3-22-2012

Local Histograms for Per-Pixel Classification

Melody R. Massar

LOCAL HISTOGRAMS FOR PER-PIXEL CLASSIFICATION

DISSERTATION

Melody L. Massar, Civilian

AFIT/DAM/ENC/12-02

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/DAM/ENC/12-02

LOCAL HISTOGRAMS FOR PER-PIXEL CLASSIFICATION

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Melody L. Massar, B.S., M.S.

Civilian

March 2012

AFIT/DAM/ENC/12-02

LOCAL HISTOGRAMS FOR PER-PIXEL CLASSIFICATION

Melody L. Massar, B.S., M.S.
Civilian

Approved:

//signed//	13 March 2012
Dr. Matthew Fickus Dissertation Advisor	Date
//signed//	13 March 2012
Dr. Mark E. Oxley Committee Member	Date
//signed//	13 March 2012
Dr. Richard K. Martin Committee Member	Date

Accepted:

//signed//	13 March 2012
M. U. Thomas Dean, Graduate School of Engineering and Management	Date

Table of Contents

AFIT/DAM/ENC/12-02

Abstract

We introduce a rigorous mathematical theory for the analysis of local histograms, and study how they interact with textures that can be modeled as occlusions of simpler components. We first show how local histograms can be computed as a system of convolutions and discuss some basic local histogram properties. We then introduce a probabilistic, occlusion-based model for textures and formally demonstrate that local histogram transforms are natural tools for analyzing the textures produced by our model. Next, we characterize all nonlinear transforms which satisfy the three key properties of local histograms and consider the appropriateness of local histogram features in the automated classification of textures commonly encountered in histological images. When classifying tissues, pathologists indicate they focus on simple, locally-defined features that essentially involve pixel counting, such as the number of cells in a region of given size, the size of the nuclei within these cells, and the distribution of color within both. We discuss how local histogram transforms can be used to produce numerical features that, when fed into mainstream classification schemes, mimic the baser aspects of a pathologist's thought process.

Acknowledgements

I would like to begin by thanking my friends and family for always being a constant source of encouragement, love, and support throughout this journey. I would also like to express my appreciation to Dr. John A. Ozolek and Dr. Carlos Castro for providing the histological images and motivating application for this dissertation. Thank you to Dr. Jelena Kovačević and Dr. Ramamurthy Bhagavatula for introducing me to this topic and being a great source of collaboration. I am also thankful for my committee members, Dr. Mark E. Oxley and Dr. Richard K. Martin, for their contributions to this dissertation. Last, but certainly not least, I want to thank my advisor, Dr. Matthew Fickus. The advice and guidance he has provided on this dissertation was indispensable. I am extremely grateful for the countless hours he has spent working with me. He has, and will continue to be, a source of inspiration to me as a mathematician.

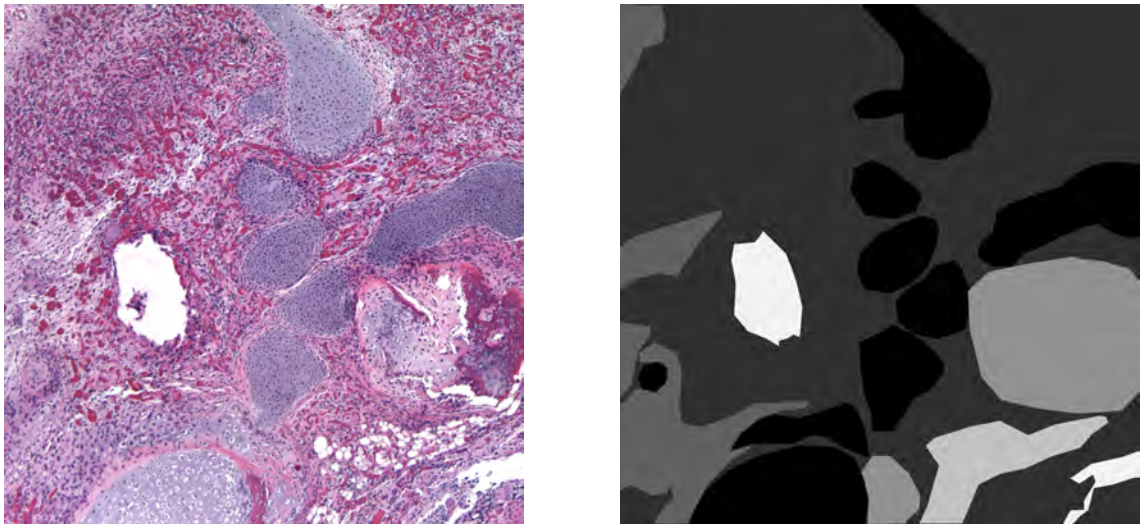
Melody L. Massar

I. Introduction

Local histograms are a well-studied signal processing tool [6, 7, 9–19, 22–24, 26–28, 32, 34–36]. They permit one to count the frequency of occurrence of a given feature over a small neighborhood of any given pixel. They have great potential in image classification, as they mimic certain visual cues that people use when hand-segmenting an image. In particular, rather than considering pixels in isolation, they account for the distribution of colors within the neighborhood of a given location [26, 27]. The focus of this research is the development of new mathematical theory, specifically that based upon local color histograms, to be used in the development of image classification systems, such as those considered in [2, 3].

Our focus is a rigorous mathematical analysis of local histograms. This theory arose during the development of an automatic classification scheme for histology (Figure 1(a)) [2, 3, 5, 26–28]. There, we considered the appropriateness of using local histograms as features in the automated classification of textures commonly encountered in images of hematoxylin and eosin (H&E) stained tissues. Even when a human user can easily delineate and identify the different tissues present in such multiple-tissue images (Figure 1(a)), segmenting them by hand (Figure 1(b)) can be time-consuming and error-prone. As such, there is a pressing need to automate as much of this process as possible.

One of the main contributions of this dissertation is the realization that local histograms—though difficult to analyze in general—can be well-understood in the special case where the images are generated according to *random* processes. Simple



(a) (b)

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. One of the purposes of this dissertation is to provide a theoretical justification for using local histograms to achieve this goal.

examples of the types of random images we consider are given in Figure 2(a), (b), and (c), denoted f_a, f_b and f_c, respectively. Each of these 128×128 images is generated by the following process: at each pixel location, a coin is flipped, resulting in either a white pixel value (heads) or a black pixel value (tails). The average intensity of each image depends on the fairness of the coin used to generate it: letting ρ denote the probability of getting heads, f_a, f_b and f_c were generated with ρ being $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$, respectively. By compositing f_a, f_b and f_c according to the values given in Figure 2(d) we obtain Figure 2(e). Though one can easily synthesize Figure 2(e) from Figure 2(a), (b), (c), and (d), the real-world goal is to do the opposite: given an image like Figure 2(e), we want to find the spatial relationships Figure 2(d) between distinct texture types Figure 2(a), (b), and (c).

The reason we care about randomly generated synthetic images like Figure 2(e) is because they seem to be a good model for the types of images that often arise

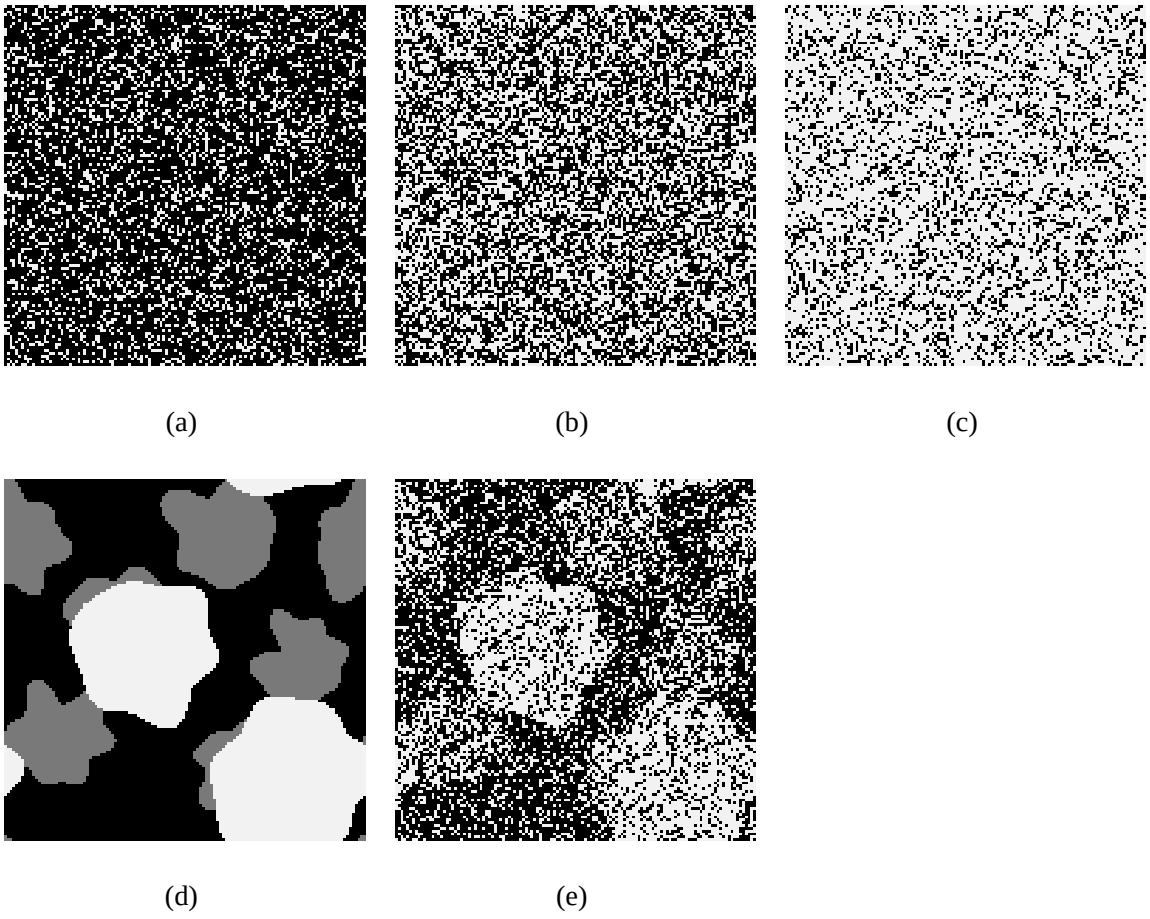


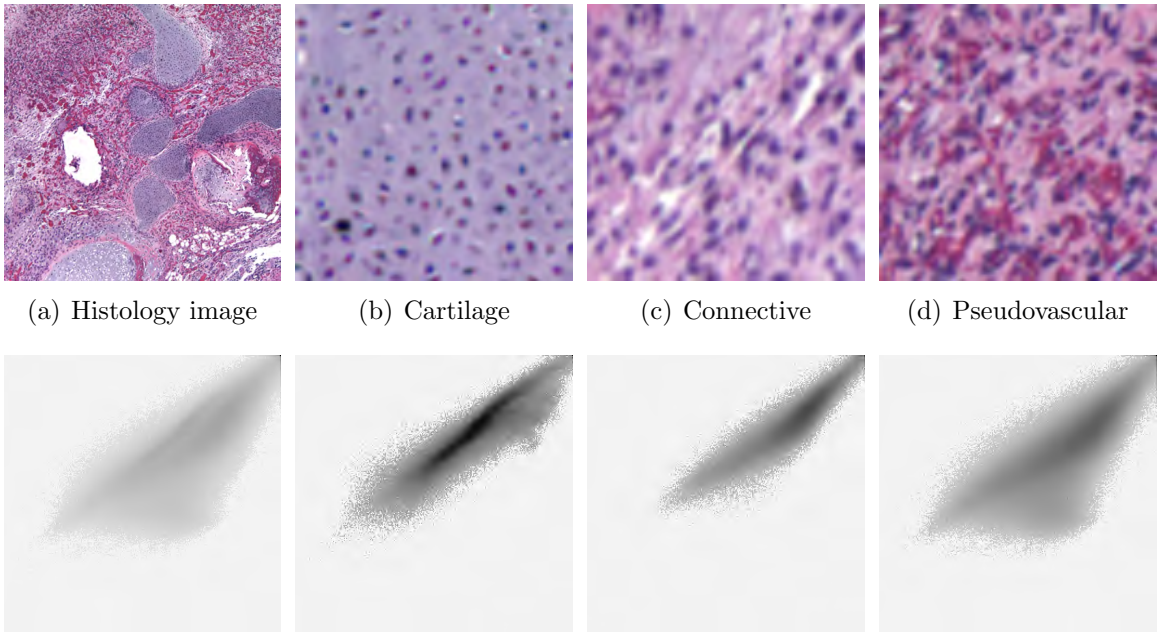
Figure 2. Three 128×128 black and white images (a), (b) and (c) that were randomly generated by a sequence of Bernoulli random variables—coin flips—with $\rho = \frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$, respectively. Compositing these three images together according to (d) produces an image (e) whose image statistics vary by location. Such images model those commonly found in digital microscopy. We want an algorithm to segment and classify such images. That is, given (e) and knowing the image statistics of each individual texture (a), (b) and (c), we want to produce a good approximation of (d). The theory presented in this dissertation provides theoretical justification for one such algorithm: assign a label to any given pixel by comparing its local histogram to the three distinct distributions of pixels found in (a), (b) and (c).

in digital microscopy of biological tissues [1, 3, 28]. Indeed, one way to think about such real-world images is that they are a mosaic of several distinct textures, each one being a relatively uniform composite of more basic texture types. For example, the histology image in Figure 1(a)—thumbnailed in Figure 3(a)—is comprised of more basic textures such as cartilage (Figure 3(b)), connective tissue (Figure 3(c)), and pseudovascular tissue (Figure 3(d)). Moreover, each tissue itself is comprised of more

basic components, such as the particular combination of dark red, dark purple, and light pink of typical pseudovascular tissue versus the dark and light purple of cartilage.

To summarize, our goal is to segment and classify real-world images such as Figure 1(a) according to texture (tissue) type; in order to have a mathematically rigorous analysis of this problem, we model these images as being the result of random processes (Figure 2(e)). In particular, we focus on a rigorous analysis of how local histograms can be used to solve this problem. We favor histograms since they are a natural way of estimating the distribution of a given texture's pixel values; as seen in Figure 3(f), (g), and (h), such histograms are a key discriminating feature between distinct texture types. It is important, however, that these histograms are computed *locally*—over a fixed neighborhood of every pixel location—since global histograms, such as the one depicted in Figure 3(e) derived from Figure 3(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 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.

All of our main results are based on the following idea: often the local histograms of a texture are, in fact, a combination of local histograms of more basic textures. For example, the local histograms of pseudovascular tissue (Figure 3(d)) are combinations of three simpler distributions—one pink, another purple and a third reddish-pink—while those of connective tissue (Figure 3(c)) are mixtures of only the first two. We formally prove that this intuition is valid, provided we work with *randomly* generated textures. This paves the way for a segmentation and classification scheme that assigns a texture label to any given pixel according to how its local histogram compares against those learned from training data.



(a) Histology image (b) Cartilage (c) Connective (d) Pseudovascular

(e) The red-blue histogram of (a). (f) The red-blue histogram of (b). (g) The red-blue histogram of (c). (h) The red-blue histogram of (d).

Figure 3. 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 Figure 1(b). In particular, the histogram (f) of cartilage (b) is computed over all points labeled in black in Figure 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.

1.1 Contributions

In this section, we highlight the main results of this dissertation and discuss the publications that have arisen from this research. Our research focuses on providing a rigorous mathematical analysis of local histograms. We first show that they can be computed as a system of convolutions (Theorem 1). In Theorem 3, we find a relationship between local histograms and images that are composites of several distinct textures: the local histograms of an image over a given neighborhood are approximately, on average, a combination of the local histograms of the tissues that comprise the image in that neighborhood. In Theorem 4, we show that, under certain hypotheses, this approximation becomes perfect, on average. Various methods for constructing probabilistic image models for which these hypotheses hold are given in Theorems 6, 7, and 8. These results and their proofs, along with a proof-of-concept segmentation-and-classification algorithm, have been submitted in the journal article "Local Histograms and Image Occlusion Models" [27].

Having found some nice relationships between local histograms and composite images, we then characterize all (possibly nonlinear) transforms which satisfy the three key properties of local histograms. In Theorem 10, we formally show that these properties are almost unique to local histograms. In particular, we show that local histograms are essentially the only transforms that distribute over random composites, on average. Having this understanding of the average behavior of such local histograms, we then turn to the subject of how closely this average approximates a typical local histogram. That is, we compute the variance of the local histograms of a composite image (Theorem 11), showing that the distance of our local histograms from the average is highly dependent on the scale of the local histogram window. To compute this variance, we necessarily make several simplifying assumptions on the nature of the random images in question; in Theorem 13, we provide a method for

constructing image models which demonstrates that these assumptions are indeed realistic. These results and their proofs are presented in the journal article "Local Histograms and Texture Classification" [29] which is in its final stages of preparation.

In addition to the two main journal articles mentioned above, the main results of this PhD research are also included in two conference proceedings. The conference proceeding "Local Histograms for Classifying H&E Stained Tissues" [26] introduces a rigorous mathematical theory for the analysis of local histograms and discusses its usefulness in the automated classification of textures commonly encountered in histological images. The conference proceeding "A Domain-Knowledge-Inspired Mathematical Framework for the Description and Classification of H&E Stained Histopathology Images" [28] presents an overview of our mathematical framework for the segmentation and classification of histology images.

In addition to the major contributions arising from this work in image processing, the journal article "Fast Computation of Spectral Centroids," [30] which represents a refinement of the MS research [25], has been published in *Advances in Computational Mathematics*. Finally, the book chapter "Finite Frames and Filter Banks" is to appear in *Finite Frames: Theory and Application*.

The following is a list of current publications that have arisen during this research:

Journal Articles

1. Massar, Melody L., Matthew Fickus, Erik Bryan, Douglas T. Petkie, and Andrew J. Terzuoli, Jr. "Fast Computation of Spectral Centroids," *Advances in Computational Mathematics*, 35: 83–97 (2011).

2. Massar, Melody L., Ramamurthy Bhagavatula, Matthew Fickus, and Jelena Kovačević. "Local Histograms and Image Occlusion Models," submitted to: *Applied and Computational Harmonic Analysis*, arXiv:1105.4069, 21 pages.

3. Massar, Melody L. and Matthew Fickus. "Local Histograms and Texture Classification," in preparation, to be submitted to: *Journal of Fourier Analysis and Applications*.

Refereed Book Chapters

5. Fickus, Matthew, Melody L. Massar, and Dustin G. Mixon. "Finite Frames and Filter Banks," to appear in: *Finite Frames: Theory and Application*, Peter G. Casazza and Gitta Kutyniok eds., Birkhäuser, 40 pages.

Conference Proceedings

6. Massar, Melody L., Ramamurthy Bhagavatula, Matthew Fickus, and Jelena Kovačević. "Local Histograms for Classifying H&E Stained Tissues," *Proceedings of the 26th Southern Biomedical Engineering Conference*, 348–352 (2010).

7. Massar, Melody L., Ramamurthy Bhagavatula, John A. Ozolek, Carlos A. Castro, Matthew Fickus, Jelena Kovačević. "A Domain-Knowledge-Inspired Mathematical Framework for the Description and Classification of H&E Stained Histopathology Images," *Proceedings of SPIE*, 8138: 81380U/1–7 (2011).

1.2 Outline

We begin in Chapter II with a short review of the background literature of local histograms, occlusions, and histology image classification. In Chapter III we discuss local histograms: their basic properties, how to efficiently compute them using convolutions, and how they interact with occlusions. In Chapter IV, we characterize all nonlinear transforms which satisfy the three key properties of local histograms. In Chapter V, we then extend our theory by computing the variance of the local histogram transform. In the final chapter, we present a preliminary segmentation-and-classification algorithm in which local histograms are decomposed using principal component analysis (PCA).

II. Literature Review

In this chapter, we discuss the literature relevant to our research. We begin our review in Section 2.1 with a discussion of the background literature of local histograms. Local histograms are used in image processing to compute the number of times a given feature occurs within a neighborhood of a given pixel [6, 7, 9–19, 22–24, 26–28, 32, 34–36]. In particular, local color histograms [12, 13, 18, 19, 26–28, 32, 35, 36] , local orientation histograms [6, 7, 9–11, 15], and local spectral histograms [22–24] allow us to compute the number of times a given pixel intensity, edge orientation, or spatial frequency, respectively, occur within a neighborhood of a given location. Background literature of each of these three types of histogram transforms are discussed in detail below.

In Section 2.2, we discuss the background literature of occlusion models [4, 20, 21, 31, 37, 38]. Such models are based on the ideas of representing an image as a collage of different objects. When a texture can be represented as an occlusion of more basic components, we expect it to interact well with local histograms, that is, we expect the local histograms of such textures to be a combination of the local histograms of the more basic components. The background literature of several occlusion models and their statistical properties are discussed below.

