x \in \mathcal{X}}$ and $\sigma' = \varphi' \star \{\psi'_x\}_{x \in \mathcal{X}}$ of Fig. 6(c) and (f) reshown here in (a) and (c), respectively. Further consider a constant function $0 : \mathcal{X} \rightarrow \mathbb{Z}_1$ that assigns label 0 to every point in \mathcal{X} . The overlay (27) of 0 over σ' is given in (b); essentially, σ' -shaped holes are cut from 0 and the result is laid over σ' , resulting in a new texture. A distinct texture can be produced by cutting σ' -shaped holes out from σ and laying the result over the constant function 0 (d). Overlaying the resulting textures with each other can produce even more complicated textures. In Theorem 7, we consider randomly generated versions of (b) and (d) arising from randomly generated versions of (a) and (c). In essence, we show that if the models that generate (a) and (c) are flat, then the induced models that generate (b) and (d) are also flat, meaning that Theorem 1 applies to them. In terms of our motivating segmentation application, the important fact here is that the model that generates (b) is distinct from the model that generates (d) – the flatness constants for (b) are different than those of (d) – and so by Theorem 7, the local histograms of (b)-generated composites are distinct from the those of (d)-generated composites. This means that local histograms can serve to discriminate regions in an image that look like (b) from those that look like (d).

Theorem 7. If Φ, Ψ and Σ are occlusion models on $\ell(X, \mathbb{Z}_{N_\Phi}), \ell(X, \mathbb{Z}_{N_\Psi})$ and $\ell(X, \mathbb{Z}_2)$, respectively, then (28) defines a probability density function on $\ell(\mathcal{X}, \mathbb{Z}_{N_\Phi+N_\Psi})$. Moreover, if Φ, Ψ , and Σ are flat, then $\Phi \#_\Sigma \Psi$ is flat.

Proof. To show that (28) defines a probability density function on $\ell(\mathcal{X}, \mathbb{Z}_{N_\Phi+N_\Psi})$, note that:

$$\begin{aligned}
 1 &= (1)(1)(1) = \left(\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Phi})} P_\Phi(\varphi) \right) \left(\sum_{\psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Psi})} P_\Psi(\psi) \right) \left(\sum_{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_\Sigma(\sigma) \right) \\
 &= \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Phi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)}} P_\Phi(\varphi) P_\Psi(\psi) P_\Sigma(\sigma).
 \end{aligned} \tag{29}$$

Noting that for each fixed φ, ψ , and σ , there exists exactly one $\nu \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Phi+N_\Psi})$ such that $\varphi \#_\sigma \psi = \nu$, (29) becomes:

$$1 = \sum_{\nu \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Phi+N_\Psi})} \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Phi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \varphi \#_\sigma \psi = \nu}} P_\Phi(\varphi) P_\Psi(\psi) P_\Sigma(\sigma) = \sum_{\nu \in \ell(\mathcal{X}, \mathbb{Z}_{N_\Phi+N_\Psi})} P_{\Phi \#_\Sigma \Psi}(\nu),$$

as claimed. For the second conclusion, assume that Φ, Ψ , and Σ are flat. Our goal is to show that $\Phi \#_\Sigma \Psi$ is flat (3), meaning that for any $n \in \mathbb{Z}_{N_\Phi+N_\Psi}$, we want to show that there exists a scalar λ_n such that:

$$\sum_{\substack{v \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ v(x) = n}} P_{\Phi \#_\Sigma \Psi}(v) = \lambda_n \tag{30}$$

for all $x \in \mathcal{X}$. To see this, note that for any such x and n , we have:

$$\begin{aligned} \sum_{\substack{v \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ v(x) = n}} P_{\Phi \#_\Sigma \Psi}(v) &= \sum_{\substack{v \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ v(x) = n}} \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \varphi \#_\sigma \psi = v}} P_\Phi(\varphi) P_\Psi(\psi) P_\Sigma(\sigma) \\ &= \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ (\varphi \#_\sigma \psi)(x) = n}} P_\Phi(\varphi) P_\Psi(\psi) P_\Sigma(\sigma). \end{aligned} \tag{31}$$

Now, in the special case where $n = 0, \dots, N_\varphi - 1$, (27) gives that $(\varphi \#_\sigma \psi)(x) = n$ if and only if $\varphi(x) = n$ and $\sigma(x) = 0$. As such, in this case (31) becomes:

$$\begin{aligned} \sum_{\substack{v \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ v(x) = n}} P_{\Phi \#_\Sigma \Psi}(v) &= \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}), \varphi(x) = n \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2), \sigma(x) = 0}} P_\Phi(\varphi) P_\Psi(\psi) P_\Sigma(\sigma) \\ &= \left(\sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \varphi(x) = n}} P_\Phi(\varphi) \right) \left(\sum_{\psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi})} P_\Psi(\psi) \right) \left(\sum_{\substack{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \sigma(x) = 0}} P_\Sigma(\sigma) \right) \\ &= \lambda_{\Phi, n} \lambda_{\Sigma, 0}. \end{aligned} \tag{32}$$

If, on the other hand $n = N_\varphi, \dots, N_\varphi + N_\psi - 1$ then (27) gives that $(\varphi \#_\sigma \psi)(x) = n$ if and only if $\psi(x) = n - N_\varphi$ and $\sigma(x) = 1$. In this case, (31) becomes:

$$\begin{aligned} \sum_{\substack{v \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ v(x) = n}} P_{\Phi \#_\Sigma \Psi}(v) &= \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}), \psi(x) = n - N_\varphi \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2), \sigma(x) = 1}} P_\Phi(\varphi) P_\Psi(\psi) P_\Sigma(\sigma) \\ &= \left(\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi})} P_\Phi(\varphi) \right) \left(\sum_{\substack{\psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \psi(x) = n - N_\varphi}} P_\Psi(\psi) \right) \left(\sum_{\substack{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \sigma(x) = 1}} P_\Sigma(\sigma) \right) \\ &= \lambda_{\Psi, n - N_\varphi} \lambda_{\Sigma, 1}. \end{aligned} \tag{33}$$

Thus, for any $x \in \mathcal{X}$ we either have (32) or (33) meaning $\Phi \#_\Sigma \Psi$ is flat (30), as claimed. \square

5. A local histogram-based segmentation-and-classification algorithm

In this section, we present a proof-of-concept segmentation-and-classification scheme that is inspired by Theorem 1. We emphasize that for the algorithm presented here, 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. For algorithms intended for real-world use, such 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 [1]; these facts are not reprinted here.

The concept of Theorem 1 is that the local histograms of certain textures should, on the whole, be able to be decomposed in terms of more basic distributions. Indeed, it is reasonable to expect a local histogram computed over a region of cartilage (Fig. 2(b)) to be a mixture of 0.8 of a “light purple” distribution – a distribution mostly supported in portions of \mathcal{Y} that correspond to light purple – with 0.2 of a darker reddish-purple one. Meanwhile, local histograms of other tissues will correspond to distinct mixtures of other distributions. For example, local histograms computed over a region of pseudo-vascular tissue (Fig. 2(d)) might be a mixture of 0.5 of a light pink distribution, with 0.25 of a dark purple one and 0.25 of a reddish-pink one.