Section 2.3 provides a review on the literature of per-pixel texture classification, specifically classification of histology images like the one shown in Figure 1(a). We consider the appropriateness of applying local histogram features to the classification problem. Details of classification systems for histology images are given in [1–3, 5]; more properties and details of the mathematical framework of local histograms and occlusion models are given in [26–28].

2.1 Local Histograms

Local histograms are a well-studied signal processing tool. They have applications in image classification, image registration, computer graphics, computer vision, etc. They are used to compute the frequency of occurrence of a given feature within a neighborhood of a given pixel. Though they are often used to compute the distribution of color within the neighborhood of a given location, they can also be used to compute local distributions of edge orientations and spatial frequencies. Before we look at these local histograms and their applications, we first look at some methods used to compute them.

In [17, 34], a method for computing the local histograms of a grayscale image is presented. The computation involves convolving level sets of the input image with a binary kernel. From this, it can be shown that local histograms can be computed by convolving each level curve of a grayscale image with a binary kernel; we generalize this result in Theorem 1(a). This method of computing smoothed, locally-weighted histograms as two-dimensional spatial convolutions is also used in [16], where a means for computing derivatives and integrals of locally-weighted histograms over large neighborhoods is presented.

In [18, 19], histograms of continuous images are studied. Though this dissertation focuses on histograms of discrete images, such continuous theory can help one understand the discrete theory at a deeper level. Histograms are usually computed over discrete images as the number of pixels in a particular bin. In order to obtain a "density" independent of the bin-width, one can divide the histogram by the bin-width [18]. In continuous images, it is impossible to simply count the number of instances a given image obtains a given pixel value, as this will occur an infinite number of times, in general. Rather, in continuous images, one instead computes areas of image regions. The result is normalized by dividing by the total available area.

In [18], these facts are illustrated by a few one- and two-dimensional examples. This idea is also applied to local histograms of continuous images. In other words, rather than looking at the histogram of the whole image, regions of interest are considered by computing histograms over a windowed portion of an image.

Because the set of points where a continuous image obtains a particular pixel value will often have measure zero, the integral of this particular indicator function is zero. In [18, 19], rather than integrating over this indicator function, they first convolve the image with a Gaussian that has a small standard deviation and then integrate the image that has value one at a given intensity, but slowly decays to zero for values close to the given intensity. That is, they replace a level curve, which has measure zero, with a diffuse ribbon which has a nontrivial integral.

2.1.1 Local Color Histograms.

Local color histograms compute the distribution of color within the neighborhood of a given location. Such local histograms have been used in [26–28, 32] for segmentation and classification. In [32], the Wasserstein distance is used to compare the similarity of normalized histograms. This comparison is used in a nonparametric region-based active contour model for segmenting cluttered scenes. Mathematical properties of this proposed model, along with their proofs and experimental verifications, are also presented.

In [18], the idea of *locally orderless images* is presented. These images discard order at the fine level and replace it with order at the coarse level. In other words, the local image structure is replaced with local histograms. They describe how to construct such images, render them, and use them in several local and global image processing operations. Meanwhile, in [12], the versatility of locally orderless images is demonstrated by considering a variety of image processing tasks, such as variations

of adaptive histogram equalization, methods for noise and scratch removal, texture rendering, classification, and segmentation. In [19], a formal framework for locally orderless images is discussed. It is shown that in image segmentation, locally orderless images have distinct advantages over blurred images. In particular, when an image is blurred, the pixels lose their origin and the boundary is not clear. For example, if we blurred an image that had a uniform green region and a uniform red region, the boundary between the two regions would be yellow and not belong to the green or red regions. However, in locally orderless images, the pixels do not change their hue and a boundary area can be determined by the overlap in regions. In addition to the applications above, since locally orderless images can be easily rendered, they may also be used in image compression.

A global mode is the maximizer of a histogram, that is, the value that occurs most frequently. Similarly, a local mode is the maximizer of a local histogram. In [36], local mode filtering is used for noise reduction. This method of using the local mode is preferred over linear filtering and global modes as it preserves edges and details in the image. Another method that preserves edges is bilateral filtering [35]. Images are smoothed by means of a nonlinear combination of nearby image values. In particular, it replaces a pixel value with an average of similar and nearby pixel values. In [13], filters satisfying certain image measurement requirements of scale and imprecision are discussed. Among these filters are the mean filter, the median filter, and the mode filter. In particular, a system based on mode filtering and its behaviour over scale and imprecision is presented.

2.1.2 Local Orientation Histograms.

Histograms of oriented gradients have been used for object detection, matching image features, gesture recognition, and texture classification [6, 7, 9–11, 15]. The

local orientations in the image are obtained by calculating the first derivatives in two orthogonal directions. Because the Gaussian has a smoothing effect and the amount of noise reduction can be controlled by the scale parameter, the partial derivatives in the image are often calculated by filtering the image in the x and y directions with the filters that implement the derivatives of the Gaussian function. Edges and corners can also be found using image moment transforms; properties of such transforms for both continuous and digital images are discussed in [8].

Local object appearance and shape can often be characterized by the local histograms of gradient directions or edge orientations. In particular, for better invariance to illumination and shadow, the local histograms can be contrast-normalized giving us histograms of oriented gradient (HOG) descriptors. In [7], HOG descriptors are used as feature sets in a linear support vector machine (SVM) and outperform the existing feature sets used for object detection. More details on the best spatial sampling, orientation sampling, and contrast-normalization method are also discussed. In [6], a method for matching image features is proposed. Rather than using the common method of scale-invariant-feature-transform-descriptors (SIFT), they propose to use a new method that will reduce the negative effect of scale error. The method involves modifying the HOG descriptor's regular grid to an irregular grid. In [10], a method using orientation histograms to recognize hand gestures is presented.

Local orientation histograms have also been used in texture classification [9, 11, 15]. In [14], histograms of multiple resolutions of an image are computed to form a multiresolution histogram. Such histograms directly encode spatial information while still preserving several desirable properties of the histogram, such as being fast to compute, achieving significant data reduction, being invariant to rigid motions, and being robust to noise. In [9, 11, 15], this idea of computing multi-resolution histograms is applied to local image orientation. Because the dominant orientation depends

strongly on the observation scale, the macro-texture is evaluated by calculating the distribution of the dominant orientations for all pixels in the image that sample the texture at a micro-level.

2.1.3 Local Spectral Histograms.

Local spectral histograms are used for image and texture segmentation and texture classification [22–24]. The method for computing local spectral histograms is discussed in [22, 23]. The spectral histogram is defined with respect to a bank of filters. A filter selection algorithm that maximizes classification performance is proposed in [22]. Filters that have shown to be effective for different kinds of textures include the intensity filter, difference or gradient filters, Laplacian of Gaussian filters, and Gabor filters with both sine and cosine components [22].

To compute the local spectral histogram, a sub-band image is computed for each filter by taking a window in an input image and linearly convolving it with each filter. Then the histogram of each of the sub-band images is computed. The collection of these histograms, appropriately scaled, is called *the spectral histogram* of an image window. In other words, the spectral histogram of an image window is a vector consisting of marginal distributions of filter responses.

The distance between two spectral histograms is computed using a χ^2-statistic. Applying this distance measure to spectral histograms provides a means for texture classification. Several properties of spectral histograms are proved in [22]: it is shown that the spectral histogram is translation-invariant and a nonlinear operator. Additionally, with sufficient filters, a spectral histogram can uniquely represent any image up to a translation. Finally, all of the images sharing a spectral histogram define an equivalence class.

The process of applying local spectral histograms for image and texture segmen-

tation in [23, 24] includes three major steps. First, an initial segmentation result is obtained by classifying local windows. This is done by computing the local spectral histograms of a subsample of pixels and then using the distance measure to determine the classification of the pixels. After this first step, an algorithm is used to iteratively update the segmentation using the derived probability models. The final step involves reducing the boundary uncertainty. In particular, the large spatial window used for spectral histograms can cause a large segmentation error to occur on the boundary of the textures. The region boundaries are localized by building refined probability models that are sensitive to spatial patterns in segmented regions.

2.2 Occlusions

In this dissertation, we formally study how local histograms interact with textures that are formed as the occlusion of simpler components. Intuitively, the local histograms of the texture are related to the local histograms of the components that comprise the texture. Indeed, if we could break apart the histogram of the texture, we could gain information about what components are present and to what degree. We now give an overview of the literature on different occlusion image models and their applications.

Certain image textures can be thought of as occlusions of more basic components, that is, images are collages of statistically independent objects [20]. Several statistical models for occlusions have been presented [4, 20, 21, 31, 37, 38]. The goal of the models is to simulate images that well-approximate the statistics of natural images.

In [20], a stochastic model that takes occlusions into account and is both translation-invariant and fully scale-invariant is presented. Comparing the statistics of the simulated images of the occlusion model with statistics of natural images shows that the single pixel statistics and the derivative statistics agree well if the simulated images

are block averaged a few times for suitable subpixel resolution. Models of images that capture the invariance properties of real scenes could be used in image compression and in the study of sensory processing in biology.

In [21], an occlusion model based on the notion that the three-dimensional world creates a collage of occluded objects is presented. This model is called the *dead leaves* model and resembles images that were formed by randomly adding independent elementary shapes in layers. A comparison of the statistics of the occlusion model with the empirical statistics of two large databases of natural images gives excellent qualitative and good quantitative agreement. Such statistics have applications in computational and biological vision. The image model comes close to duplicating several elementary statistics of natural images and, at that time, was the only image model to do so. A result characterizing the distribution of the boundary between the shapes generated by the dead leaves model is given in [4].

The apparent scale-invariance exhibited by the statistics of natural images is discussed in [31]. In particular, there do not exist any scale-invariant probability measures supported on image functions. In order to construct such probability measures, their samples must be generalized functions. The basic idea of the paper is to assume that an image having clutter $c_1 + c_2$ can be constructed by adding independent images with clutter c_1 and c_2. Image models having this property are studied and a few axioms for such models are presented. These axioms are shown to be satisfied using the convergence of random wavelet expansions. The authors of [31] believe that the axioms present a natural model for images that is closer to the truth than Gaussian models, but nevertheless do not capture all the basic qualitative properties of such images.

In [37, 38], a statistical approach for partially occluded object recognition is presented. This approach is based on matching extracted local features to template

features. Two different statistical occlusion models are introduced and experiments illustrating the differences between the two occlusion models are presented. In [38], model-based statistical algorithms for recognition are used on examples from synthetic aperture radar imagery to evaluate the recognition performance and illustrate their advantages over other algorithms.

2.3 Per-Pixel Texture Classification

Per-pixel classification is important in many applications. Classification of aerial images has applications in automatic target recognition (ATR) and image registration. In particular, one of the long-term goals of this research is to apply local histogram features to a problem of passive navigation and surveillance, namely the development of an image-based passive navigation system which performs a real-time, per-pixel classification of the surrounding terrain, and registers the result to an internally stored map. Up to this point, our work in this area has focused on a similar biomedical image processing problem [1–3, 5], and has been performed in conjunction with Professor Jelena Kovacevic's group at Carnegie Mellon University. The focus of [26–28] is to present a rigorous mathematical theory for the analysis of local histograms and consider the appropriateness of their use in the classification of histology images, whereas [1–3, 5] focus more on the classification system and the method used to automatically identify and delineate histology images.

In [26–28], we discuss some of the many image features that pathologists indicate they use when classifying tissues by hand. We present a probabilistic, occlusion-based model for textures which shows how tissue-similar textures can be built from simpler ones; details of this work are given in Chapter III. After proving some properties of local histograms, we discuss how they relate to our model. We also discuss how local histogram transforms can be used as numerical features in the classification system

to mimic the pathologist's thought process.

In [1–3, 5], a framework and methodology for the automatic identification and delineation of histology images is presented. Although pathologists can accurately and consistently identify and delineate tissues and their pathologies, it is an expensive and time-consuming task, therefore presenting the need for an automated algorithm. In [3], an example of such an algorithm is validated on a clinical and research application and experimental results show great promise towards developing an automated tool for digital histopathology. We now present our work on the analysis of local histograms.

III. Local Histograms and Image Occlusion Models

In this chapter, we discuss some of the basic properties of local histograms. In particular, we show that local histograms can be computed using convolutions. We also introduce a model for textures that relies on the concept of image occlusions, and present one of our main results, that is, on average, the local histograms of the composite of a sequence of images can be approximated by a convex combination—a linear combination with nonnegative coefficients that sum to one—of the local histograms of each image. We discuss the conditions under which the approximation becomes perfect, namely, when the expected value of the graph's characteristic function is independent of pixel location, a property we call *flatness*. We also consider what properties of occlusion models give us flatness and what operations can be used to build intricate flat occlusion models from more simple examples.

3.1 Local Histograms

In this section, we formally introduce local histograms, discuss an efficient means of computing them, and discuss several of their basic properties. 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, an image f is a member of the set $\mathcal{Y}^{\mathcal{X}} := \{f : \mathcal{X} \to \mathcal{Y}\}$. For example, the 1200×1200, 8-bit red-green-blue (RGB) image given in Figure 1(a) has $\mathcal{X} = \mathbb{Z}_{1200}^2 = \mathbb{Z}_{1200} \oplus \mathbb{Z}_{1200}$ and $\mathcal{Y} = \mathbb{Z}_{256}^3$, where $\mathbb{Z}_N$ denotes the cyclic group of integers modulo N. Here, we note that $\oplus$ denotes the direct sum of abelian groups. That is, it is the Cartesian product equipped with pointwise addition; formally, $\mathcal{X} \oplus \mathcal{Y} := \{(x, y) : x \in \mathcal{X}, y \in \mathcal{Y}\}$, where $(x, y) + (x', y') := (x + x', y + y')$. 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* $w \in \mathbb{R}^{\mathcal{X}}$. In practice, this function is usually chosen to be a discretized Gaussian or some nonnegative-valued function whose values sum to one. However, for the purpose of generality, we assume w to be arbitrary unless stated otherwise. Specifically, the *local histogram transform* of f with respect to w is the function $\mathrm{LH}_w f : \mathcal{X} \oplus \mathcal{Y} \to \mathbb{R}$,

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

where $\delta_y(y') := 1$ if $y' = y$ and is otherwise zero. For any fixed $x \in \mathcal{X}$, the corresponding cross-section of this function, namely $(\mathrm{LH}_w f)(x, \cdot) : \mathcal{Y} \to \mathbb{R}$, counts the number of instances at which f obtains a given value y in a w-neighborhood of x.

Computing local histograms can be time consuming, especially as $\mathcal{X}$ and $\mathcal{Y}$ become large. Noting that $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$, we have that for a general weighting function w, a direct computation of (1) requires $\mathcal{O}(|\mathcal{X}|^2|\mathcal{Y}|)$ operations: summing the real values $w(x')$ for all $x' \in \mathcal{X}$ such that $f(x + x') = y$. A more efficient method is given in Theorem 1 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 $1_f : \mathcal{X} \oplus \mathcal{Y} \to \mathbb{R}$,

$$1_f(x, y) := 1_{f^{-1}\{y\}}(x) = \delta_y(f(x)) = \begin{cases} 1, & f(x) = y, \\ 0, & f(x) \neq y, \end{cases} \tag{2}$$

with the *reversal* of $w \in \mathbb{R}^{\mathcal{X}}$, namely $\tilde{w}(x) := w(-x)$. Here, we let $f^{-1}\{y\}$ denote the preimage of $\{y\}$ under f, that is, the set of all $x \in \mathcal{X}$ such that $f(x) = y$. We note that the characteristic function does not distribute over addition and is thus nonlinear. Therefore, we have that local histograms are also nonlinear; a fact which can easily be

verified. This method for computing local histograms using convolutions is illustrated in Figure 4. Alternatively, (1) can be computed as a single convolution over $\mathcal{X} \oplus \mathcal{Y}$; here, the *tensor product* of $w \in \mathbb{R}^{\mathcal{X}}$ with $\omega \in \mathbb{R}^{\mathcal{Y}}$ is defined as $w \otimes \omega \in \mathbb{R}^{\mathcal{X} \oplus \mathcal{Y}}$, $(w \otimes \omega)(x, y) := w(x)\omega(y)$.

Theorem 1. *For any $w \in \mathbb{R}^{\mathcal{X}}$, $\omega \in \mathbb{R}^{\mathcal{Y}}$, $f \in \mathcal{Y}^{\mathcal{X}}$, $x \in \mathcal{X}$, and $y \in \mathcal{Y}$:*

(a) Local histograms (1) *can be evaluated as a system of $|\mathcal{Y}|$ convolutions over $\mathcal{X}$:*

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

(b) Alternatively, (1) *may be computed as a single convolution over $\mathcal{X} \oplus \mathcal{Y}$:*

$$(\delta_0 \otimes \omega) * \mathrm{LH}_w f = 1_f * (\tilde{w} \otimes \omega).$$

In particular, taking $\omega = \delta_0$ gives $\mathrm{LH}_w f = 1_f * (\tilde{w} \otimes \delta_0)$.

Proof. For (a), replacing x' with $-x'$, and substituting $\delta_y(f(x-x')) = 1_{f^{-1}\{y\}}(x-x')$, a relation which follows from (2), into (1) yields:

$$\begin{aligned}
(\mathrm{LH}_w f)(x, y) &= \sum_{x' \in \mathcal{X}} w(x')\delta_y(f(x + x')) \\
&= \sum_{x' \in \mathcal{X}} w(-x')1_{f^{-1}\{y\}}(x - x') \\
&= \sum_{x' \in \mathcal{X}} \tilde{w}(x')1_{f^{-1}\{y\}}(x - x') \\
&= (\tilde{w} * 1_{f^{-1}\{y\}})(x). \qquad (3)
\end{aligned}$$

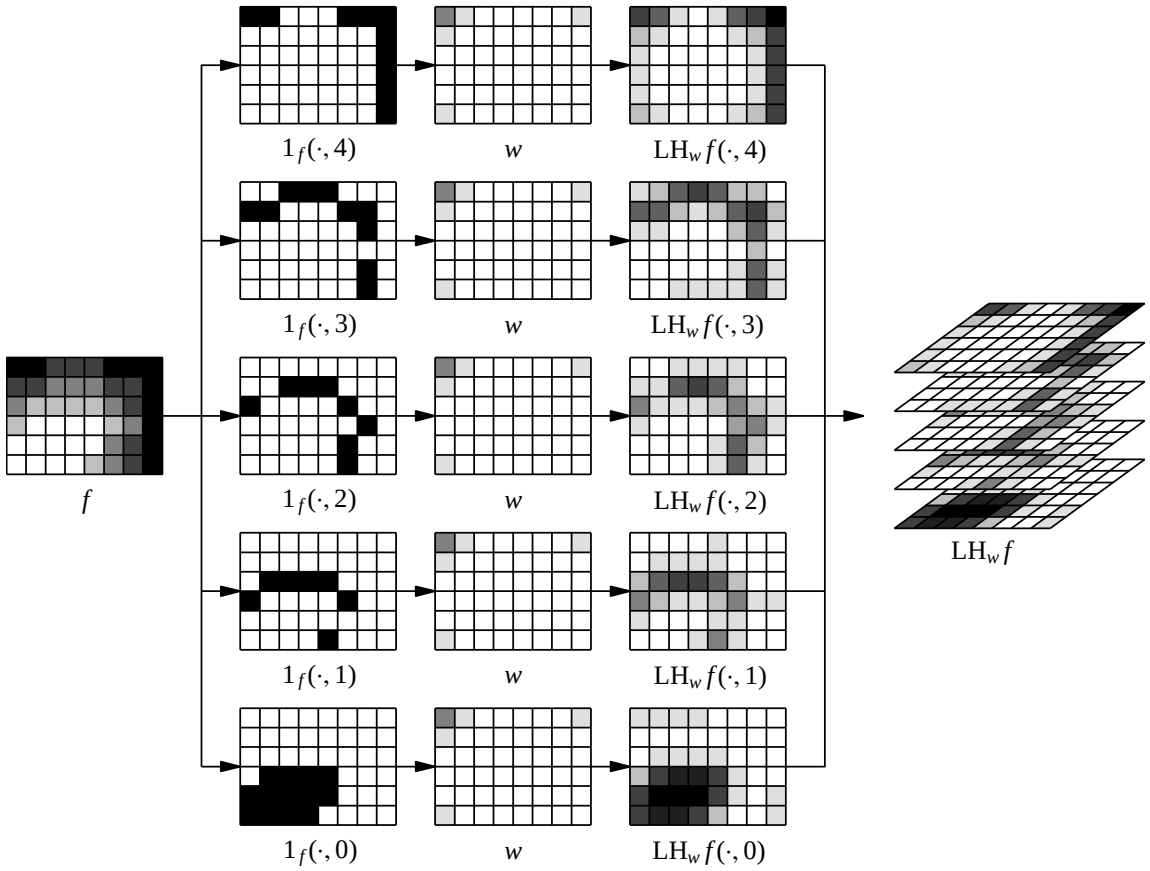


Figure 4. An example of how to compute local histograms using Theorem 1(a). For the sake of readability, larger numerical values are represented by darker shades throughout. The source image (far left) f is 6×8 and has grayscale values ranging from 0 to 4. That, is $f \in \mathcal{Y}^{\mathcal{X}}$ where $\mathcal{X} = \mathbb{Z}_6 \oplus \mathbb{Z}_8$ and $\mathcal{Y} = \mathbb{Z}_5$. Its characteristic function (2) 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 1(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.