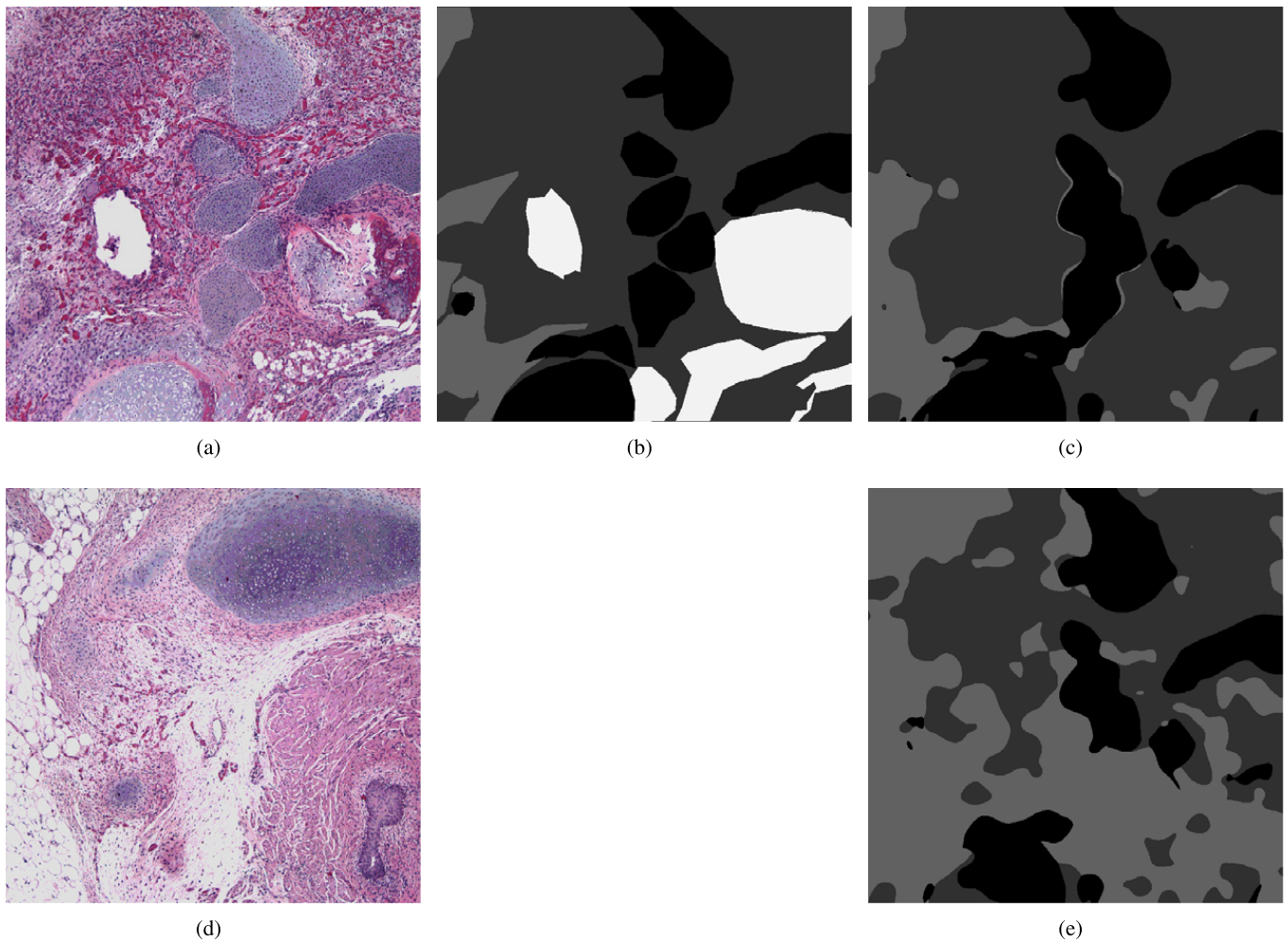


Fig. 8. An example of using PCA of local histograms to perform segmentation and classification of the image given in (a), which is a 3-bit quantized version of Fig. 1(a). A manually segmented and labeled version of (a) is shown in (b) where black represents cartilage, light gray represents connective 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 a PCA-based classification scheme (37), 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). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

The algorithm we present here exploits this concept. The first step is to train our classifier. To do so, let K be the number of distinct tissue types found in a training image such as Fig. 8(a) or (d). For any tissue type $k = 1, \dots, K$, we compute local histograms $\{h_{k;m}\}_{m=1}^{M_k}$ about pixel locations $\{x_{k;m}\}_{m=1}^{M_k}$ that have been labeled as being of that type by medical experts. Each $h_{k,m}$ is a nonnegatively-valued function over \mathcal{Y} that sums to one. There are several ways to pick the $x_{k,m}$'s. One approach is to have the expert choose each point individually. Alternatively, if the expert has manually segmented and labeled the entire image (Fig. 8(b)), then the $x_{k,m}$'s can be chosen at random from regions of type k . The number M_k of local histograms that we compute for type k is somewhat arbitrary; we used repeated experimentation to find a sample size large enough to guarantee reliably-decent performance.