For (b), the definition of δ_0 gives:

$$\begin{aligned}[(\delta_0 \otimes \omega) * \mathrm{LH}_w f](x,y) &= \sum_{(x',y') \in \mathcal{X} \oplus \mathcal{Y}} (\delta_0 \otimes \omega)(x',y')(\mathrm{LH}_w f)(x-x', y-y') \\ &= \sum_{y' \in \mathcal{Y}} \omega(y')(\mathrm{LH}_w f)(x, y-y'). \end{aligned} \tag{4}$$

Substituting (3) into (4) and using (2), gives our result:

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

For the final conclusion, we note that when $\omega = \delta_0$, it follows from this result that $\mathrm{LH}_w f = (\tilde{w} \otimes \delta_0) * 1_f$. In other words, the local histogram can be computed by convolving the graph's characteristic function (2) with a filter $\tilde{w} \otimes \delta_0$ which is completely supported on one cross-section of $\mathcal{X} \oplus \mathcal{Y}$ where $y = 0$. □

In our next result, we make use of the translation operator $\mathrm{T}^x : \mathbb{Z}_N^{\mathcal{X}} \to \mathbb{Z}_N^{\mathcal{X}}$, $\mathrm{T}^x \varphi(x') := \varphi(x' - x)$. Here, we summarize several other basic properties of local histograms:

Proposition 2. *For any $w \in \mathbb{R}^{\mathcal{X}}$ and $f \in \mathcal{Y}^{\mathcal{X}}$:*

(a) If the values of w sum to 1, then the levels of a local histogram transform also sum to 1:

$$\sum_{y \in \mathcal{Y}} (\mathrm{LH}_w f)(x,y) = 1, \, \forall x \in \mathcal{X}.$$

(b) Local histograms commute with spatial translation T^x*:*

$$\mathrm{LH}_w \mathrm{T}^x = \mathrm{T}^{(x,0)} \mathrm{LH}_w, \, \forall x \in \mathcal{X}.$$

(c) Adding constants to images shifts their local histograms along $\mathcal{Y}$*:*

$$\mathrm{LH}_w(f + y) = \mathrm{T}^{(0,y)} \mathrm{LH}_w f, \, \forall y \in \mathcal{Y}.$$

(d) Quantizing an image will bin its local histograms:

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

Proof. For any $w \in \mathbb{R}^{\mathcal{X}}$ and $f \in \mathcal{Y}^{\mathcal{X}}$:

(a) Let $x \in \mathcal{X}$ and assume that $\sum_{x' \in \mathcal{X}} w(x') = 1$. By interchanging sums and noting that for some fixed $x' \in \mathcal{X}$, $f(x + x')$ is equal to one and only one $y \in \mathcal{Y}$, it follows that:

$$\sum_{y \in \mathcal{Y}} (\mathrm{LH}_w f)(x, y) = \sum_{y \in \mathcal{Y}} \sum_{x' \in \mathcal{X}} w(x') \delta_y(f(x + x')) = \sum_{x' \in \mathcal{X}} w(x') = 1.$$

(b) Let $(x', y) \in \mathcal{X} \oplus \mathcal{Y}$. The result follows directly from the definitions of local histogram and translation:

$$\begin{aligned} (\mathrm{LH}_w \mathrm{T}^x f)(x', y) &= \sum_{x'' \in \mathcal{X}} w(x'') \delta_y(\mathrm{T}^x f(x' + x'')) \\ &= (\mathrm{LH}_w f)(x' - x, y) \\ &= (\mathrm{T}^{(x,0)} \mathrm{LH}_w f)(x', y). \end{aligned}$$

(c) Let $(x, y') \in \mathcal{X} \oplus \mathcal{Y}$. First note that

$$\begin{aligned}
\delta_{y'}((f+y)(x+x')) &= \begin{cases} 1, & (f+y)(x+x') = y', \\ 0, & (f+y)(x+x') \neq y', \end{cases} \\
&= \begin{cases} 1, & f(x+x') = y' - y, \\ 0, & f(x+x') \neq y' - y, \end{cases} \\
&= \delta_{y'-y}(f(x+x')). \qquad (5)
\end{aligned}$$

Combining the definition of the local histogram with (5) gives our result:

$$\begin{aligned}
[\mathrm{LH}_w(f+y)](x, y') &= \sum_{x' \in \mathcal{X}} w(x') \delta_{y'-y}(f(x+x')) \\
&= (\mathrm{LH}_w f)(x, y' - y) \\
&= (\mathrm{T}^{(0,y)} \mathrm{LH}_w f)(x, y').
\end{aligned}$$

(d) We claim that:

$$\delta_{y'}((q \circ f)(x+x')) = \sum_{\substack{y \in \mathcal{Y} \\ q(y) = y'}} \delta_y(f(x+x')). \qquad (6)$$

For fixed x, x', y', f, and q, the left-hand-side of (6) equals one if and only if $(q \circ f)(x + x') = y'$, that is, if and only if there exists some $y \in \mathcal{Y}$ such that $f(x + x') = y$ and $q(y) = y'$, implying the right-hand-side of (6) is also one. Alternatively, if the left-hand-side of (6) is zero, then $f(x + x')$ does not equal any y such that $q(y) = y'$ and, therefore, the right-hand-side of (6) is also zero.

This claim, along with (1), gives our result:

$$[\mathrm{LH}_w(q \circ f)](x, y') = \sum_{\substack{y \in \mathcal{Y} \\ q(y)=y'}} \sum_{x' \in \mathcal{X}} w(x')\delta_y(f(x+x')) = \sum_{\substack{y \in \mathcal{Y} \\ q(y)=y'}} (\mathrm{LH}_w f)(x, y). \quad \square$$

In the following section, we define a model for the occlusion of a sequence of images and present one of our main results: on average, the local histogram of the occlusion of a sequence of images can be approximated as a convex combination of the local histograms of each image.

3.2 A Probabilistic Image Occlusion Model

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 Figure 3(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. We rigorously show that such decompositions of local histograms indeed exist for textures arising from a certain class of probabilistic image models.

To see how to formalize these ideas, we consider the example given in the introduction (Figure 2) where at each pixel location, a coin is flipped, resulting in either a white pixel value (heads) or a black pixel value (tails). One expects that, on average, the local histogram at any point will consist of two peaks: one in the white portion of $\mathcal{Y}$, and one in the black. Such an image can be regarded as the result of *occluding* a solid black image f_0 with a solid white 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 $\mathcal{Y}^{\mathcal{X}}$ with respect to a given *label function*

$\varphi \in \mathbb{Z}_N^{\mathcal{X}}$ is:

$$\big(\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}\big)(x) := f_{\varphi(x)}(x). \tag{7}$$

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 $\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}$ at that point.

The main results of this dissertation are concerned with when the local histograms (1) of a composite image (7) are related to the local histograms of the individual images f_n. 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 white and black peaks of the composite image's local histograms to be ρ and $1-\rho$, respectively. That is, $\mathrm{LH}_w\mathrm{occ}_\varphi\{f_0, f_1\}$ should be $(1-\rho)\mathrm{LH}_w f_0 + \rho\mathrm{LH}_w f_1$, 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 of all possible composite images $\{\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}\}_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}}$ as defined in (7) and obtained by letting φ be any one of the $N^{|\mathcal{X}|}$ elements of $\mathbb{Z}_N^{\mathcal{X}}$. 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 the label function φ. That is, we work in the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where the sample space Ω is the set $\mathbb{Z}_N^{\mathcal{X}}$ of all functions from $\mathcal{X}$ to $\mathbb{Z}_N$, the σ-algebra $\mathcal{F}$ is the power set of $\mathbb{Z}_N^{\mathcal{X}}$, and the probability measure $\mathcal{P}$ is defined as:

$$\mathcal{P}(\{\varphi_i\}_{i\in\mathcal{I}}) := \sum_{i\in\mathcal{I}} \mathrm{P}_\Phi(\varphi_i),$$

where P_Φ is the corresponding probability density function on $\mathbb{Z}_N^{\mathcal{X}}$, that is, it is a

nonnegatively-valued function such that $\sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) = 1$. Going back to our coin-flipping example, the role of the occlusion model is to determine the *likelihood* of choosing a particular $\varphi : \mathcal{X} \to \mathbb{Z}_2$. In particular, in a *Bernoulli model*, we fix $\rho_1 \in [0,1]$, let $\rho_0 := (1 - \rho_1)$ and define the probability of picking any particular $\varphi \in \mathbb{Z}_2^{\mathcal{X}}$ as $\mathrm{P}_\Phi(\varphi) = \prod_{x \in \mathcal{X}} \rho_{\varphi(x)}$. We emphasize here that though this method of producing images is random, it is by no means uniform: the role of the model is to emphasize certain images over others. In particular, when $\rho_1 = \frac{1}{4}$ the Bernoulli model will often produce images similar to Figure 2(a), but will almost never produce images like Figure 2(b) and (c). Indeed, Figure 2(b) and (c) arise from distinct Bernoulli models in which $\rho_1 = \frac{1}{2}$ and $\rho_1 = \frac{3}{4}$, respectively. Images such as Figure 2(e) are therefore regarded as composites of images arising from several distinct models. From this perspective, our segmentation and classification algorithm boils down to using local histograms to estimate which of these three models is being exhibited at any pixel.

For a more realistic example, imagine three 128×128 images f_0, f_1 and f_2 which exhibit nearly constant shades of pink, purple and red, respectively. Given any label function $\varphi : \mathbb{Z}_{128}^2 \to \mathbb{Z}_3$ we can produce a corresponding 128×128 composite image $\mathrm{occ}_\varphi\{f_0, f_1, f_2\}$ 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 Figure 3(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.

To formally confirm our intuition that local histograms, on average, should distribute over random occlusions, fix any set of N source images $\{f_n\}_{n=0}^{N-1}$ and let Φ

be any occlusion model as defined in (7). That is, let Φ be a random variable version of a label function $\varphi : \mathcal{X} \to \mathbb{Z}_N$, as defined by a probability density function $\mathrm{P}_\Phi : \mathbb{Z}_N^{\mathcal{X}} \to [0,1]$ where $\sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) = 1$. 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 (2):

$$\overline{1}_\Phi(x,n) := \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x,n) = \sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x)=n}} \mathrm{P}_\Phi(\varphi). \tag{8}$$

Essentially, $\overline{1}_\Phi(x,n)$ is the probability that a random label function φ generated by the occlusion model Φ will assign label n to pixel location x. That is, for any fixed x and n, $\overline{1}_\Phi(x,n)$ is the expected value of $1_\varphi(x,n)$. More generally, for any real-valued function β over $\mathbb{Z}_N^{\mathcal{X}}$, the *expected value* of β is:

$$\mathrm{E}_\Phi(\beta) := \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)\beta(\varphi) \tag{9}$$

Having these concepts, we present one of our main results:

Theorem 3. *For any sequence of images $\{f_n\}_{n=0}^{N-1}$ in $\mathcal{Y}^{\mathcal{X}}$, weighting function w and any N-image occlusion model Φ, the expected value* (9) *of the local histogram* (1) *at any given $(x,y) \in \mathcal{X} \oplus \mathcal{Y}$ of the composite image* (7) *with respect to w is:*

$$\mathrm{E}_\Phi(\mathrm{LH}_w \mathrm{occ}_\Phi \{f_n\}_{n=0}^{N-1})(x,y) = \sum_{n=0}^{N-1} \overline{1}_\Phi(x,n)(\mathrm{LH}_w f_n)(x,y) + \varepsilon, \tag{10}$$

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

$$\sum_{n=0}^{N-1} \overline{1}_\Phi(x,n) = 1, \tag{11}$$

and so (10) *states that, the local histograms of the composite image* $\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}$ *can, on average, 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 (7) is:

$$\mathrm{E}_\Phi(\mathrm{LH}_w\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})(x,y) = \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) \sum_{x'\in\mathcal{X}} w(x')\delta_y((\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})(x+x')). \quad (12)$$

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 $n \neq \varphi(x+x')$:

$$\begin{aligned}\sum_{n=0}^{N-1} 1_\varphi(x+x',n)\delta_y(f_n(x+x')) &= (1)\delta_y(f_{\varphi(x+x')}(x+x')) + \sum_{n\neq\varphi(x+x')} (0)\delta_y(f_n(x+x')) \\ &= \delta_y((\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})(x+x')), \quad (13)\end{aligned}$$

where the final equality follows immediately from (7). Substituting (13) into (12) and using (8) yields:

$$\begin{aligned}\mathrm{E}_\Phi(\mathrm{LH}_w\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})(x,y) &= \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) \sum_{x'\in\mathcal{X}} w(x') \sum_{n=0}^{N-1} 1_\varphi(x+x',n)\delta_y(f_n(x+x')) \\ &= \sum_{n=0}^{N-1} \sum_{x'\in\mathcal{X}} w(x')\delta_y(f_n(x+x')) \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x+x',n) \\ &= \sum_{n=0}^{N-1} \sum_{x'\in\mathcal{X}} w(x')\delta_y(f_n(x+x'))\overline{1}_\Phi(x+x',n). \quad (14)\end{aligned}$$

Rewriting (14) in terms of $\varepsilon := \sum_{n=0}^{N-1} \sum_{x'\in\mathcal{X}} w(x')\delta_y(f_n(x+x'))[\overline{1}_\Phi(x+x',n) - \overline{1}_\Phi(x,n)]$

gives our first claim (10):

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

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

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

Finally, to prove our third claim (11), note that for any fixed $x \in \mathcal{X}$, (8) gives:

$$\sum_{n=0}^{N-1} \overline{1}_\Phi(x,n) = \sum_{n=0}^{N-1}\sum_{\varphi\in\mathbb{Z}_N^\mathcal{X}} \mathrm{P}_\Phi(\varphi)1_\varphi(x,n) = \sum_{\varphi\in\mathbb{Z}_N^\mathcal{X}} \mathrm{P}_\Phi(\varphi) \sum_{n=0}^{N-1} \begin{cases} 1, & \varphi(x) = n, \\ 0, & \varphi(x) \neq n. \end{cases} \tag{15}$$

Since, as previously noted, we have $\varphi(x) = n$ for exactly one n, (15) becomes:

$$\sum_{n=0}^{N-1} \overline{1}_\Phi(x,n) = \sum_{\varphi\in\mathbb{Z}_N^\mathcal{X}} \mathrm{P}_\Phi(\varphi) = 1. \qquad \square$$

An example illustrating the direct computation of the left-hand side of (10) is given in Figure 5. Note that Theorem 3 implies that the error term ε in (10) will be small provided the probability $\overline{1}_\Phi(x,n)$ of assigning label n to x changes little as x varies over regions smaller than the the support of w. In the next section, we discuss

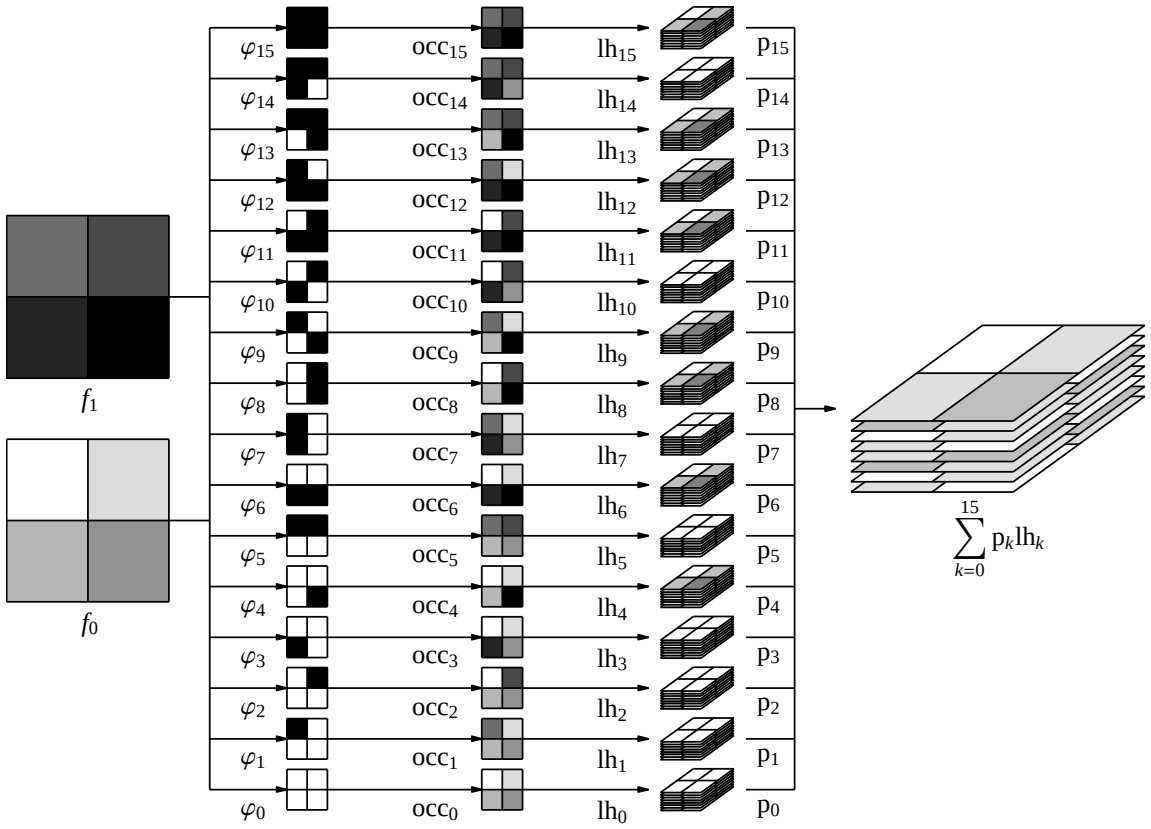


Figure 5. An example of how to compute the left-hand side of (10) explicitly as a probability-weighted sum. For the sake of readability, larger numerical values are represented by darker shades throughout. We consider two 2×2, 3-bit images, namely $\{f_n\}_{n=0}^{N-1}$ in $\mathcal{Y}^{\mathcal{X}}$ where $N = 2$, $\mathcal{X} = \mathbb{Z}_2 \oplus \mathbb{Z}_2$ 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 $\varphi : \mathbb{Z}_2 \oplus \mathbb{Z}_2 \to \mathbb{Z}_2$ (left column) each yielding a composite image $\mathrm{occ}_\varphi\{f_0, f_1\}$ (center column); in accordance with (7), 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) which was computed using the weighting function $w = \frac{1}{2}\delta_{0,0} + \frac{1}{4}(\delta_{1,0} + \delta_{0,1})$. Since occlusion (7) 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 $\mathbb{Z}_2^{\mathbb{Z}_2 \oplus \mathbb{Z}_2}$ of all possible φ's; in this example, we let each φ_k have equal probability, that is, we let $\mathrm{P}_\Phi(\varphi_k) := p_k = \frac{1}{16}$. In particular, if the occlusion model Φ is flat (16), meaning in this case that the probability-weighted-sum of all φ's is a constant function of x, then Theorem 4 states that this average is a convex combination of the local histograms of f_0 and f_1 as depicted in Figure 6.

what hypotheses on Φ will make the ε in Theorem 3 vanish completely.

3.3 Constructing Flat Occlusion Models

In this section, we provide a sufficient hypothesis on the occlusion model Φ so as to ensure that the local histograms (1) of a composite image (7) 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 for each $n \in \mathbb{Z}_N$:

$$\sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x)=n}} \mathrm{P}_\Phi(\varphi) = \lambda_n, \quad \forall x \in \mathcal{X}. \tag{16}$$

Equivalently, Φ is flat if $\overline{1}_\Phi(x, n)$, defined in (8), is constant with respect to x. 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 (16) over all n yields that $\sum_{n=1}^{N} \lambda_n = 1$. 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 example where the values of φ are determined by $|\mathcal{X}|$ spatially-independent coin flips, the probability of getting any particular $\varphi \in \mathbb{Z}_2^{\mathcal{X}}$ is $\mathrm{P}_\Phi(\varphi) = \rho^{|\varphi^{-1}\{1\}|}(1-\rho)^{|\mathcal{X}|-|\varphi^{-1}\{1\}|}$; substituting this expression into (16), we find that this model is indeed flat with $\lambda_0 = 1-\rho$ and $\lambda_1 = \rho$, as detailed below. Note that, if $\rho > \frac{1}{2}$, the resulting random image $\mathrm{occ}_\Phi\{f_0, f_1\}$ will be more white than black; flatness does not mean that each label is equally likely, but rather that the chance of being white 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 4. *If* Φ *is flat as in* (16)*, then the expected value* (9) *of the local histogram transform* (1) *of a composite image* (7) *is a convex combination of the local histograms of each individual image:*

$$\sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)(\mathrm{LH}_w \mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})(x,y) = \sum_{n=0}^{N-1} \lambda_n(\mathrm{LH}_w f_n)(x,y). \tag{17}$$

Proof. If Φ is flat, $\overline{1}_\Phi(x+x',n) = \overline{1}_\Phi(x,n)$ for all $x,x' \in \mathcal{X}$. The error bound in Theorem 3 then gives $\varepsilon = 0$. Denoting $\overline{1}_\Phi(x,n)$ as λ_n in (10) thus yields our claim. □

Therefore, when Φ is flat, (10) simplifies to (17), and so the in-depth computation of Figure 5 can be replaced by the much simpler one depicted in Figure 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.

We now 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.

Theorem 5. *If* Φ *is a Bernoulli random variable, then* Φ *is flat.*

Proof. Let $K = |\mathcal{X}|$ and claim that for any $F : \mathbb{Z}_N^{\mathcal{X}} \to \mathbb{R}$, we have:

$$\sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} F(\varphi) = \sum_{n_1=0}^{N-1} \cdots \sum_{n_K=0}^{N-1} \sum_{\substack{\varphi\in\mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x_k)=n_k,\,\forall k}} F(\varphi). \tag{18}$$

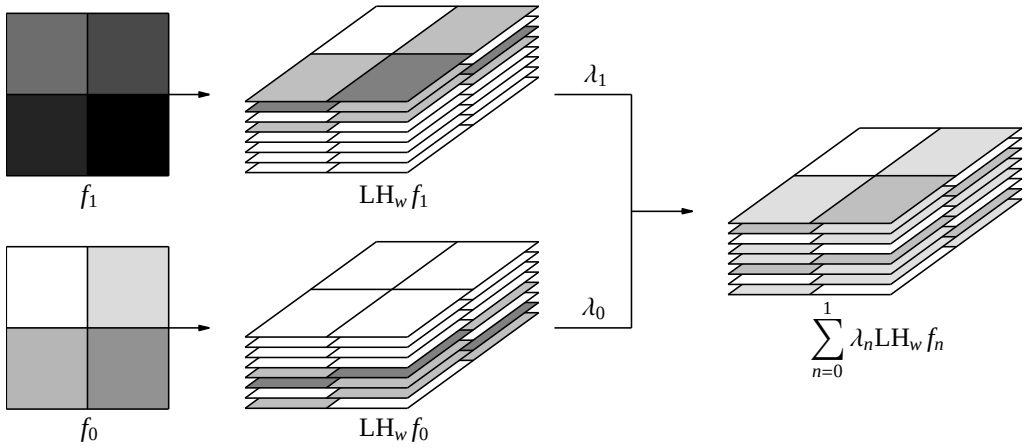


Figure 6. A continuation of the example of Figure 5. When the occlusion model Φ is flat, Theorem 3 becomes Theorem 4, with (10) simplifying to (17). Though each of the 16 distinct composite images shown in Figure 5 has a distinct local histogram transform, the average of these 16 local histogram transforms 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 Figure 5 is equal to the average-over-all-sources local histogram computed above.

To show this, we enumerate $\mathcal{X}$ as $\mathcal{X} = \{x_k\}_{k=1}^K$ and note that for any x_k, we have $\sum_{n_k=0}^{N-1} 1_\varphi(x_k, n_k) = 1$ thus giving our claim:

$$\begin{aligned}
\sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} F(\varphi) &= \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} F(\varphi) \prod_{k=1}^{K} \sum_{n_k=0}^{N-1} 1_\varphi(x_k, n_k) \\
&= \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} F(\varphi) \sum_{n_1=0}^{N-1} 1_\varphi(x_1, n_1) \cdots \sum_{n_K=0}^{N-1} 1_\varphi(x_K, n_K) \\
&= \sum_{n_1=0}^{N-1} \cdots \sum_{n_K=0}^{N-1} \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} F(\varphi) \prod_{k=1}^{K} 1_\varphi(x_k, n_k) \\
&= \sum_{n_1=0}^{N-1} \cdots \sum_{n_K=0}^{N-1} \sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x_k)=n_k, \forall k}} F(\varphi).
\end{aligned}$$

To show Φ is flat, fix some $x_{k_0} \in \mathcal{X}$ and note that enumerating $\mathcal{X}$ gives that:

$$\mathrm{P}_\Phi(\varphi) = \prod_{x \in \mathcal{X}} \rho_{\varphi(x)} = \prod_{k=1}^{K} \rho_{\varphi(x_k)}.$$

Using this, and letting $F(\varphi) = \mathrm{P}_\Phi(\varphi)1_\varphi(x_{k_0}, n)$ in our claim (18), gives:

$$\begin{aligned}\sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)1_\varphi(x_{k_0}, n) &= \sum_{n_1=0}^{N-1}\cdots\sum_{n_K=0}^{N-1} \sum_{\substack{\varphi\in\mathbb{Z}_N^{\mathcal{X}}\\ \varphi(x_k)=n_k,\,\forall k}} \mathrm{P}_\Phi(\varphi)1_\varphi(x_{k_0}, n)\\ &= \sum_{n_1=0}^{N-1}\cdots\sum_{n_K=0}^{N-1} \sum_{\substack{\varphi\in\mathbb{Z}_N^{\mathcal{X}}\\ \varphi(x_k)=n_k,\,\forall k}} \prod_{k=1}^{K} \rho_{\varphi(x_k)} 1_\varphi(x_{k_0}, n).\end{aligned}$$

Note that since $\varphi(x_k) = n_k$ for all k, we have $\rho_{\varphi(x_k)} = \rho_{n_k}$ for all k, thus giving:

$$\begin{aligned}\sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)1_\varphi(x_{k_0}, n) &= \sum_{n_1=0}^{N-1}\cdots\sum_{n_K=0}^{N-1} \sum_{\substack{\varphi\in\mathbb{Z}_N^{\mathcal{X}}\\ \varphi(x_k)=n_k,\,\forall k}} \prod_{k=1}^{K} \rho_{n_k} 1_\varphi(x_{k_0}, n)\\ &= \sum_{n_1=0}^{N-1}\rho_{n_1}\cdots\sum_{n_K=0}^{N-1}\rho_{n_K} \sum_{\substack{\varphi\in\mathbb{Z}_N^{\mathcal{X}}\\ \varphi(x_k)=n_k,\,\forall k}} 1_\varphi(x_{k_0}, n).\end{aligned}$$

Here, we note that for the sum over $\varphi \in \mathbb{Z}_N^{\mathcal{X}}$, the constraints that $\varphi(x_k) = n_k$ for all k uniquely define a single φ. Therefore, this sum is one if $n = n_{k_0}$:

$$\begin{aligned}\sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)1_\varphi(x_{k_0}, n) &= \sum_{n_1=0}^{N-1}\rho_{n_1}\cdots\sum_{n_K=0}^{N-1}\rho_{n_K}\delta_n(n_{k_0})\\ &= \Bigg(\sum_{n_{k_0}=0}^{N-1}\rho_{n_{k_0}}\delta_n(n_{k_0})\Bigg)\prod_{k\neq k_0}\Bigg(\sum_{n_k=0}^{N-1}\rho_{n_k}\Bigg)\\ &= \rho_n.\end{aligned}$$

□

Next, we show that an occlusion model Φ is flat if it is *translation-invariant*, meaning that its probability density function P_Φ satisfies:

$$\mathrm{P}_\Phi(\mathrm{T}^x\varphi) = \mathrm{P}_\Phi(\varphi), \quad \forall\varphi\in\mathbb{Z}_N^{\mathcal{X}}, x\in\mathcal{X}. \tag{19}$$

Note that, if we consider the value of φ at each given $x \in \mathcal{X}$ as a $\mathbb{Z}_N$-valued random variable, translation-invariance corresponds to the associated stochastic process being stationary.

Theorem 6. *If* Φ *is translation-invariant* (19)*, then* Φ *is flat* (16)*.*

Proof. To show that Φ is flat (16), we must show that for any $n \in \mathbb{Z}_N$, $\overline{1}(x, n)$ is constant with respect to x. Using (8), and noting that the definition of the translation operator gives that $\varphi(x) = \mathrm{T}^{x_0 - x}\varphi(x_0)$ for any $x_0 \in \mathcal{X}$, we have:

$$\overline{1}(x, n) = \sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x) = n}} \mathrm{P}_\Phi(\varphi) = \sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \mathrm{T}^{x_0 - x}\varphi(x_0) = n}} \mathrm{P}_\Phi(\varphi).$$

Now, making a change of variables by letting $\psi = \mathrm{T}^{x_0 - x}\varphi$, and using our assumption that Φ is translation-invariant (19) gives:

$$\overline{1}_\Phi(x, n) = \sum_{\substack{\mathrm{T}^{x - x_0}\psi \in \mathbb{Z}_N^{\mathcal{X}} \\ \psi(x_0) = n}} \mathrm{P}_\Phi(\mathrm{T}^{x - x_0}\psi) = \sum_{\substack{\psi \in \mathbb{Z}_N^{\mathcal{X}} \\ \psi(x_0) = n}} \mathrm{P}_\Phi(\psi) = \overline{1}_\Phi(x_0, n). \qquad \square$$

Theorem 6 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 $\mathrm{P}_\Phi(\varphi)$ is solely determined by the number of times that φ achieves each given value n. Note this implies that the Bernoulli model is flat, a fact we independently observed in Theorem 5. 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 translations, and assign probability 0 to all others; if the source images $\{f_n\}_{n=0}^{N-1}$ are constant, the composite images (7) produced

by such a model are all translates of a single image. More generally, we can always partition the $N^{|\mathcal{X}|}$ elements of $\mathbb{Z}_N^{\mathcal{X}}$ 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 $\mathcal{X} = \mathbb{Z}_2 \oplus \mathbb{Z}_2$ depicted in Figure 5, we may partition the 16 possible φ's into 7 such classes, and pick any probabilities $\{\mathrm{p}_k\}_{k=0}^{15}$ such that $\mathrm{p}_1 = \mathrm{p}_2 = \mathrm{p}_3 = \mathrm{p}_4$, $\mathrm{p}_5 = \mathrm{p}_6$, $\mathrm{p}_7 = \mathrm{p}_8$, $\mathrm{p}_9 = \mathrm{p}_{10}$, $\mathrm{p}_{11} = \mathrm{p}_{12} = \mathrm{p}_{13} = \mathrm{p}_{14}$. Armed with one general method—translation-invariance—for producing flat models Φ, we now turn to ways of combining known models to produce more complicated and realistic ones.

3.3.1 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 4, doing so would formally justify the decomposition 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 $\varphi \in \mathbb{Z}_2^{\mathcal{X}}$ 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 $\psi_x \in \mathbb{Z}_2^{\mathcal{X}}$. The new texture will be the union of all such blobs.

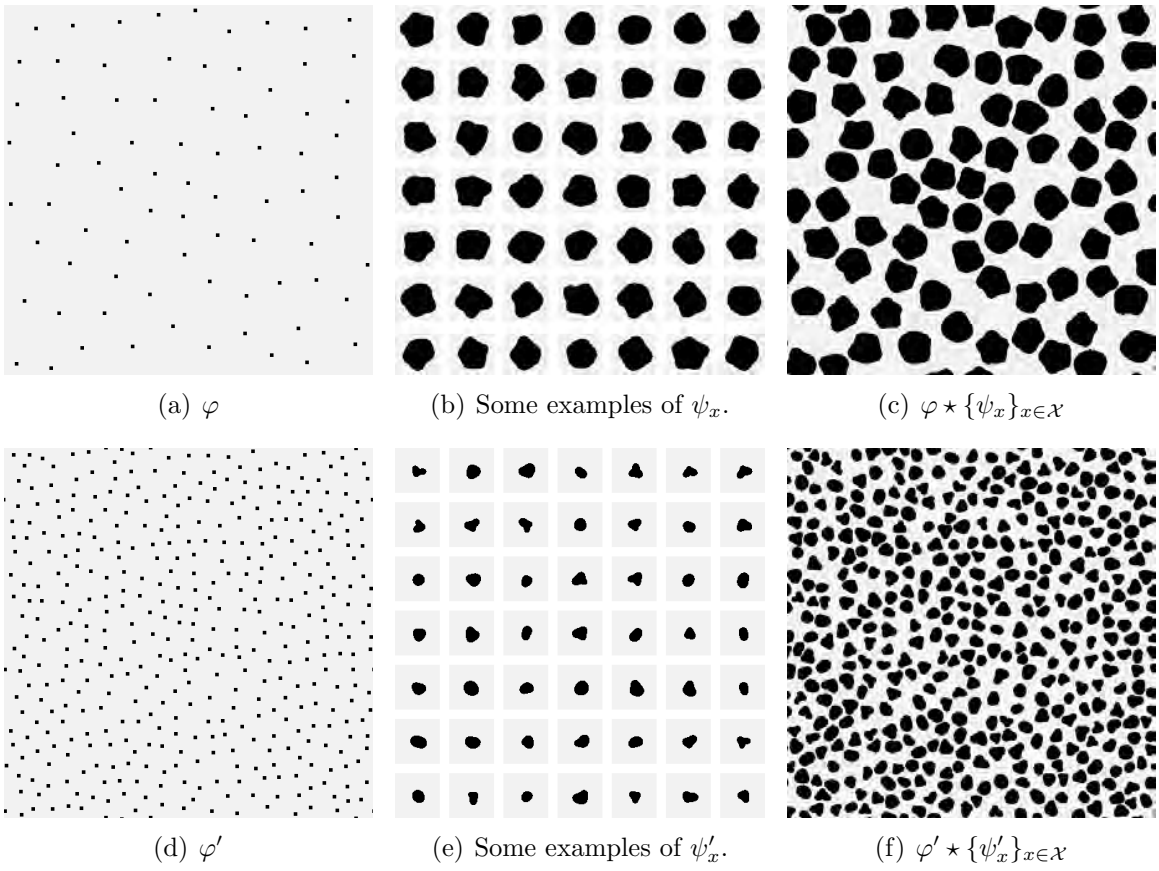
Formally, given any $\varphi \in \mathbb{Z}_2^{\mathcal{X}}$ and $\{\psi_x\}_{x\in\mathcal{X}} \in [\mathbb{Z}_2^{\mathcal{X}}]^{\mathcal{X}}$, we define the *expansion* of φ by $\{\psi_x\}_{x\in\mathcal{X}}$ to be $\varphi \star \{\psi_x\}_{x\in\mathcal{X}} \in \mathbb{Z}_2^{\mathcal{X}}$,

$$(\varphi \star \{\psi_{x'}\}_{x'\in\mathcal{X}})(x) := \begin{cases} 1, & x = x' + x'', \varphi(x') = 1, \psi_{x'}(x'') = 1, \\ 0, & \text{else.} \end{cases} \tag{20}$$

Two examples of this expansion operation are given in Figure 7. Note that expansion itself (20) is not an occlusion model. Indeed, (20) is but a way of combining functions in $\mathbb{Z}_2^{\mathcal{X}}$ to produce other ones, whereas an occlusion model is a random variable Φ defined by a probability density function P_Φ over $\mathbb{Z}_2^{\mathcal{X}}$. This fact notwithstanding, the expansion operation (20) 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 $\mathrm{P}_{\Phi\star\Psi} : \mathbb{Z}_2^{\mathcal{X}} \to [0, 1]$,

$$\mathrm{P}_{\Phi\star\Psi}(\sigma) := \sum_{\substack{\varphi\in\mathbb{Z}_2^{\mathcal{X}} \\ \{\psi_x\}_{x\in\mathcal{X}}\in[\mathbb{Z}_2^{\mathcal{X}}]^{\mathcal{X}} \\ \varphi\star\{\psi_x\}_{x\in\mathcal{X}}=\sigma}} \mathrm{P}_\Phi(\varphi) \prod_{x\in\mathcal{X}} \mathrm{P}_\Psi(\psi_x). \tag{21}$$

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 (21) indeed defines a probability density function on $\mathbb{Z}_2^{\mathcal{X}}$. We further show that if Φ is translation-invariant (19), then $\Phi \star \Psi$ is translation-invariant which implies that $\Phi \star \Psi$ is flat by Theorem 6. 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



(a) φ (b) Some examples of ψ_x. (c) $\varphi \star \{\psi_x\}_{x\in\mathcal{X}}$

(d) φ' (e) Some examples of ψ'_x. (f) $\varphi' \star \{\psi'_x\}_{x\in\mathcal{X}}$

Figure 7. Examples of the expansion operation (20), where black denotes the value of 1, and the lighter shade denotes the value of 0. A function $\varphi : \mathcal{X} \to \{0, 1\}$ 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 $\psi_x : \mathcal{X} \to \{0, 1\}$. 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 (20) 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 7.

requirement that Φ be translation-invariant so as to only require that Φ is itself flat, provided Φ and Ψ are *effectively disjoint*:

$$\text{If } \mathrm{P}_\Phi(\varphi) > 0 \text{ and } \mathrm{P}_\Psi(\psi_x) > 0 \text{ for all } x \in \mathcal{X}, \text{ then } \varphi \star \{\psi_x\}_{x\in\mathcal{X}} = \sum_{\substack{x\in\mathcal{X}\\ \varphi(x)=1}} \mathrm{T}^x\psi_x. \quad (22)$$

Put another way, (22) means that there is only at most one way, with nontrivial probability, in which the x in (20) can be written as $x = x' + x''$ where both $\varphi(x') = 1$ and $\psi_{x'}(x'') = 1$.

Theorem 7. *If Φ and Ψ are occlusion models from $\mathcal{X}$ into $\mathbb{Z}_2$, then their expansion $\Phi\star\Psi$, with probability density function* (21)*, is as well. Moreover, if Φ is translation-invariant* (19)*, then $\Phi \star \Psi$ is translation-invariant. Furthermore, if Φ and Ψ are effectively disjoint* (22) *and either Φ or Ψ is flat* (16)*, then $\Phi \star \Psi$ is flat.*

Proof. We first show that (21) defines a probability density function, namely that values of $\mathrm{P}_{\Phi\star\Psi}(\sigma)$ over all σ in $\mathbb{Z}_2^{\mathcal{X}}$ sum to 1. Since P_Φ is a probability density function by assumption, we have:

$$1 = \sum_{\varphi\in\mathbb{Z}_2^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi). \quad (23)$$

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

$$1 = \sum_{\psi_x\in\mathbb{Z}_2^{\mathcal{X}}} \mathrm{P}_\Psi(\psi_x), \quad (24)$$

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 (23) with the product of (24) over all x yields:

$$1 = 1(1)^{|\mathcal{X}|} = \sum_{\varphi \in \mathbb{Z}_2^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) \prod_{x \in \mathcal{X}} \sum_{\psi_x \in \mathbb{Z}_2^{\mathcal{X}}} \mathrm{P}_\Psi(\psi_x) = \sum_{\substack{\varphi \in \mathbb{Z}_2^{\mathcal{X}} \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\mathbb{Z}_2^{\mathcal{X}}]^{\mathcal{X}}}} \mathrm{P}_\Phi(\varphi) \prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x), \qquad (25)$$

where the final quantity in (25) contains all of the cross terms resulting from distributing the product over all sums of the form (24). 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 (25) in terms of the definition (21) of $\mathrm{P}_{\Phi \star \Psi}$, obtaining our claim:

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

Thus, (21) 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 first claim that if $\mathrm{T}^{\tilde{x}}\sigma = \varphi \star \{\psi_x\}_{x \in \mathcal{X}}$, then $\sigma = (\mathrm{T}^{-\tilde{x}}\varphi) \star \{\psi_{x+\tilde{x}}\}_{x \in \mathcal{X}}$. To see this claim, note that

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

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 $(\mathrm{T}^{-\tilde{x}}\varphi)(x' - \tilde{x}) = \varphi(x' - \tilde{x} + \tilde{x}) = \varphi(x') = 1$ and $\psi_{(x'-\tilde{x})+\tilde{x}}(x'') = \psi_{x'}(x'') = 1$, implying $\sigma = (\mathrm{T}^{-\tilde{x}}\varphi) \star \{\psi_{x+\tilde{x}}\}_{x \in \mathcal{X}}$, as claimed. Having

the claim, (21) implies:

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

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

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

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

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

and so $\Phi\star\Psi$ is indeed translation-invariant (19), as claimed.