In light of Theorem 1, it would be nice to demix the training local histograms $\{h_{k;m}\}_{m=1}^{M_k}$ in terms of a type-dependent class of more basic distributions $\{g_{k;n}\}_{n=1}^{N_k}$. That is, we would like to find nonnegatively-valued functions $g_{k;n}$ over \mathcal{Y} that sum to one and have the property that for each training local histogram $h_{k,m}$ there exists nonnegative scalars $\{\lambda_{k;m,n}\}_{n=1}^{N_k}$ that themselves sum to one and such that:

$$h_{k;m} \approx \sum_{n=1}^{N_k} \lambda_{k;m,n} g_{k;n}, \quad \forall m = 1, \dots, M_k. \quad (34)$$

Unfortunately, computing the $g_{k;n}$'s that minimize the approximation error in (34) is a nontrivial optimization problem. As such, we leave this approach for future work, and instead consider a mathematically-simpler problem in which the $\lambda_{k;m,n}$'s and $g_{k;n}$'s are permitted to be arbitrary real scalars and vectors, respectively. That is, we perform PCA for each tissue type k . To be precise, for each type, we form a $|\mathcal{Y}| \times M_k$ matrix H_k whose columns are the (vectorized) local histograms $h_{k;m}$ less their average \bar{h}_k :

$$H_k(:, m) = h_{k;m} - \bar{h}_k \quad \text{where } \bar{h}_k = \frac{1}{M_k} \sum_{m=1}^{M_k} h_{k;m}. \quad (35)$$

We then compute the singular value decompositions $H_k = U_k \Sigma_k V_k^T$ and identify those left-singular vectors $\{u_{k;n}\}_{n=1}^{N_k}$ that correspond to some experimentally-determined number N_k of dominant singular values $\{\sigma_{k;n}\}_{n=1}^{N_k}$. In this setting, the approximation (34) is replaced by:

$$h_{k;m} \approx \bar{h}_k + \sum_{n=1}^{N_k} \langle h_{k;m} - \bar{h}_k, u_{k;n} \rangle u_{k;n}, \quad \forall m = 1, \dots, M_k. \quad (36)$$

The classical theory of PCA states that the approximation error in (36) is optimally small in the sense that these specific $u_{k;n}$'s span the particular N_k -dimensional subspace of $\ell(\mathcal{Y}, \mathbb{R})$ whose orthogonal projection operator P_k minimizes the total squared-error $\sum_{m=1}^{M_k} \|h_{k;m} - \bar{h}_k - P_k(h_{k;m} - \bar{h}_k)\|^2$. The vectors \bar{h}_k and $\{u_{k;n}\}_{n=1}^{N_k}$ in hand, we store them in memory, completing the training phase of our classification algorithm.

To segment and label a given image f , we compute its local histograms (1), obtaining local distributions of color $h_x : \mathcal{Y} \rightarrow \mathbb{R}$, $h_x(y) = (\text{LH}_w f)(x, y)$ about every pixel location $x \in \mathcal{X}$. At any given x , we then assign a tissue label $k(x)$ by finding the tissue type k whose shifted subspace $\bar{h}_k + \text{span}\{u_{k;n}\}_{n=1}^{N_k}$ is nearest to h_x . Specifically, we let:

$$\begin{aligned} k(x) &= \underset{k=1, \dots, K}{\operatorname{argmin}} \left\| h_x - \bar{h}_k - \sum_{n=1}^{N_k} \langle h_x - \bar{h}_k, u_{k;n} \rangle u_{k;n} \right\|^2 \\ &= \underset{k=1, \dots, K}{\operatorname{argmin}} \left(\|h_x - \bar{h}_k\|^2 - \sum_{n=1}^{N_k} |\langle h_x - \bar{h}_k, u_{k;n} \rangle|^2 \right) \\ &= \underset{k=1, \dots, K}{\operatorname{argmin}} \left(\sum_{y \in \mathcal{Y}} |h_x(y) - \bar{h}_k(y)|^2 - \sum_{n=1}^{N_k} \left| \sum_{y \in \mathcal{Y}} (h_x(y) - \bar{h}_k(y)) u_{k;n}(y) \right|^2 \right). \end{aligned} \quad (37)$$