For our final claim, we assume that Φ and Ψ are effectively disjoint (22) 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 $\overline{\Phi} := \sum_{\varphi\in\mathbb{Z}_2^{\mathcal{X}}}\mathrm{P}_\Phi(\varphi)\varphi$. Indeed, for any φ from $\mathcal{X}$ to $\mathbb{Z}_2$, (2) may be rewritten as $1_\varphi(x,1) = \varphi(x)$ and so:

$$\overline{1}_\Phi(x,1) = \sum_{\varphi\in\mathbb{Z}_2^{\mathcal{X}}}\mathrm{P}_\Phi(\varphi)1_\varphi(x,1) = \sum_{\varphi\in\mathbb{Z}_2^{\mathcal{X}}}\mathrm{P}_\Phi(\varphi)\varphi(x) = \overline{\Phi}(x). \tag{26}$$

In light of (26), we claim that Φ is flat if and only if $\overline{\Phi}$ is constant. Indeed, if Φ is flat, then there exists λ_1 such that $\overline{\Phi}(x) = \overline{1}_\Phi(x,1) = \lambda_1$ for all $x\in\mathcal{X}$. Conversely, if $\overline{\Phi}(x)$ is constant, then there exists λ_1 such that $\overline{1}_\Phi(x,1) = \overline{\Phi}(x) = \lambda_1$ for all $x\in\mathcal{X}$;

by (11), this further implies that $\overline{1}_\Phi(x,0) = 1 - \overline{1}_\Phi(x,1) = 1 - \lambda_1$ 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 $\overline{\Phi \star \Psi} = \overline{\Phi} * \overline{\Psi}$ where "$*$" denotes standard convolution over $\mathcal{X}$. According to the definition of $\Phi \star \Psi$ (21) we have:

$$\overline{\Phi \star \Psi} = \sum_{\sigma \in \mathbb{Z}_2^{\mathcal{X}}} \mathrm{P}_{\Phi\star\Psi}(\sigma)\sigma = \sum_{\sigma \in \mathbb{Z}_2^{\mathcal{X}}} \sum_{\substack{\varphi \in \mathbb{Z}_2^{\mathcal{X}} \\ \{\psi_x\}_{x\in\mathcal{X}} \in [\mathbb{Z}_2^{\mathcal{X}}]^{\mathcal{X}} \\ \varphi\star\{\psi_x\}_{x\in\mathcal{X}} = \sigma}} \mathrm{P}_\Phi(\varphi) \left(\prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x) \right) (\varphi \star \{\psi_x\}_{x\in\mathcal{X}}). \quad (27)$$

Since any particular choice of φ and $\{\psi_x\}_{x\in\mathcal{X}}$ produces a unique σ via $\star$ we can simplify (27) to

$$\overline{\Phi \star \Psi} = \sum_{\substack{\varphi \in \mathbb{Z}_2^{\mathcal{X}} \\ \{\psi_x\}_{x\in\mathcal{X}} \in [\mathbb{Z}_2^{\mathcal{X}}]^{\mathcal{X}}}} \mathrm{P}_\Phi(\varphi) \left(\prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x) \right) (\varphi \star \{\psi_x\}_{x\in\mathcal{X}}). \quad (28)$$

Moreover, since Φ and Ψ are effectively disjoint (22) we have $\varphi\star\{\psi_x\}_{x\in\mathcal{X}} = \sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} \mathrm{T}^{x'}\psi_{x'}$ meaning (28) becomes:

$$\begin{aligned} \overline{\Phi \star \Psi} &= \sum_{\substack{\varphi \in \mathbb{Z}_2^{\mathcal{X}} \\ \{\psi_x\}_{x\in\mathcal{X}} \in [\mathbb{Z}_2^{\mathcal{X}}]^{\mathcal{X}}}} \mathrm{P}_\Phi(\varphi) \left(\prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x) \right) \left(\sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} \mathrm{T}^{x'}\psi_{x'} \right) \\ &= \sum_{\varphi \in \mathbb{Z}_2^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) \sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} \mathrm{T}^{x'} \left[\sum_{\{\psi_x\}_{x\in\mathcal{X}} \in [\mathbb{Z}_2^{\mathcal{X}}]^{\mathcal{X}}} \left(\prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x) \right) \psi_{x'} \right]. \end{aligned} \quad (29)$$

Now, for any fixed $x' \in \mathcal{X}$ such that $\varphi(x') = 1$, we factor the corresponding innermost

sum in (29) into a product of $|\mathcal{X}|$ distinct sums—one for each $x \in \mathcal{X}$—to obtain:

$$\begin{aligned}\sum_{\{\psi_x\}_{x\in\mathcal{X}}\in[\mathbb{Z}_2^{\mathcal{X}}]^{\mathcal{X}}}\left(\prod_{x\in\mathcal{X}}\mathrm{P}_\Psi(\psi_x)\right)\psi_{x'} &= \left[\prod_{x\neq x'}\left(\sum_{\psi_x\in\mathbb{Z}_2^{\mathcal{X}}}\mathrm{P}_\Psi(\psi_x)\right)\right]\sum_{\psi_{x'}\in\mathbb{Z}_2^{\mathcal{X}}}\mathrm{P}_\Psi(\psi_{x'})\psi_{x'} \\ &= \left(\prod_{x\neq x'}1\right)\overline{\Psi} \\ &= \overline{\Psi}. \end{aligned} \tag{30}$$

Substituting (30) into (29) then gives:

$$\begin{aligned}\overline{\Phi\star\Psi} &= \sum_{\varphi\in\mathbb{Z}_2^{\mathcal{X}}}\mathrm{P}_\Phi(\varphi)\sum_{\substack{x'\in\mathcal{X}\\ \varphi(x')=1}}\mathrm{T}^{x'}\overline{\Psi} \\ &= \sum_{\varphi\in\mathbb{Z}_2^{\mathcal{X}}}\mathrm{P}_\Phi(\varphi)\left(\sum_{\substack{x'\in\mathcal{X}\\ \varphi(x')=1}}\delta_{x'}\right)*\overline{\Psi} \\ &= \left(\sum_{\varphi\in\mathbb{Z}_2^{\mathcal{X}}}\mathrm{P}_\Phi(\varphi)\varphi\right)*\overline{\Psi} \\ &= \overline{\Phi}*\overline{\Psi}.\end{aligned}$$

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. □

3.3.2 Overlay.

Above, we discussed how the expansion (21) 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 (20). Under certain hypotheses, Theorem 7 gives that such models $\Phi\star\Psi$ are flat, meaning their local histograms can be understood in terms of Theorem 4. Moreover, some examples of these models produce

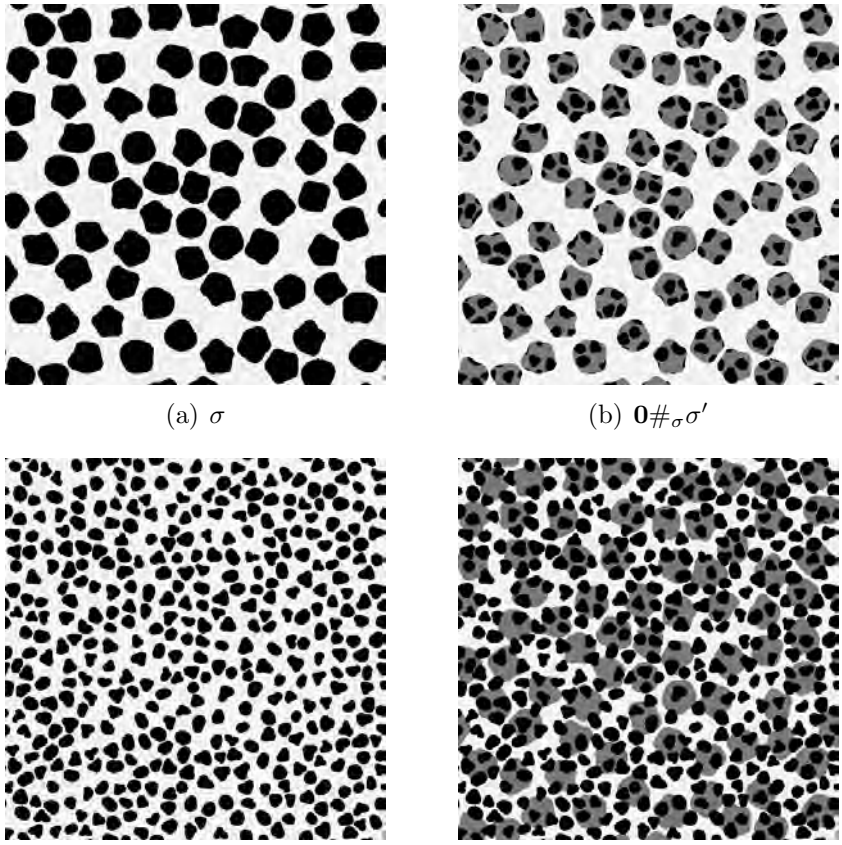
textures that resemble those encountered in histological tissues: if f_0 and f_1 are roughly constant light purple and dark purple fields, respectively, then the composite image $\mathrm{occ}_\varphi\{f_0, f_1\}$ obtained by picking φ as in Figure 7(f) bears some similarity to an actual image of cartilage, such as the one given in Figure 3(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 (20), models produced by expansion (21) 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 Figure 3(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 $\varphi \in \mathbb{Z}_{N_1}^{\mathcal{X}}$, $\psi \in \mathbb{Z}_{N_2}^{\mathcal{X}}$ and $\sigma \in \mathbb{Z}_2^{\mathcal{X}}$, we define the *overlay* of φ over ψ with respect to σ to be $\varphi \#_\sigma \psi \in \mathbb{Z}_{N_1+N_2}^{\mathcal{X}}$,

$$(\varphi \#_\sigma \psi)(x) := \begin{cases} \varphi(x), & \sigma(x) = 0, \\ \psi(x) + N_1, & \sigma(x) = 1. \end{cases} \tag{31}$$

Essentially, an overlay (31) 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_1 so that they cannot be confused with those of φ. Examples of this overlay operation are given in Figure 8.

In a manner similar to the relationship between (20) and (21), we have that (31) naturally induces a parallel operation on occlusion models: given probability density functions P_Φ, P_Ψ and P_Σ on $\mathbb{Z}_{N_\Phi}^{\mathcal{X}}$, $\mathbb{Z}_{N_\Psi}^{\mathcal{X}}$ and $\mathbb{Z}_2^{\mathcal{X}}$, respectively, we define the overlay of the occlusion model Φ over Ψ with respect to Σ to be the new occlusion model



(a) σ (b) $\mathbf{0}\#_\sigma\sigma'$

(c) σ' (d) $\sigma\#_{\sigma'}\mathbf{0}$

Figure 8. Two examples of the overlay operation (31). 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 Figure 7(c) and (f) reshown here in (a) and (c), respectively. Further consider a constant function $\mathbf{0} : \mathcal{X} \to \mathbb{Z}_1$ that assigns label 0 to every point in $\mathcal{X}$. The overlay (31) of $\mathbf{0}$ over σ' is given in (b); essentially, σ-shaped holes are cut from $\mathbf{0}$ 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 $\mathbf{0}$ (d). Overlaying the resulting textures with each other can produce even more complicated textures.

$\Phi\#_\Sigma\Psi$ whose probability density function is $\mathrm{P}_{\Phi\#_\Sigma\Psi} : \mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}} \to [0,1]$,

$$\mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon) := \sum_{\substack{\varphi\in\mathbb{Z}_{N_\Phi}^{\mathcal{X}} \\ \psi\in\mathbb{Z}_{N_\Psi}^{\mathcal{X}} \\ \sigma\in\mathbb{Z}_2^{\mathcal{X}} \\ \varphi\#_\sigma\psi=\upsilon}} \mathrm{P}_\Phi(\varphi)\mathrm{P}_\Psi(\psi)\mathrm{P}_\Sigma(\sigma). \tag{32}$$

In the next result, we verify that (32) 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 (7) produced by such a model will behave according to Theorem 4.

Theorem 8. *If Φ, Ψ and Σ are occlusion models on $\mathbb{Z}_{N_\Phi}^{\mathcal{X}}$, $\mathbb{Z}_{N_\Psi}^{\mathcal{X}}$ and $\mathbb{Z}_2^{\mathcal{X}}$, respectively, then* (32) *defines a probability density function on $\mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}}$. Moreover, if Φ, Ψ, and Σ are flat, then $\Phi\#_\Sigma\Psi$ is flat.*

Proof. To show that (32) defines a probability density function on $\mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}}$, note that:

$$1 = (1)(1)(1) = \sum_{\varphi\in\mathbb{Z}_{N_\Phi}^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) \sum_{\psi\in\mathbb{Z}_{N_\Psi}^{\mathcal{X}}} \mathrm{P}_\Psi(\psi) \sum_{\sigma\in\mathbb{Z}_2^{\mathcal{X}}} \mathrm{P}_\Sigma(\sigma) = \sum_{\substack{\varphi\in\mathbb{Z}_{N_\Phi}^{\mathcal{X}} \\ \psi\in\mathbb{Z}_{N_\Psi}^{\mathcal{X}} \\ \sigma\in\mathbb{Z}_2^{\mathcal{X}}}} \mathrm{P}_\Phi(\varphi)\mathrm{P}_\Psi(\psi)\mathrm{P}_\Sigma(\sigma). \tag{33}$$

Noting that for each fixed φ, ψ, and σ, there exists exactly one $\upsilon \in \mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}}$ such that $\varphi\#_\sigma\psi = \upsilon$, (33) becomes:

$$1 = \sum_{\upsilon\in\mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}}} \sum_{\substack{\varphi\in\mathbb{Z}_{N_\Phi}^{\mathcal{X}} \\ \psi\in\mathbb{Z}_{N_\Psi}^{\mathcal{X}} \\ \sigma\in\mathbb{Z}_2^{\mathcal{X}} \\ \varphi\#_\sigma\psi=\upsilon}} \mathrm{P}_\Phi(\varphi)\mathrm{P}_\Psi(\psi)\mathrm{P}_\Sigma(\sigma) = \sum_{\upsilon\in\mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}}} \mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon),$$

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 (16), 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:

$$\sum_{\substack{\upsilon\in\mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}}\\ \upsilon(x)=n}} \mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon) = \lambda_n \tag{34}$$

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

$$\begin{aligned}\sum_{\substack{\upsilon\in\mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}}\\ \upsilon(x)=n}} \mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon) &= \sum_{\substack{\upsilon\in\mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}}\\ \upsilon(x)=n}} \sum_{\substack{\varphi\in\mathbb{Z}_{N_\Phi}^{\mathcal{X}}\\ \psi\in\mathbb{Z}_{N_\Psi}^{\mathcal{X}}\\ \sigma\in\mathbb{Z}_2^{\mathcal{X}}\\ \varphi\#_\sigma\psi=\upsilon}} \mathrm{P}_\Phi(\varphi)\mathrm{P}_\Psi(\psi)\mathrm{P}_\Sigma(\sigma)\\ &= \sum_{\substack{\varphi\in\mathbb{Z}_{N_\Phi}^{\mathcal{X}}\\ \psi\in\mathbb{Z}_{N_\Psi}^{\mathcal{X}}\\ \sigma\in\mathbb{Z}_2^{\mathcal{X}}\\ (\varphi\#_\sigma\psi)(x)=n}} \mathrm{P}_\Phi(\varphi)\mathrm{P}_\Psi(\psi)\mathrm{P}_\Sigma(\sigma). \end{aligned}\tag{35}$$

Now, in the special case where $n = 0, \ldots, N_\Phi - 1$, (31) gives that $(\varphi\#_\sigma\psi)(x) = n$ if and only if $\varphi(x) = n$ and $\sigma(x) = 0$. As such, in this case (35) becomes:

$$\begin{aligned}\sum_{\substack{\upsilon\in\mathbb{Z}_{N_\Phi+N_\Psi}^{\mathcal{X}}\\ \upsilon(x)=n}} \mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon) &= \sum_{\substack{\varphi\in\mathbb{Z}_{N_\Phi}^{\mathcal{X}},\varphi(x)=n\\ \psi\in\mathbb{Z}_{N_\Psi}^{\mathcal{X}}\\ \sigma\in\mathbb{Z}_2^{\mathcal{X}},\sigma(x)=0}} \mathrm{P}_\Phi(\varphi)\mathrm{P}_\Psi(\psi)\mathrm{P}_\Sigma(\sigma)\\ &= \sum_{\substack{\varphi\in\mathbb{Z}_{N_\Phi}^{\mathcal{X}}\\ \varphi(x)=n}} \mathrm{P}_\Phi(\varphi) \sum_{\psi\in\mathbb{Z}_{N_\Psi}^{\mathcal{X}}} \mathrm{P}_\Psi(\psi) \sum_{\substack{\sigma\in\mathbb{Z}_2^{\mathcal{X}}\\ \sigma(x)=0}} \mathrm{P}_\Sigma(\sigma)\\ &= \lambda_{\Phi,n}\lambda_{\Sigma,0}. \end{aligned}\tag{36}$$

If, on the other hand $n = N_\Phi, \ldots, N_\Phi + N_\Psi - 1$ then (31) gives that $(\varphi\#_\sigma\psi)(x) = n$

if and only if $\psi(x) = n - N_\Phi$ and $\sigma(x) = 1$. In this case, (35) becomes:

$$\begin{aligned}\sum_{\substack{\upsilon\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi+N_\Psi}\\ \upsilon(x)=n}} \mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon) &= \sum_{\substack{\varphi\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi}\\ \psi\in\mathbb{Z}^{\mathcal{X}}_{N_\Psi},\psi(x)=n-N_\Phi\\ \sigma\in\mathbb{Z}^{\mathcal{X}}_{2},\sigma(x)=1}} \mathrm{P}_\Phi(\varphi)\mathrm{P}_\Psi(\psi)\mathrm{P}_\Sigma(\sigma)\\ &= \sum_{\varphi\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi}} \mathrm{P}_\Phi(\varphi) \sum_{\substack{\psi\in\mathbb{Z}^{\mathcal{X}}_{N_\Psi}\\ \psi(x)=n-N_\Phi}} \mathrm{P}_\Psi(\psi) \sum_{\substack{\sigma\in\mathbb{Z}^{\mathcal{X}}_{2}\\ \sigma(x)=1}} \mathrm{P}_\Sigma(\sigma)\\ &= \lambda_{\Psi,n-N_\Phi}\lambda_{\Sigma,1}. \end{aligned} \tag{37}$$

Thus, for any $x \in \mathcal{X}$ we either have (36) or (37) meaning $\Phi\#_\Sigma\Psi$ is flat (34), as claimed. □

Up to this point, we have considered the appropriateness of using local histograms to segment and classifiy histology images like the one in Figure 1(a). In particular, we have shown that local histograms have several useful properties: they commute with translations on $\mathcal{X}$ and $\mathcal{Y}$, and they distribute, on average, over occlusions provided that our occlusion model is flat. In the next chapter, we characterize all transforms which satisfy these three key properties of local histograms. In doing so, we demonstrate the significance that local histograms have in analyzing textures commonly found in histology images.

IV. Characterizing Local Histogram Transforms

In this chapter, we characterize all (possibly nonlinear) transforms which possess the nice occlusion-handling properties of local histograms. Our main result here is Theorem 10, which formally shows that these properties are almost unique to local histograms. In particular, Theorem 4 gives that the local histogram transform of a random composite image (7) produced by a flat occlusion model Φ is, on average, a convex combination of the local histogram transforms of the source images $\{f_n\}_{n=0}^{N-1}$. Here, the coefficients in this convex combination are the values $\{\lambda_n\}_{n=0}^{N-1}$ defined in (16). In the next result, we show that this is not an accident, namely that if any transform distributes on average over occlusions in this manner, the resulting coefficients are necessarily these particular values:

Theorem 9. *Let Θ be any (possibly nonlinear) transformation from $\mathcal{Y}^{\mathcal{X}}$ into any real vector space. If $\Theta(\delta_x) \neq \Theta(0)$ for some $x \in \mathcal{X}$ and if, for some flat occlusion model Φ, there exists scalars $\{c_n\}_{n=0}^{N-1}$ such that:*

$$\mathrm{E}_{\Phi}[\Theta(\mathrm{occ}_{\Phi}\{f_n\}_{n=0}^{N-1})] = \sum_{n=0}^{N-1} c_n \Theta(f_n), \tag{38}$$

for all $\{f_n\}_{n=0}^{N-1}$ in $\mathcal{Y}^{\mathcal{X}}$, then these scalars are necessarily:

$$c_n = \lambda_n = \sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x)=n}} \mathrm{P}_{\Phi}(\varphi) \qquad \forall\ n = 0, \ldots, N-1. \tag{39}$$

Proof. Fixing any such flat Φ, our assumption (38) states:

$$\sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_{\Phi}(\varphi) \Theta(\mathrm{occ}_{\varphi}\{f_n\}_{n=0}^{N-1}) = \sum_{n=0}^{N-1} c_n \Theta(f_n) \tag{40}$$

for all $\{f_n\}_{n=0}^{N-1}$ in $\mathcal{Y}^{\mathcal{X}}$. Note that $\Theta(f) \neq 0$ for some $f \in \mathcal{Y}^{\mathcal{X}}$: since $\Theta(\delta_x) \neq \Theta(0)$ for some $x \in \mathcal{X}$, at least one of the selections $f = 0$ or $f = \delta_x$ will suffice. Letting $f_n = f$ for all n gives that $\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1} = f$ for all $\varphi \in \mathbb{Z}_N^{\mathcal{X}}$. Noting that the values $\mathrm{P}_\Phi(\varphi)$ sum to one, (40) in this special case becomes:

$$\Theta(f) = \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)\Theta(f) = \sum_{n=0}^{N-1} c_n \Theta(f) = \left(\sum_{n=0}^{N-1} c_n \right) \Theta(f),$$

which, since $\Theta(f) \neq 0$, necessarily implies $\sum_{n=0}^{N-1} c_n = 1$. Having this fact, we again let the f_n's be arbitrary, and subtract $\Theta(0)$ from (40):

$$\sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)[\Theta(\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}) - \Theta(0)] = \sum_{n=0}^{N-1} c_n[\Theta(f_n) - \Theta(0)]. \tag{41}$$

Now, fix $x_0 \in \mathcal{X}$ such that $\Theta(\delta_{x_0}) \neq \Theta(0)$, that is, $\Theta(\delta_{x_0}) - \Theta(0) \neq 0$. Also, fix any $n_0 = 0, \ldots, N-1$, and let

$$f_n = \begin{cases} \delta_{x_0}, & n = n_0, \\ 0, & n \neq n_0. \end{cases} \tag{42}$$

That is, for any $x \in \mathcal{X}$:

$$f_n(x) = \begin{cases} 1, & n = n_0, x = x_0, \\ 0, & \text{else.} \end{cases} \tag{43}$$

We claim that:

$$\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1} = \begin{cases} \delta_{x_0}, & \varphi(x_0) = n_0, \\ 0, & \varphi(x_0) \neq n_0. \end{cases} \tag{44}$$

Note this is equivalent to showing that for any $x \in \mathcal{X}$,

$$\begin{aligned} f_{\varphi(x)}(x) &= (\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})(x) \\ &= \begin{cases} \delta_{x_0}(x), & \varphi(x_0) = n_0, \\ 0, & \varphi(x_0) \neq n_0, \end{cases} \\ &= \begin{cases} 1, & x = x_0, \varphi(x_0) = n_0, \\ 0, & \text{else.} \end{cases} \end{aligned} \tag{45}$$

For $x \neq x_0$, the left-hand-side of (45) is $f_{\varphi(x)}(x) = 0$ by (43), while the right-hand-side of (45) is clearly zero. Meanwhile for $x = x_0$, evaluating (43) at $x = x_0$ and $n = \varphi(x_0)$ gives the left-hand-side of (45) to be:

$$f_{\varphi(x_0)}(x_0) = \begin{cases} 1, & \varphi(x_0) = n_0, \\ 0, & \text{else,} \end{cases}$$

which equals the right-hand-side of (45) at $x = x_0$.

Having (44), we can decompose the left-hand-sum of (41) as:

$$\begin{aligned} &\sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)[\Theta(\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}) - \Theta(0)] \\ &\qquad = \sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x_0) = n_0}} \mathrm{P}_\Phi(\varphi)[\Theta(\delta_{x_0}) - \Theta(0)] + \sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x_0) \neq n_0}} \mathrm{P}_\Phi(\varphi)[\Theta(0) - \Theta(0)] \\ &\qquad = \sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x_0) = n_0}} \mathrm{P}_\Phi(\varphi)[\Theta(\delta_{x_0}) - \Theta(0)]. \end{aligned}$$

Meanwhile, in light of (42), the right-hand-sum of (41) can be decomposed as:

$$\sum_{n=0}^{N-1} c_n[\Theta(f_n) - \Theta(0)] = c_{n_0}[\Theta(\delta_{x_0}) - \Theta(0)] + \sum_{n \neq n_0} c_n[\Theta(0) - \Theta(0)] = c_{n_0}[\Theta(\delta_{x_0}) - \Theta(0)].$$

That is, in the special case (42), the relation (41) becomes:

$$\sum_{\substack{\varphi\in\mathbb{Z}_N^{\mathcal{X}}\\ \varphi(x_0)=n_0}} \mathrm{P}_\Phi(\varphi)[\Theta(\delta_{x_0})-\Theta(0)] = c_{n_0}[\Theta(\delta_{x_0})-\Theta(0)]. \tag{46}$$

Dividing both sides of (46) by $\Theta(\delta_{x_0})-\Theta(0)\neq 0$ gives our claim (39). □

In light of Theorem 9, we therefore only seek to characterize those transforms Θ which satisfy $\mathrm{E}_\Phi[\Theta(\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})] = \sum_{n=0}^{N-1}\lambda_n\Theta(f_n)$. Though a complete characterization of such transforms remains elusive, we are able to characterize them in the special case where the codomain of Θ is $\mathbb{R}^{\mathcal{X}\oplus\mathcal{Y}}$ and Θ is further assumed to commute with translations in $\mathcal{X}$ and $\mathcal{Y}$; see Theorem 10 below for a precise statement. That is, we show that if Θ transforms images $f:\mathcal{X}\to\mathcal{Y}$ into data cubes $\Theta(f):\mathcal{X}\oplus\mathcal{Y}\to\mathbb{R}$ in a way that distributes over occlusions and commutes with translation, then Θ is necessarily a filtering operation on the graph (2) of f. The proof of this fact uses discrete Fourier transforms (DFTs) over the finite abelian groups $\mathcal{X}$ and $\mathcal{Y}$ of pixel locations and pixel values, respectively.

To be precise, the Fundamental Theorem of Finite Abelian Groups gives that $\mathcal{Y}$ can be written as a direct sum of cyclic groups:

$$\mathcal{Y} := \bigoplus_{k=1}^{K}\mathbb{Z}_{M_k} = \big\{\oplus_{k=1}^{K}y(k) : y(k)\in\mathbb{Z}_{M_k}\ \forall\ k=1,\ldots,K\big\}, \tag{47}$$

that is, $y := (y(1),\ldots,y(K))$. In 8-bit red-green-blue (RGB) images for example, we have $\mathcal{Y}=\mathbb{Z}_{256}\oplus\mathbb{Z}_{256}\oplus\mathbb{Z}_{256}$, that is, $K=3$ and $M_1=M_2=M_3=256$. The DFT over $\mathcal{Y}$ is $\mathrm{F}_\mathcal{Y}:\mathbb{C}^\mathcal{Y}\to\mathbb{C}^\mathcal{Y}$,

$$(\mathrm{F}_\mathcal{Y}g)(y) := \sum_{y'\in\mathcal{Y}} g(y')\mathrm{e}^{-2\pi\mathrm{i}(y\cdot y')},$$

where the (nonstandard) dot product of $y, y' \in \mathcal{Y}$ is defined as:

$$y \cdot y' := \sum_{k=1}^{K} \frac{y(k)y'(k)}{M_k}. \tag{48}$$

Note since each value $y(k)$ is only unique modulo M_k, the value of the dot product (48) is only unique modulo $\mathbb{Z}$. That is, this dot product is not a unique real number, but is rather an equivalence class of all integer translations of that number. We define the DFT over $\mathcal{X}$ similarly. Taking the tensor product of these two DFT's yields the DFT over $\mathcal{X} \oplus \mathcal{Y}$, namely the operator:

$$(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}} W)(x, y) = \sum_{x' \in \mathcal{X}} \sum_{y' \in \mathcal{Y}} W(x', y') \mathrm{e}^{-2\pi \mathrm{i}[(x \cdot x') + (y \cdot y')]}. \tag{49}$$

These facts in hand, we present the main result of this chapter.

Theorem 10. *Let Θ be any (possibly nonlinear) transformation from $\mathcal{Y}^{\mathcal{X}}$ into $\mathbb{R}^{\mathcal{X}\oplus\mathcal{Y}}$. Then Θ is of the form $\Theta(f) = 1_f * W$ for some $W \in \mathbb{R}^{\mathcal{X}\oplus\mathcal{Y}}$ if and only if Θ satisfies the following three properties:*

(i) Θ commutes with translation on $\mathcal{X}$: $\Theta(\mathrm{T}^x f) = \mathrm{T}^{(x,0)}\Theta(f)$,

(ii) Θ commutes with translation on $\mathcal{Y}$: $\Theta(f + y\mathbf{1}) = \mathrm{T}^{(0,y)}\Theta(f)$,

(iii) on average, Θ distributes over all flat occlusion models Φ:

$$\mathrm{E}_\Phi[\Theta(\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})] = \sum_{n=0}^{N-1} \lambda_n \Theta(f_n). \tag{50}$$

Moreover, W is not unique: we have $1_f * W_1 = 1_f * W_2$ for all $f \in \mathcal{Y}^{\mathcal{X}}$ precisely when $(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}} W_1)(x, y) = (\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}} W_2)(x, y)$ whenever $y \neq 0$ or $(x, y) = (0, 0)$.

Proof. We begin by noting that Θ is a generalization of the local histogram transform. Indeed, in Theorem 1(b) we showed that the local histogram transform (1) can be

computed as $\mathrm{LH}_w f = 1_f * (\tilde{w} \otimes \delta_0)$ for any $w \in \mathbb{R}^{\mathcal{X}}$. We generalize this by letting $\Theta(f) = 1_f * W$ where W is any function from $\mathcal{X} \oplus \mathcal{Y}$ into $\mathbb{R}$; here, we note that when $W = \tilde{w} \otimes \delta_0$ this transformation is equal to the local histogram transform. ($\Rightarrow$) Before showing that (i), (ii), and (iii) hold, we begin by noting that they are generalizations of Proposition 2(b) and (c) and Theorem 4, respectively; although the proofs of these facts are straightforward generalizations of the proofs of Proposition 2(b) and (c) and Theorem 4, we present them here for the sake of completeness. To show (i), we note that $1_{\mathrm{T}^x f} = \mathrm{T}^{(x,0)} 1_f$, implying:

$$\Theta(\mathrm{T}^x f) = 1_{\mathrm{T}^x f} * W = (\mathrm{T}^{(x,0)} 1_f) * W = \mathrm{T}^{(x,0)}(1_f * W) = \mathrm{T}^{(x,0)}\Theta(f).$$

Similarly, the fact that $\mathrm{T}^{(0,y)} 1_f = 1_{f+y\mathbf{1}}$ implies (ii):

$$\mathrm{T}^{(0,y)}\Theta(f) = \mathrm{T}^{(0,y)}(1_f * W) = (\mathrm{T}^{(0,y)} 1_f) * W = 1_{f+y\mathbf{1}} * W = \Theta(f + y\mathbf{1}).$$

For (iii), let $(x, y) \in \mathcal{X} \oplus \mathcal{Y}$ and let Φ be any flat occlusion model. We have:

$$\begin{aligned}
\mathrm{E}_\Phi[\Theta(\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})](x,y) &= \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)[\Theta(\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})](x,y) \\
&= \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)(1_{\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}} * W)(x,y) \\
&= \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) \sum_{(x',y') \in \mathcal{X} \oplus \mathcal{Y}} 1_{\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}}(x',y') W(x-x', y-y').
\end{aligned}$$

Noting that (13) gives $\sum_{n=0}^{N-1} 1_\varphi(x',n)1_{f_n}(x',y') = 1_{\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}}(x',y')$, this becomes:

$$\begin{aligned}
&\mathrm{E}_\Phi[\Theta(\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})](x,y) \\
&\quad= \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) \sum_{(x',y')\in\mathcal{X}\oplus\mathcal{Y}} \sum_{n=0}^{N-1} 1_\varphi(x',n)1_{f_n}(x',y')W(x-x',y-y') \\
&\quad= \sum_{n=0}^{N-1} \sum_{(x',y')\in\mathcal{X}\oplus\mathcal{Y}} \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)1_\varphi(x',n)1_{f_n}(x',y')W(x-x',y-y').
\end{aligned}$$

Since Φ is flat, (16) simplifies this to:

$$\begin{aligned}
\mathrm{E}_\Phi[\Theta(\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})](x,y) &= \sum_{n=0}^{N-1} \lambda_n \sum_{(x',y')\in\mathcal{X}\oplus\mathcal{Y}} 1_{f_n}(x',y')W(x-x',y-y') \\
&= \sum_{n=0}^{N-1} \lambda_n (1_{f_n} * W)(x,y) \\
&= \sum_{n=0}^{N-1} \lambda_n \Theta(f_n)(x,y).
\end{aligned}$$

($\Leftarrow$) This direction of the proof is substantially more involved than its converse. We first show how (iii) implies that $\Theta(f) - \Theta(0)$ distributes additively over the standard coordinate vectors:

$$\Theta(f) - \Theta(0) = \sum_{x\in\mathcal{X}} [\Theta(f(x)\delta_x) - \Theta(0)]. \tag{51}$$

Here, for any $y \in \mathcal{Y}$, $y\delta_x$ denotes the function from $\mathcal{X}$ to $\mathcal{Y}$ that has value y at x and is otherwise zero. To prove that (51) holds, we first show that:

$$\Theta(f) - \Theta((1-\delta_x)f) = \Theta(f(x)\delta_x) - \Theta(0). \tag{52}$$

Note, we assume (50) holds for *all* flat Φ and all $\{f_n\}_{n=0}^{N-1}$ in $\mathcal{Y}^{\mathcal{X}}$. In particular, (50)

holds for the specific image occlusion model Φ whose probability density function is:

$$\mathrm{P}_\Phi(\varphi) = \begin{cases} \frac{1}{|\mathcal{X}|}, & \varphi = \delta_x \text{ for some } x \in \mathcal{X}, \\ 0, & \text{else.} \end{cases} \tag{53}$$

We claim this Φ is flat with $\lambda_0 = 1 - \frac{1}{|\mathcal{X}|}$, $\lambda_1 = \frac{1}{|\mathcal{X}|}$, and $\lambda_n = 0$ for all $n \geq 2$. To see this, let $\Delta = \{\delta_{x'} : x' \in \mathcal{X}\}$, and note:

$$\lambda_n = \sum_{\substack{\varphi \in \mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x) = n}} \mathrm{P}_\Phi(\varphi) = \sum_{\substack{\varphi \in \Delta \\ \varphi(x) = n}} \mathrm{P}_\Phi(\varphi) + \sum_{\substack{\varphi \notin \Delta \\ \varphi(x) = n}} \mathrm{P}_\Phi(\varphi) = \sum_{\substack{x' \in \mathcal{X} \\ \delta_{x'}(x) = n}} \mathrm{P}_\Phi(\delta_{x'}).$$

The λ_n's may then be computed as:

$$\begin{aligned} \lambda_n &= \begin{cases} \displaystyle\sum_{\substack{x' \in \mathcal{X} \\ x' \neq x}} \mathrm{P}_\Phi(\delta_{x'}), & n = 0, \\ \mathrm{P}_\Phi(\delta_x), & n = 1, \\ 0, & n > 1, \end{cases} \\ &= \begin{cases} 1 - \frac{1}{|\mathcal{X}|}, & n = 0, \\ \frac{1}{|\mathcal{X}|}, & n = 1, \\ 0, & n > 1. \end{cases} \end{aligned}$$

Having that the Φ defined in (53) is flat, we apply our assumption (50) to it in a special case. Specifically, fixing any f in $\mathcal{Y}^{\mathcal{X}}$ and any $x \in \mathcal{X}$, $y \in \mathcal{Y}$, letting $f_0 = f$ and $f_1 = y\delta_x$, the right-hand-side of (50) is:

$$\sum_{n=0}^{N-1} \lambda_n \Theta(f_n) = \left(1 - \tfrac{1}{|\mathcal{X}|}\right)\Theta(f) + \tfrac{1}{|\mathcal{X}|}\Theta(y\delta_x). \tag{54}$$

Meanwhile, it follows from (53) that the left-hand-side of (50) is:

$$\begin{aligned}\mathrm{E}_\Phi[\Theta(\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})] &= \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)\Theta(\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}) \\ &= \sum_{\varphi\in\Delta} \mathrm{P}_\Phi(\varphi)\Theta(\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}) + \sum_{\varphi\notin\Delta} \mathrm{P}_\Phi(\varphi)\Theta(\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1}) \\ &= \tfrac{1}{|\mathcal{X}|}\sum_{x'\in\mathcal{X}} \Theta(\mathrm{occ}_{\delta_{x'}}\{f_n\}_{n=0}^{N-1}).\end{aligned}$$

Using the easily verified fact that $\mathrm{occ}_{\delta_{x'}}\{f_n\}_{n=0}^{N-1} = (1-\delta_{x'})f_0 + \delta_{x'}f_1$ gives:

$$\begin{aligned}\mathrm{E}_\Phi[\Theta(\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})] &= \tfrac{1}{|\mathcal{X}|}\sum_{x'\in\mathcal{X}} \Theta((1-\delta_{x'})f + \delta_{x'}(y\delta_x)) \\ &= \tfrac{1}{|\mathcal{X}|}\Bigg[\Theta((1-\delta_x)f + y\delta_x) + \sum_{x'\neq x} \Theta((1-\delta_{x'})f)\Bigg]. \qquad (55)\end{aligned}$$

Substituting (54) and (55) into (50), and multiplying by $|\mathcal{X}|$ gives:

$$\Theta((1-\delta_x)f + y\delta_x) + \sum_{x'\neq x} \Theta((1-\delta_{x'})f) = (|\mathcal{X}|-1)\Theta(f) + \Theta(y\delta_x). \qquad (56)$$

Rearranging (56) so that all "y" terms lie on the left-hand-side yields:

$$\Theta((1-\delta_x)f + y\delta_x) - \Theta(y\delta_x) = (|\mathcal{X}|-1)\Theta(f) - \sum_{x'\neq x} \Theta((1-\delta_{x'})f). \qquad (57)$$

As the right-hand-side of (57) is obviously constant with respect to y, then the left-hand-side is implicitly so. In particular, the left-hand-side of (57) is equal to itself at $y = 0$:

$$\Theta((1-\delta_x)f + y\delta_x) - \Theta(y\delta_x) = \Theta((1-\delta_x)f) - \Theta(0). \qquad (58)$$

Now, letting $y = f(x)$ in (58) gives:

$$\Theta(f) - \Theta(f(x)\delta_x) = \Theta((1-\delta_x)f) - \Theta(0),$$

which rearranged, gives our claim (52). Now, using (52) to prove (51), we enumerate $\mathcal{X}$ as $\mathcal{X} = \{x_k\}_{k=1}^{|\mathcal{X}|}$, and write the left-hand-side of (51) as a telescoping sum:

$$\begin{aligned}\Theta(f) - \Theta(0) &= \sum_{j=1}^{|\mathcal{X}|}\left[\Theta\left(f\prod_{k=1}^{j-1}(1-\delta_{x_k})\right) - \Theta\left(f\prod_{k=1}^{j}(1-\delta_{x_k})\right)\right] \\ &= \sum_{j=1}^{|\mathcal{X}|}\left[\Theta\left(f\prod_{k=1}^{j-1}(1-\delta_{x_k})\right) - \Theta\left((1-\delta_{x_j})f\prod_{k=1}^{j-1}(1-\delta_{x_k})\right)\right]. \qquad (59)\end{aligned}$$

Applying (52) to (59) where "f" is $f\prod_{k=1}^{j-1}(1-\delta_{x_k})$ and "x" is x_j then yields (51):

$$\begin{aligned}\Theta(f) - \Theta(0) &= \sum_{j=1}^{|\mathcal{X}|}\left[\Theta\left(\left(f\prod_{k=1}^{j-1}(1-\delta_{x_k})\right)(x_j)\delta_{x_j}\right) - \Theta(0)\right] \\ &= \sum_{j=1}^{|\mathcal{X}|}[\Theta(f(x_j)\delta_{x_j}) - \Theta(0)] \\ &= \sum_{x\in\mathcal{X}}[\Theta(f(x)\delta_x) - \Theta(0)].\end{aligned}$$

Having (51), we now use it, along with our assumptions (i) and (ii), to show that there exists $W \in \mathbb{R}^{\mathcal{X}\oplus\mathcal{Y}}$ such that $\Theta(f) = 1_f * W$ for all $f \in \mathcal{Y}^{\mathcal{X}}$. Indeed, we have:

$$\begin{aligned}\Theta(f) - \Theta(0) &= \sum_{x\in\mathcal{X}}[\Theta(f(x)\delta_x) - \Theta(0)] \\ &= \sum_{x\in\mathcal{X}}[\Theta(\mathrm{T}^x f(x)\delta_0) - \Theta(\mathrm{T}^x 0)] \\ &= \sum_{x\in\mathcal{X}}\mathrm{T}^{(x,0)}[\Theta(f(x)\delta_0) - \Theta(0)].\end{aligned}$$

For any fixed f, writing our sum over $\mathcal{X}$ as a sum over $\mathcal{Y}$ and $\mathcal{X}$ subject to the

constraint that $f(x) = y$ gives:

$$\begin{aligned}\Theta(f) - \Theta(0) &= \sum_{y\in\mathcal{Y}} \sum_{\substack{x\in\mathcal{X}\\ f(x)=y}} \mathrm{T}^{(x,0)}[\Theta(y\delta_0) - \Theta(0)] \\ &= \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} 1_f(x,y)\mathrm{T}^{(x,0)}[\Theta(y\delta_0) - \Theta(0)]. \end{aligned} \tag{60}$$

We now claim that there exists a $W \in \mathbb{R}^{\mathcal{X}\oplus\mathcal{Y}}$ such that:

$$\Theta(y\delta_0) - \Theta(0) = \big(\mathrm{T}^{(0,y)} - \mathrm{I}\big)W, \qquad \forall\, y \in \mathcal{Y}. \tag{61}$$

To be precise, we define W in the frequency domain as:

$$(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W)(x',y') = \begin{cases} \dfrac{\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)]\big\}(x',y')}{\mathrm{e}^{-2\pi\mathrm{i}(y\cdot y')} - 1}, & y' \neq 0, \\ \displaystyle\sum_{y\in\mathcal{Y}} [\Theta(0)](0,y), & (x',y') = (0,0), \\ 0, & x' \neq 0, y' = 0, \end{cases} \tag{62}$$

where, for any $y' \neq 0$, y is chosen such that $y\cdot y' \notin \mathbb{Z}$, where this dot product is defined in (48) according to the factorization (47) of $\mathcal{Y}$ as a direct sum of cyclic groups. To see that W is well-defined, we first note that for any $y' \neq 0$ there always exists at least one such y. Indeed, since $y' \neq 0$, there exists at least one index k_0 such that $y'(k_0) \neq 0 \in \mathbb{Z}_{M_{k_0}}$. Picking y such that $y(k)$ is 1 when $k = k_0$ and is otherwise zero, we have:

$$y \cdot y' = \sum_{k=1}^{K} \frac{y(k)y'(k)}{M_k} = \frac{y'(k_0)}{M_{k_0}} \notin \mathbb{Z}.$$

To show that W is well-defined, we must further show that for $y' \neq 0$, the value of $(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W)(x',y')$ does not depend on the particular y one chooses such that $y\cdot y' \notin \mathbb{Z}$.