In implementation, we compute the summations over \mathcal{Y} in (37) as running sums, looping over all $y \in \mathcal{Y}$. This computational trick greatly reduces our memory requirements: at any given time, we only store a single level of $\text{LH}_w f$. By Theorem 2, such a level can be obtained by filtering an indicator function; in the following experimental results, we avoided edge artifacts by using a weighted noncyclic method of filtering, namely the \star -convolution of [17]. Without such a trick, one must store the entire local histogram transform in memory, a daunting task for even modestly-sized images: the full local histogram transform of the 1200×1200 , 8-bit RGB image given in Fig. 1(a) is a $1200 \times 1200 \times 256 \times 256 \times 256$ array.

Further computational advantages may be gained by quantizing the image and reducing the dimension of the color space. 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. That is, we quantize \mathcal{Y} from \mathbb{Z}_{256}^3 to \mathbb{Z}_8^2 . By Proposition 3, this is equivalent to binning the original $1200 \times 1200 \times 256 \times 256 \times 256$ local histogram array down to a new one of size $1200 \times 1200 \times 8 \times 8$. The quantized version of Fig. 1(a) is given in Fig. 8(a); for the sake of readability, a 3-bit quantized version of the unused green channel was included in this rendering. As a result of this quantization, it only takes a few seconds to assign per-pixel labels to a 1200×1200 histology image using a MATLAB-based implementation of (37), running on standard desktop hardware. To be precise, the classification algorithm requires approximately 75 seconds to assign a label to every pixel of the 1200×1200 3-bit RB image. Alternatively, using 4-bit and 5-bit RB images requires approximately 4 and 15 minutes of run time, respectively, while 3-bit and 4-bit RGB images need 4 and 27 minutes, respectively. For this particular set of images, further color quantization, such as using 2-bit colors ($\mathcal{Y} = \mathbb{Z}_2^2$) or converting the original image to grayscale ($\mathcal{Y} = \mathbb{Z}_{256}$), results in an unacceptable loss in classification accuracy, as do attempts at spatial quantization ($\mathcal{X} = \mathbb{Z}_{600}^2$). Throughout all of this experimentation, we chose the weighting function w to be a discretized Gaussian of a scale that we experimentally determined to yield good classification performance. In the numerical results presented below, we found $16\sqrt{2}$ to be a good standard deviation for this Gaussian, meaning the window had an effective radius of about 68 pixels. Multiscale methods were left for future work.

Two runs of this classification algorithm are depicted in Fig. 8. In the first run, we train the classifier on the 3-bit 1200×1200 red-blue image given in Fig. 8(a). For the sake of simplicity, we restrict ourselves to $K = 3$ tissue types: cartilage, connective tissue and pseudovascular tissue; all other tissue types are ignored in the confusion matrices given below. For each type $k = 1, 2, 3$, we randomly choose $M_k = 64$ points of that type, making use of a small number of the 1200^2 ground truth labels given in Fig. 8(b); edge artifacts are avoided by not picking points near the border. For each type, we then perform PCA on the 64 local histograms $h_{k;m}$ of that type, computing an average local histogram \bar{h}_k as well as the dominant left-singular vectors of H_k (35). For the sake of simplicity, in a given experiment we will use the same number of principal components for each of the three types, that is, $N_k = N$ for $k = 1, 2, 3$. At the same time, we experiment with this number itself, letting N be either 1, 2, 3 or 4. With the training complete, we then segment and classify Fig. 8(a) using

the decision rule (37), resulting in per-pixel labels such as the ones given in Fig. 8(c) for $N = 4$. Comparing Fig. 8(c) and the ground truth of Fig. 8(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 N are given by a confusion matrix:

	$N = 1$			$N = 2$			$N = 3$			$N = 4$		
	Ca	Co	Ps	Ca	Co	Ps	Ca	Co	Ps	Ca	Co	Ps
Ca	79	16	5	88	7	5	92	3	5	95	3	2
Co	0	93	7	1	91	8	4	93	3	3	93	4
Ps	2	8	90	2	8	90	3	8	89	3	7	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 tell us that when using a single principal component, those points labeled as cartilage by a medical expert in Fig. 8(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 in which a medical expert handpicks 64 points of each given type and lets the algorithm automatically assign labels to the rest.