To do this, take any $y, \tilde{y}, y' \in \mathcal{Y}$ such that $y \cdot y', \tilde{y} \cdot y' \notin \mathbb{Z}$. We need to show that:

$$\frac{\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)]\big\}(x', y')}{\mathrm{e}^{-2\pi\mathrm{i}(y\cdot y')} - 1} = \frac{\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(\tilde{y}\delta_0) - \Theta(0)]\big\}(x', y')}{\mathrm{e}^{-2\pi\mathrm{i}(\tilde{y}\cdot y')} - 1}. \tag{63}$$

To prove that (63) holds, we first show that:

$$\Theta(y\delta_0 + \tilde{y}\mathbf{1}) - \Theta(\tilde{y}\mathbf{1}) - \Theta(y\delta_0) = \Theta(\tilde{y}\delta_0 + y\mathbf{1}) - \Theta(y\mathbf{1}) - \Theta(\tilde{y}\delta_0). \tag{64}$$

To prove (64) holds, note (51) gives:

$$\begin{aligned}\Theta(y\delta_0 + \tilde{y}\mathbf{1}) - \Theta(\tilde{y}\mathbf{1}) &= [\Theta(y\delta_0 + \tilde{y}\mathbf{1}) - \Theta(0)] - [\Theta(\tilde{y}\mathbf{1}) - \Theta(0)] \\ &= \sum_{x\in\mathcal{X}}[\Theta((y\delta_0 + \tilde{y}\mathbf{1})(x)\delta_x) - \Theta(0)] - \sum_{x\in\mathcal{X}}[\Theta(\tilde{y}\delta_x) - \Theta(0)].\end{aligned}$$

Noting that $(y\delta_0 + \tilde{y}\mathbf{1})(x)$ equals $y + \tilde{y}$ when x is zero and is $\tilde{y}$ otherwise, this becomes:

$$\begin{aligned}&\Theta(y\delta_0 + \tilde{y}\mathbf{1}) - \Theta(\tilde{y}\mathbf{1}) \\ &\quad= [\Theta((y + \tilde{y})\delta_0) - \Theta(0)] + \sum_{x\neq 0}[\Theta(\tilde{y}\delta_x) - \Theta(0)] - \sum_{x\in\mathcal{X}}[\Theta(\tilde{y}\delta_x) - \Theta(0)] \\ &\quad= \Theta((y + \tilde{y})\delta_0) - \Theta(0) - [\Theta(\tilde{y}\delta_0) - \Theta(0)] \\ &\quad= \Theta((y + \tilde{y})\delta_0) - \Theta(\tilde{y}\delta_0).\end{aligned} \tag{65}$$

Interchanging the roles of y and $\tilde{y}$ in (65) gives:

$$\Theta(\tilde{y}\delta_0 + y\mathbf{1}) - \Theta(y\mathbf{1}) = \Theta((\tilde{y} + y)\delta_0) - \Theta(y\delta_0). \tag{66}$$

Solving for $\Theta((\tilde{y} + y)\delta_0)$ in (65) and (66) gives:

$$\Theta(y\delta_0 + \tilde{y}\mathbf{1}) - \Theta(\tilde{y}\mathbf{1}) + \Theta(\tilde{y}\delta_0) = \Theta(\tilde{y}\delta_0 + y\mathbf{1}) - \Theta(y\mathbf{1}) + \Theta(y\delta_0).$$

Subtracting $\Theta(\tilde{y}\delta_0) + \Theta(y\delta_0)$ from this yields our claim (64). Next, by adding $\Theta(0)$ to both sides of (64), and using the assumption (ii) that Θ commutes with translation on $\mathcal{Y}$, we can write (64) as:

$$\begin{aligned}&\mathrm{T}^{(0,\tilde{y})}\Theta(y\delta_0) - \mathrm{T}^{(0,\tilde{y})}\Theta(0) - \Theta(y\delta_0) + \Theta(0)\\ &\qquad\qquad = \mathrm{T}^{(0,y)}\Theta(\tilde{y}\delta_0) - \mathrm{T}^{(0,y)}\Theta(0) - \Theta(\tilde{y}\delta_0) + \Theta(0),\end{aligned}$$

or more concisely, as:

$$\big(\mathrm{T}^{(0,\tilde{y})} - \mathrm{I}\big)[\Theta(y\delta_0) - \Theta(0)] = \big(\mathrm{T}^{(0,y)} - \mathrm{I}\big)[\Theta(\tilde{y}\delta_0) - \Theta(0)].$$

Taking Fourier transforms then gives:

$$\begin{aligned}&(\mathrm{e}^{-2\pi\mathrm{i}(\tilde{y}\cdot y')} - 1)\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)]\big\}(x', y')\\ &\qquad\qquad = (\mathrm{e}^{-2\pi\mathrm{i}(y\cdot y')} - 1)\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(\tilde{y}\delta_0) - \Theta(0)]\big\}(x', y'),\end{aligned}$$

for any $(x', y') \in \mathcal{X} \oplus \mathcal{Y}$. Since $y \cdot y' \notin \mathbb{Z}$ and $y \cdot \tilde{y} \notin \mathbb{Z}$, dividing gives (63). This concludes the argument showing that W is well-defined by (62).

We now note that W—implicitly defined in (62)—is real-valued, as claimed in the statement of the result. Since $W \in \mathbb{C}^{\mathcal{X}\oplus\mathcal{Y}}$ is real-valued precisely when its Fourier transform is skew-symmetric about the origin, it suffices to show that:

$$\big(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W\big)(-x', -y') = [(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W)(x', y')]^*, \qquad \forall\ (x', y') \in \mathcal{X} \oplus \mathcal{Y}. \tag{67}$$

We recall our assumption that Θ is any transformation from $\mathcal{Y}^{\mathcal{X}}$ into $\mathbb{R}^{\mathcal{X}\oplus\mathcal{Y}}$. As such, $\Theta(0)$ is real-valued, and so (62) gives $(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W)(0,0)$ is real, meaning (67) holds at $(x', y') = (0, 0)$. Moreover, for $y' = 0$, (62) gives both sides of (67) to be 0. Meanwhile,

for $y' \neq 0$, since $\Theta(y\delta_0) - \Theta(0)$ is real-valued, its Fourier transform satisfies:

$$\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta) - \Theta(0)]\big\}(-x', -y') = \big(\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta) - \Theta(0)]\big\}(x', y')\big)^*.$$

Taking any y such that $y \cdot y' \notin \mathbb{Z}$, note we also have $y \cdot (-y') \notin \mathbb{Z}$ and so this fact, along with (67), gives:

$$\begin{aligned}(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W)(-x', -y') &= \frac{\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta) - \Theta(0)]\big\}(-x', -y')}{\mathrm{e}^{-2\pi\mathrm{i}(y\cdot(-y'))} - 1} \\ &= \frac{\big(\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta) - \Theta(0)]\big\}(x', y')\big)^*}{(\mathrm{e}^{-2\pi\mathrm{i}(y\cdot y')} - 1)^*} \\ &= [(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W)(x', y')]^*.\end{aligned}$$

Having that the W defined (62) is real-valued, we next claim that W indeed satisfies (61). In the frequency domain, (61) is equivalent to having:

$$\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)]\big\}(x', y') = (\mathrm{e}^{-2\pi\mathrm{i}(y\cdot y')} - 1)(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W)(x', y'), \tag{68}$$

for all $x' \in \mathcal{X}$, $y, y' \in \mathcal{Y}$. For $y' \in \mathcal{Y}$ such that $y \cdot y' \notin \mathbb{Z}$, note that (62) immediately gives (68). Meanwhile, for $y' \in \mathcal{Y}$ such that $y \cdot y' \in \mathbb{Z}$, the right-hand-side of (68) is necessarily 0, meaning we must show that:

$$\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)]\big\}(x', y') = 0, \quad \forall\ (x', y)' \in \mathcal{X} \oplus \mathcal{Y} \text{ s.t. } y \cdot y' \in \mathbb{Z}. \tag{69}$$

To do this, note that for any $y \cdot y' \in \mathbb{Z}$, we have:

$$\begin{aligned}\big\{\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)]\big\}(x', y') &= \sum_{\tilde{x}\in\mathcal{X}}\sum_{\tilde{y}\in\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)](\tilde{x}, \tilde{y})\mathrm{e}^{-2\pi\mathrm{i}(\tilde{x}\cdot x')}\mathrm{e}^{-2\pi\mathrm{i}(\tilde{y}\cdot y')} \\ &= \sum_{\tilde{x}\in\mathcal{X}}\mathrm{e}^{-2\pi\mathrm{i}(\tilde{x}\cdot x')}\sum_{\tilde{y}\in\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)](\tilde{x}, \tilde{y}).\end{aligned}$$

It thus suffices to prove that for any fixed $\tilde{x} \in \mathcal{X}$,

$$\sum_{\tilde{y}\in\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)](\tilde{x},\tilde{y}) = 0. \tag{70}$$

Note that a change of variables, along with our assumption (ii), gives:

$$\begin{aligned}\sum_{\tilde{y}\in\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)](\tilde{x},\tilde{y}) &= \sum_{(\tilde{y}-y')\in\mathcal{Y}}[\Theta(y\delta_0) - \Theta(0)](\tilde{x},\tilde{y}-y') \\ &= \sum_{y'\in\mathcal{Y}}\left\{\mathrm{T}^{(0,y')}[\Theta(y\delta_0) - \Theta(0)]\right\}(\tilde{x},\tilde{y}) \\ &= \sum_{y'\in\mathcal{Y}}[\Theta(y\delta_0 + y'\mathbf{1}) - \Theta(y'\mathbf{1})](\tilde{x},\tilde{y}). \end{aligned} \tag{71}$$

We now recall (65) which gives the telescoping sum:

$$\sum_{y'\in\mathcal{Y}}[\Theta(y\delta_0 + y'\mathbf{1}) - \Theta(y'\mathbf{1})] = \sum_{y'\in\mathcal{Y}}[\Theta((y+y')\delta_0) - \Theta(y'\delta_0)] = 0. \tag{72}$$

Substituting (72) into (71) gives (70), which in turn implies (69). Having that (68) holds in all cases, we obtain (61).

Having the claim (61), we substitute it into (60):

$$\begin{aligned}\Theta(f) - \Theta(0) &= \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} 1_f(x,y)\mathrm{T}^{(x,0)}(\mathrm{T}^{(0,y)} - \mathrm{I})W \\ &= \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} 1_f(x,y)(\mathrm{T}^{(x,y)}W - \mathrm{T}^{(x,0)}W). \end{aligned} \tag{73}$$

Since for any fixed $x \in \mathcal{X}$, we have that $\sum_{y\in\mathcal{Y}} 1_f(x,y) = 1$, evaluating (73) at some

$(x', y') \in \mathcal{X} \oplus \mathcal{Y}$ gives:

$$\begin{aligned}
[\Theta(f) - \Theta(0)](x', y') &= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} 1_f(x, y)[(\mathrm{T}^{(x,y)} W)(x', y') - (\mathrm{T}^{(x,0)} W)(x', y')] \\
&= (1_f * W)(x', y') - \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} 1_f(x, y) W(x' - x, y') \\
&= (1_f * W)(x', y') - \sum_{x \in \mathcal{X}} W(x' - x, y') \\
&= (1_f * W)(x', y') - \sum_{x \in \mathcal{X}} (\mathrm{T}^{(x,0)} W)(x', y'). \qquad (74)
\end{aligned}$$

We now claim that

$$\Theta(0) = \sum_{x \in \mathcal{X}} \mathrm{T}^{(x,0)} W. \qquad (75)$$

To see this, note that both $(\sum_{x \in \mathcal{X}} \mathrm{T}^{(x,0)} W)(x', y')$ and $[\Theta(0)](x', y')$ are constant with respect to x': for the former, note $(\sum_{x \in \mathcal{X}} \mathrm{T}^{(x,0)} W)(x', y') = \sum_{x \in \mathcal{X}} W(x, y')$; for the latter, note that for any $x \in \mathcal{X}$, assumption (i) gives:

$$[\Theta(0)](x', y') = [\Theta(\mathrm{T}^{x'-x} 0)](x', y') = \left\{\mathrm{T}^{(x'-x,0)} [\Theta(0)]\right\}(x', y') = [\Theta(0)](x, y').$$

We therefore treat both of these functions as functions solely of y'. In particular, in order to prove (75), it suffices to show that it holds in the frequency domain of $\mathcal{Y}$:

$$[\mathrm{F}_{\mathcal{Y}} \Theta(0)](y') = \left[\mathrm{F}_{\mathcal{Y}} \left(\sum_{x \in \mathcal{X}} \mathrm{T}^{(x,0)} W \right)\right](y'), \qquad \forall\, y' \in \mathcal{Y}. \qquad (76)$$

To show that (76) holds for any $y' \neq 0$, pick y such that $y \cdot y' \notin \mathbb{Z}$ and let $f = y\mathbf{1}$ in (74):

$$(\mathrm{T}^{y} - \mathrm{I})\Theta(0) = \Theta(y\mathbf{1}) - \Theta(0) = 1_{y\mathbf{1}} * W - \sum_{x \in \mathcal{X}} \mathrm{T}^{(x,0)} W.$$

Using the easily verified fact that $1_{y\mathbf{1}} * W = \sum_{x\in\mathcal{X}} \mathrm{T}^{(x,y)}W$, this becomes:

$$(\mathrm{T}^y - \mathrm{I})\Theta(0) = \sum_{x\in\mathcal{X}} \mathrm{T}^{(x,y)}W - \sum_{x\in\mathcal{X}} \mathrm{T}^{(x,0)}W = (\mathrm{T}^y - \mathrm{I}) \sum_{x\in\mathcal{X}} \mathrm{T}^{(x,0)}W.$$

Taking Fourier transforms of this over $\mathcal{Y}$ gives that:

$$(\mathrm{e}^{-2\pi \mathrm{i} y\cdot y'} - 1)[\mathrm{F}_{\mathcal{Y}}\Theta(0)](y') = (\mathrm{e}^{-2\pi \mathrm{i} y\cdot y'} - 1)\left[\mathrm{F}_{\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} \mathrm{T}^{(x,0)}W\right)\right](y'). \tag{77}$$

Since $y \cdot y' \notin \mathbb{Z}$, we may divide by $(\mathrm{e}^{-2\pi \mathrm{i} y\cdot y'} - 1)$ to obtain (76) in this case. In the remaining case where $y' = 0$, the right-hand-side of (76) becomes:

$$\left[\mathrm{F}_{\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} \mathrm{T}^{(x,0)}W\right)\right](0) = \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} (\mathrm{T}^{(x,0)}W)(y) = \sum_{(x,y)\in\mathcal{X}\oplus\mathcal{Y}} W(x,y) = (\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W)(0,0).$$

From the definition (62) of W, this becomes:

$$\left[\mathrm{F}_{\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} \mathrm{T}^{(x,0)}W\right)\right](0) = \sum_{y\in\mathcal{Y}} [\Theta(0)](y) = [\mathrm{F}_{\mathcal{Y}}\Theta(0)](0),$$

namely (76) where $y' = 0$. Therefore, (76) is satisfied giving (75). Substituting (75) into (74) gives our result that $\Theta(f) = 1_f * W$.

All that remains to be shown is that our choice of W is not unique. To be precise, the argument up to this point has only made use of the values of (62) where either $y' \neq 0$ or $(x', y') = (0, 0)$. This suggests the values of $(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W)(x', 0)$ where $x' \neq 0$ are, in fact, arbitrary, even though they were chosen to be 0 in (62). This is indeed the case: if $1_f * W_1 = 1_f * W_2$, then $0 = 1_f * (W_1 - W_2)$ for all f. Taking the Fourier transform over $\mathcal{X} \oplus \mathcal{Y}$ gives:

$$0 = (\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}1_f)[\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}(W_1 - W_2)], \tag{78}$$

where:

$$(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}1_f)(x',0) = \sum_{(x,y)\in\mathcal{X}\oplus\mathcal{Y}} 1_f(x,y)\mathrm{e}^{-2\pi\mathrm{i}(x'\cdot x)} = \sum_{x\in\mathcal{X}} \mathrm{e}^{-2\pi\mathrm{i}(x'\cdot x)} \sum_{y\in\mathcal{Y}} 1_f(x,y).$$

Since for any fixed x, $\sum_{y\in\mathcal{Y}} 1_f(x,y) = 1$, we have:

$$(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}1_f)(x',0) = \sum_{x\in\mathcal{X}} \mathrm{e}^{-2\pi\mathrm{i}(x'\cdot x)} = (\mathrm{F}_{\mathcal{X}}1)(x') = |\mathcal{X}|\delta_0(x'),$$

which equals 0 for any $x' \neq 0$. In light of (78), we thus have that $(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W_1)(x,0)$ need not be equal to $(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W_2)(x,0)$ for $x \neq 0$, thus giving our result: $1_f * W_1 = 1_f * W_2$ for all $f \in \mathcal{Y}^{\mathcal{X}}$ precisely when $(\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W_1)(x,y) = (\mathrm{F}_{\mathcal{X}\oplus\mathcal{Y}}W_2)(x,y)$ whenever $y \neq 0$ or $(x,y) = (0,0)$. □

In summary, in this chapter, we showed that local histograms have unique capabilities with regards to the analysis of composite images. To be precise, we let Θ be any (possibly nonlinear) function that transforms images in $\mathcal{Y}^{\mathcal{X}}$ into "joint distributions" of pixel location and value in $\mathbb{R}^{\mathcal{X}\oplus\mathcal{Y}}$. We showed that if such a Θ commutes with translations on $\mathcal{X}$ and $\mathcal{Y}$ and, like local histograms, distributes on average over occlusions (7), then Θ is necessarily of *local histogram type*, that is, is of the form $\Theta(f) = 1_f * W$ for some $W \in \mathbb{R}^{\mathcal{X}\oplus\mathcal{Y}}$. In short, Theorem 10 establishes that such operators—ones that filter the *graph* of an image as opposed to the image itself—are uniquely well-suited to analyze composite images of a certain type.

V. Variance of Local Histogram Transforms

5.1 Variance and Two-Flatness

In the previous two chapters, we discussed how local histograms distribute, on average, over flat occlusion models. Having this understanding of the average behavior of local histograms, in this chapter, we focus on the subject of how closely this average approximates a typical local histogram. That is, we compute the variance of the local histograms of a composite image (Theorem 11), and demonstrate that the distance of our local histograms from the average is highly dependent on the scale of the local histogram window.