The second run of this algorithm is almost identical to the first, with the exception that we use a distinct image in the training phase. To be precise, for each of the three tissue types, we perform PCA on the local histograms of 64 randomly-chosen points of that type in Fig. 8(d), making use of its ground truth labels (not pictured). We then apply the principal components obtained from Fig. 8(d) to generate labels (Fig. 8(e)) for Fig. 8(a) using the decision rule (37). 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	Co	Ps	Ca	Co	Ps	Ca	Co	Ps
Ca	70	17	13	87	6	7	86	6	8	76	16	8
Co	5	88	7	17	75	8	7	93	0	1	99	0
Ps	8	27	65	3	57	40	3	51	46	3	44	53

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.

Acknowledgments

The authors are extremely grateful to Dr. Carlos Castro and Dr. John A. Ozolek for introducing us to the motivating application and for providing us with raw image data and manually segmented ground truth labels used throughout this article. We also thank the two anonymous reviewers for their many helpful comments and suggestions. This work is supported by NSF DMS 1042701 and CCF 1017278, AFOSR F1ATA01103J001 and F1ATA00183G003NIH, NIH-R03-EB009875 and 5P01HD047675-02, the PA State Tobacco Settlement and the Kamlet–Smith Bioinformatics Grant. Parts of the work were presented at ISBI 2010 [2] and SBEC 2010 [14]. 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, Towards automatic identification and delineation of tissues and pathologies in H&E stained images, PhD Thesis, Carnegie Mellon University, 2011.

[2] 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.

[3] C. Bordenave, Y. Gousseau, F. Roueff, The dead leaves model: A general tessellation modeling occlusion, Adv. in Appl. Probab. 38 (2006) 31–46.

[4] 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.

[5] B. van Ginneken, B.M. ter Haar Romeny, Applications of locally orderless images, J. Vis. Commun. Image Represent. 11 (2000) 196–208.

[6] E. Hadjidemetriou, M.D. Grossberg, S.K. Nayar, Multiresolution histograms and their use for recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 831–847.

[7] M. Kass, J. Solomon, Smoothed local histogram filters, ACM Trans. Graph. 29 (2010) 100:1–100:10.

[8] V. Kober, T. Cichocki, M. Gedziowski, T. Szoplik, Optical-digital method of local histogram calculation by threshold decomposition, Appl. Opt. 32 (1993) 692–698.

[9] J.J. Koenderink, A.J. van Doorn, The structure of locally orderless images, Int. J. Comput. Vis. 31 (1999) 159–168.

[10] J.J. Koenderink, A.J. van Doorn, Blur and disorder, J. Vis. Commun. Image Represent. 11 (2000) 237–244.

- [11] A.B. Lee, D. Mumford, J. Huang, Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model, *Int. J. Comput. Vis.* 41 (2001) 35–59.
- [12] X. Liu, D. Wang, Image and texture segmentation using local spectral histograms, *IEEE Trans. Image Process.* 15 (2006) 3066–3077.
- [13] M.L. Massar, Local histograms for per-pixel classification, PhD Thesis, Air Force Institute of Technology, 2012.
- [14] M.L. Massar, R. Bhagavatula, M. Fickus, J. Kovačević, Local histograms for classifying H&E stained tissues, in: *Proc. 26th Southern Biomed. Eng. Conf.*, 2010, pp. 348–352.
- [15] D. Mumford, B. Gidas, Stochastic models for generic images, *Quart. Appl. Math.* 59 (2000) 85–111.
- [16] K. Ni, X. Bresson, T. Chan, S. Esedoglu, Local histogram based segmentation using the Wasserstein distance, *Int. J. Comput. Vis.* 84 (2009) 97–111.
- [17] G. Srinivasa, M. Fickus, Y. Guo, A. Linstedt, J. Kovačević, Active mask segmentation of fluorescence microscope images, *IEEE Trans. Image Process.* 18 (2009) 1817–1829.
- [18] T. Szoplik, M. Gedziorowski, Optical-digital processors for morphological and rank order filtering, *Opt. Laser Technol.* 28 (1996) 73–82.