To compute this variance, we necessarily make several simplifying assumptions on the nature of the random images in question. In particular, we assume that our occlusion model Φ is *two-flat*, meaning that 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; formally, Φ is *two-flat* if there exists nonnegative scalars $\{\lambda_n\}_{n=0}^{N-1}$ such that for any n', n'' and $x', x'' \in \mathcal{X}$, $x' \neq x''$:

$$\sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x', n') 1_\varphi(x'', n'') = \lambda_{n'} \lambda_{n''}. \tag{79}$$

Note that two-flatness is a stronger form of flatness, that is, if Φ is two-flat (79), then Φ is flat (16). To show this, we begin by summing up (79) over all n'':

$$\lambda_{n'} \sum_{n''=0}^{N-1} \lambda_{n''} = \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x', n') \sum_{n''=0}^{N-1} 1_\varphi(x'', n'') = \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x', n'), \tag{80}$$

where the last equality follows directly from the fact that for any fixed φ, we have

that $\sum_{n''=0}^{N-1} 1_\varphi(x'', n'') = 1$. Now, summing over all n' gives:

$$\left(\sum_{n'=0}^{N-1} \lambda_{n'}\right)^2 = \sum_{n'=0}^{N-1} \lambda_{n'} \sum_{n''=0}^{N-1} \lambda_{n''} = \sum_{\varphi \in \mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) \sum_{n'=0}^{N-1} 1_\varphi(x', n') = 1,$$

which, since the $\lambda_{n'}$'s are nonnegative, gives that $\sum_{n'=0}^{N-1} \lambda_{n'} = 1$. Note that in light of this fact, (80) reduces to (16), namely that Φ is flat.

Having this property of two-flatness, we can 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 11. *Let $\{f_n\}_{n=0}^{N-1}$ be any sequence of images in $\mathcal{Y}^{\mathcal{X}}$ which exhibits unique pixel values, that is, $f_{n'}(x) = f_{n''}(x)$ implies $n' = n''$. Then for any two-flat N-image occlusion model Φ, and any function w in $\mathbb{R}^{\mathcal{X}}$, the variance of the local histogram transform* (1) *of a composite image* (7) *is*

$$\mathrm{V}_\Phi(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})(x, y) = \sum_{n=0}^{N-1} \lambda_n(1 - \lambda_n)(\mathrm{LH}_{w^2} f_n)(x, y). \tag{81}$$

Proof. We assume that Φ is two-flat, which, as noted above, implies that Φ is also flat. As shown in Theorem 4 and generalized in Theorem 10, since Φ is flat, we have:

$$\mathrm{E}_\Phi(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})(x, y) = \sum_{n=0}^{N-1} \lambda_n(\mathrm{LH}_w f_n)(x, y). \tag{82}$$

Therefore, we can compute the variance as:

$$\begin{aligned} \mathrm{V}_\Phi(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1}) &= \mathrm{E}_\Phi[(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})^2] - [\mathrm{E}_\Phi(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})]^2 \\ &= \mathrm{E}_\Phi[(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})^2] - \left[\sum_{n=0}^{N-1} \lambda_n(\mathrm{LH}_w f_n)(x, y)\right]^2. \end{aligned} \tag{83}$$

To compute $\mathrm{E}_\Phi[(\mathrm{LH}_w\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})^2]$, it follows from (1) that:

$$\begin{aligned}
&\mathrm{E}_\Phi[(\mathrm{LH}_w\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})^2](x,y) \\
&\qquad = \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)(\mathrm{LH}_w\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})^2(x,y) \\
&\qquad = \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)\Big[\sum_{x'\in\mathcal{X}} w(x')\delta_y((\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})(x+x'))\Big]^2 \\
&\qquad = \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)\Big[\sum_{x'\in\mathcal{X}} w(x'-x)\delta_y((\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})(x'))\Big]^2. \qquad (84)
\end{aligned}$$

Recalling (13), we have that:

$$\delta_y((\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})(x')) = \sum_{n=0}^{N-1} 1_\varphi(x',n)\delta_y(f_n(x')),$$

and so (84) becomes:

$$\begin{aligned}
&\mathrm{E}_\Phi[(\mathrm{LH}_w\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})^2](x,y) \\
&\qquad = \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)\Big[\sum_{x'\in\mathcal{X}} w(x'-x)\sum_{n'=0}^{N-1} 1_\varphi(x',n')\delta_y(f_{n'}(x'))\Big]^2 \\
&\qquad = \sum_{x'\in\mathcal{X}}\sum_{x''\in\mathcal{X}} w(x'-x)w(x''-x)\sum_{n'=0}^{N-1}\sum_{n''=0}^{N-1}\delta_y(f_{n'}(x'))\delta_y(f_{n''}(x'')) \\
&\qquad\qquad \times \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi)1_\varphi(x',n')1_\varphi(x'',n'').
\end{aligned}$$

We now break up the sums over $x', x'' \in \mathcal{X}$ into a sum where $x' = x''$ and a sum where $x' \neq x''$, and note that for $x' \neq x''$ the sum over $\varphi \in \mathbb{Z}_N^{\mathcal{X}}$ reduces to (79) since

Φ is two-flat:

$$\begin{aligned}
&\mathrm{E}_\Phi[(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})^2](x,y) \\
&\quad = \sum_{x'\in\mathcal{X}}[w(x'-x)]^2 \sum_{n'=0}^{N-1}\sum_{n''=0}^{N-1} \delta_y(f_{n'}(x'))\delta_y(f_{n''}(x')) \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x',n') 1_\varphi(x',n'') \\
&\qquad + \sum_{x'\neq x''} w(x'-x)w(x''-x) \sum_{n'=0}^{N-1}\sum_{n''=0}^{N-1} \delta_y(f_{n'}(x'))\delta_y(f_{n''}(x''))\lambda_{n'}\lambda_{n''}. \qquad (85)
\end{aligned}$$

For the first summand in (85), we have that $1_\varphi(x',n')1_\varphi(x',n'')$ is zero when $n' \neq n''$ and is $1_\varphi(x',n')$ when $n' = n''$. This, along with the fact that Φ is flat, reduces the first term in (85) to:

$$\begin{aligned}
&\sum_{x'\in\mathcal{X}}[w(x'-x)]^2 \sum_{n'=0}^{N-1}\sum_{n''=0}^{N-1} \delta_y(f_{n'}(x'))\delta_y(f_{n''}(x')) \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x',n') 1_\varphi(x',n'') \\
&\qquad = \sum_{x'\in\mathcal{X}}[w(x'-x)]^2 \sum_{n'=0}^{N-1} \delta_y(f_{n'}(x')) \sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x',n') \\
&\qquad = \sum_{n'=0}^{N-1} \lambda_{n'} \sum_{x'\in\mathcal{X}} [w(x'-x)]^2 \delta_y(f_{n'}(x')) \\
&\qquad = \sum_{n'=0}^{N-1} \lambda_{n'} (\mathrm{LH}_{w^2} f_{n'})(x,y). \qquad (86)
\end{aligned}$$

Meanwhile, we write the second term in (85) as the sums over all $x', x'' \in \mathcal{X}$ minus the sum over $x' = x''$:

$$\begin{aligned}
&\sum_{x'\neq x''} w(x'-x)w(x''-x) \sum_{n'=0}^{N-1}\sum_{n''=0}^{N-1} \delta_y(f_{n'}(x'))\delta_y(f_{n''}(x''))\lambda_{n'}\lambda_{n''} \\
&\qquad = \sum_{x'\in\mathcal{X}}\sum_{x''\in\mathcal{X}} w(x'-x)w(x''-x) \sum_{n'=0}^{N-1}\sum_{n''=0}^{N-1} \delta_y(f_{n'}(x'))\delta_y(f_{n''}(x''))\lambda_{n'}\lambda_{n''} \\
&\qquad\quad - \sum_{x'\in\mathcal{X}}[w(x'-x)]^2 \sum_{n'=0}^{N-1}\sum_{n''=0}^{N-1} \delta_y(f_{n'}(x'))\delta_y(f_{n''}(x'))\lambda_{n'}\lambda_{n''}. \qquad (87)
\end{aligned}$$

Here, the last term in (87) is zero unless $f_{n'}(x') = f_{n''}(x')$ which, from our assumptions, implies that $n' = n''$. As such, (87) becomes:

$$\begin{aligned}
&\sum_{x' \neq x''} w(x' - x)w(x'' - x) \sum_{n'=0}^{N-1} \sum_{n''=0}^{N-1} \delta_y(f_{n'}(x'))\delta_y(f_{n''}(x''))\lambda_{n'}\lambda_{n''} \\
&\qquad = \Big[\sum_{n'=0}^{N-1} \lambda_{n'} \sum_{x' \in \mathcal{X}} w(x' - x)\delta_y(f_{n'}(x')) \Big]^2 - \sum_{x' \in \mathcal{X}} [w(x' - x)]^2 \sum_{n'=0}^{N-1} \delta_y(f_{n'}(x'))\lambda_{n'}^2 \\
&\qquad = \Big[\sum_{n'=0}^{N-1} \lambda_{n'} (\mathrm{LH}_w f_{n'})(x, y) \Big]^2 - \sum_{n'=0}^{N-1} \lambda_{n'}^2 (\mathrm{LH}_{w^2} f_{n'})(x, y). \qquad (88)
\end{aligned}$$

Substituting (86) and (88) into (85), and combining like terms gives:

$$\begin{aligned}
&\mathrm{E}_\Phi[(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})^2](x, y) \\
&\qquad = \sum_{n'=0}^{N-1} \lambda_{n'}(1 - \lambda_{n'})(\mathrm{LH}_{w^2} f_{n'})(x, y) + \Big[\sum_{n'=0}^{N-1} \lambda_{n'} (\mathrm{LH}_w f_{n'})(x, y) \Big]^2.
\end{aligned}$$

Substituting this into (83) gives our result (81). □

Theorem 11 shows that the variance of the local histogram transform of a composite image depends on the scale of the support of w, a fact which is demonstrated in Figure 9. To see this more clearly, consider an example where $\mathcal{Y} = \mathbb{Z}_M$ and $\mathcal{X} = \mathbb{Z}_L \oplus \mathbb{Z}_L$. In particular, let the f_n's be distinct, constant functions, that is, $f_n = m_n \mathbf{1}$ where $m_n \in \mathbb{Z}_M$ and $m_{n'} = m_{n''}$ if and only if $n' = n''$. Let $w \in \mathbb{R}^{\mathcal{X}}$ be the square weighting function defined by:

$$w(l_1, l_2) = \begin{cases} \frac{1}{(2J+1)^2}, & l_1, l_2 \in [-J, J], \\ 0, & \text{else}, \end{cases} \qquad (89)$$

where $J < L$. Let ρ_n denote the probability of choosing label $n \in \mathbb{Z}_N$. Noting that

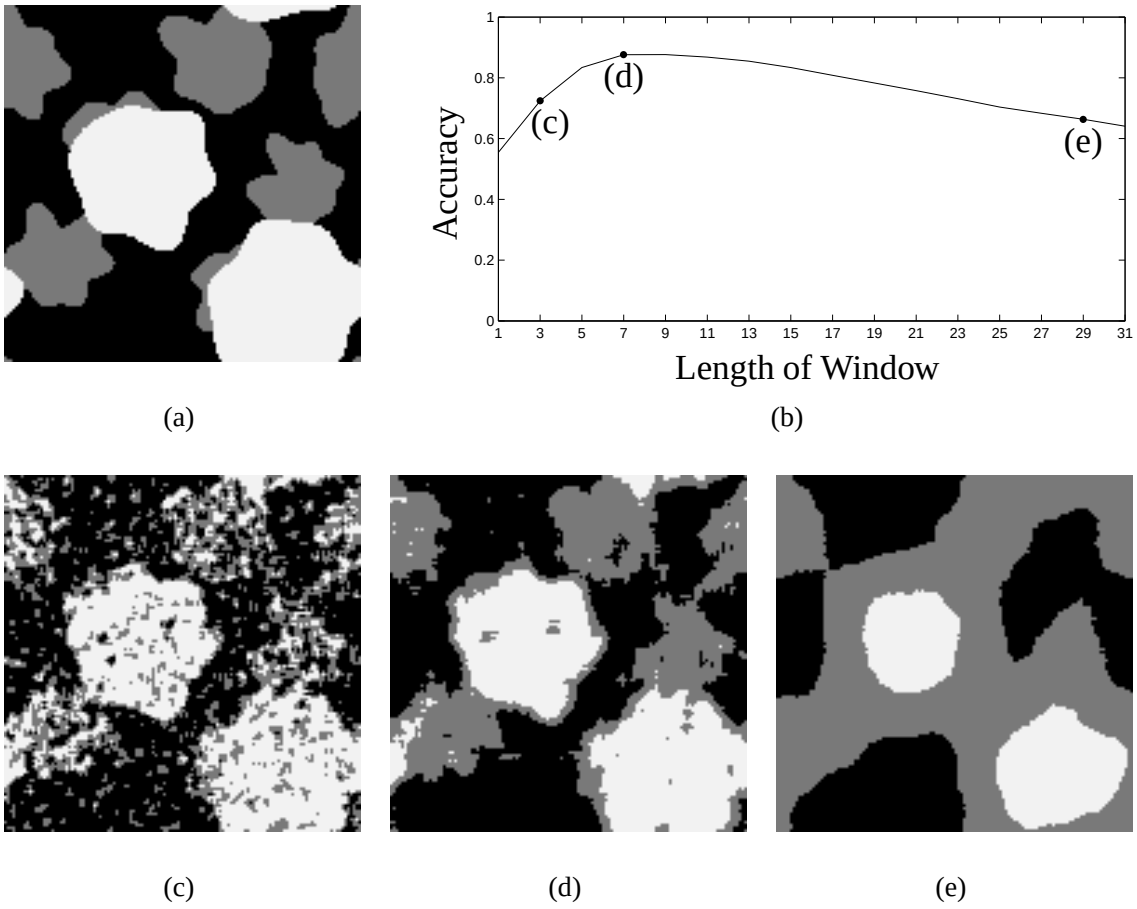


Figure 9. An example demonstrating the importance of the scale of the local histogram window. Recall the example given in Figure 2. The theory presented in this dissertation has provided the theoretical justification for a certain type of segmentation-and-classification algorithm. That is, we assign a label to any given pixel in Figure 2(e) by comparing its local histogram to the three distinct distributions of pixels found in Figure 2(a), (b) and (c). In doing so, we hope to obtain results that compare well against the ground truth image (a). In (b), we see that the support of the window plays an important role in how accurately our algorithm segments-and-labels our image Figure 2(e)—a fact which is justified in Theorem 11. In this example, our window is defined as a box filter (89) of "radius" J. That is, the box has sides of length $2J+1$; (b) gives the accuracy for this classification for windows of length $2J+1 = 1, \ldots, 31$, that is, $J = 0, \ldots, 15$. When our window is too small, we get a "noisy" segmented-and-labeled image, such as the one given in (c) where $2J+1 = 3$. In (b), we see that our classification accuracy is highest when $2J+1 = 7$. Indeed, using this window in our "nearest histogram" classification scheme results in (d) which compares well against ground truth (a). Finally, if we choose our window to be too large, our classification accuracy suffers from oversmoothing as seen in (e) which was obtained by letting $2J+1 = 29$.

$w = \tilde{w}$, it follows from Theorem 1(a) that:

$$(\mathrm{LH}_{w^2} f_n)(x,y) = (w^2 * 1_{f_n^{-1}\{y\}})(x) = \sum_{x' \in \mathcal{X}} w^2(x') 1_{f_n^{-1}\{y\}}(x - x').$$

Noting that our definition of f_n gives that $f_n(x - x') = y$ if and only if $m_n = y$, it follows from (2) that:

$$(\mathrm{LH}_{w^2} f_n)(x,y) = \sum_{x' \in \mathcal{X}} w^2(x') \delta_y(m_n) = \delta_y(m_n) \sum_{x' \in \mathrm{supp}(w)} \frac{1}{(2J+1)^4} = \frac{\delta_y(m_n)}{(2J+1)^2},$$

where the last equality follows from the fact that $|\mathrm{supp}(w)| = (2J+1)^2$. This then gives that (81) becomes:

$$\mathrm{V}_\Phi(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})(x,y) = \frac{1}{(2J+1)^2} \sum_{n=0}^{N-1} \rho_n(1-\rho_n)\delta_y(m_n).$$

Here, we indeed see that as the support of the window increases, our variance becomes smaller.

We note, however, that this is under the assumption that the window lies completely within a region for which the two-flat hypothesis holds. Indeed, as the scale of the window grows large, we expect this result to not hold when applied to images that are composites of several distinct textures, each arising from its own unique image model, such as Figure 2(e). In particular, when the scale of the window is large, the local histogram at a pixel will likely mix in the histograms of the neighboring textures. Therefore, the local histogram at that particular pixel will not compare well against the histogram of any one given texture, causing the accuracy of our "nearest histogram" classification scheme to decrease. This intuition is confirmed by the decreased performance of our classification algorithm as the scale of the window grows large, see Figure 9(b) and (e). Having seen that two-flatness is important in under-

standing the variance of the local histograms of a composite image, we, in the next section, consider if two-flatness is a realistic assumption in terms of our motivating application.

5.2 Constructing Two-Flat Occlusion Models

In this section, we demonstrate that two-flatness is a reasonable assumption by providing a couple methods for constructing two-flat occlusion models. Recall that in a Bernoulli model, we define the probability of picking any particular $\varphi \in \mathbb{Z}_2^{\mathcal{X}}$ as $\mathrm{P}_\Phi(\varphi) = \prod_{x\in\mathcal{X}} \rho_{\varphi(x)}$. Not only is the Bernoulli random variable Φ flat as shown in Theorem 5, but it is also two-flat:

Theorem 12. *If Φ is a Bernoulli random variable, then Φ is two-flat.*

Proof. To show this, first fix some $x_{k_0'}, x_{k_0''} \in \mathcal{X}$, $x_{k_0'} \neq x_{k_0''}$. Paralleling the proof of Φ being flat (Theorem 5), it follows that:

$$\sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x_{k_0'}, n') 1_\varphi(x_{k_0''}, n'') = \sum_{n_1=0}^{N-1} \rho_{n_1} \cdots \sum_{n_K=0}^{N-1} \rho_{n_K} \sum_{\substack{\varphi\in\mathbb{Z}_N^{\mathcal{X}} \\ \varphi(x_k)=n_k,\forall k}} 1_\varphi(x_{k_0'}, n') 1_\varphi(x_{k_0''}, n'').$$

As the constraints on the sum over $\varphi \in \mathbb{Z}_N^{\mathcal{X}}$ uniquely define a single φ, we have that the sum is one if $n' = n_{k_0'}$ and $n'' = n_{k_0''}$ thus giving:

$$\begin{aligned}
&\sum_{\varphi\in\mathbb{Z}_N^{\mathcal{X}}} \mathrm{P}_\Phi(\varphi) 1_\varphi(x_{k_0'}, n') 1_\varphi(x_{k_0''}, n'') \\
&\qquad = \sum_{n_1=0}^{N-1} \rho_{n_1} \cdots \sum_{n_K=0}^{N-1} \rho_{n_K} \delta_{n'}(n_{k_0'}) \delta_{n''}(n_{k_0''}) \\
&\qquad = \Bigg(\sum_{n_{k_0'}=0}^{N-1} \rho_{n_{k_0'}} \delta_{n'}(n_{k_0'}) \Bigg) \Bigg(\sum_{n_{k_0''}=0}^{N-1} \rho_{n_{k_0''}} \delta_{n''}(n_{k_0''}) \Bigg) \prod_{k\neq k_0', k_0''} \Bigg(\sum_{n_k=0}^{N-1} \rho_{n_k} \Bigg) \\
&\qquad = \rho_{n'} \rho_{n''}.
\end{aligned}$$

□

Having shown that Bernoulli random variables are two-flat, we now consider another method for constructing two-flat occlusion models which uses the overlay operator (31). In Theorem 8, we showed that if Φ, Ψ, and Σ were flat, then $\Phi \#_\Sigma \Psi$ is also flat. Here, we show a similar result holds when flatness is replaced with two-flatness:

Theorem 13. *If Φ, Ψ, and Σ are two-flat* (79)*, then $\Phi \#_\Sigma \Psi$ is two-flat.*

Proof. We begin by noting that since Φ, Ψ, and Σ are two-flat, they are also flat. We want to show that $\Phi \#_\Sigma \Psi$ is two-flat (79), meaning that for any $n', n'' \in \mathbb{Z}_{N_\Phi + N_\Psi}$ and $x', x'' \in \mathcal{X}$ such that $x' \neq x''$, our goal is to show that there exists scalars $\{\lambda_{\Phi \#_\Sigma \Psi, n}\}_{n=0}^{N_\Phi + N_\Psi - 1}$ such that:

$$\sum_{\substack{\upsilon \in \mathbb{Z}_{N_\Phi + N_\Psi}^{\mathcal{X}} \\ \upsilon(x') = n' \\ \upsilon(x'') = n''}} \mathrm{P}_{\Phi \#_\Sigma \Psi}(\upsilon) = \lambda_{\Phi \#_\Sigma \Psi, n'} \lambda_{\Phi \#_\Sigma \Psi, n''}. \tag{90}$$

To see this, note that for any $n', n'' = 0, \cdots, N_\Phi + N_\Psi - 1$ and $x', x'' \in \mathcal{X}$ such that $x' \neq x''$, it follows from the ideas used in (35) that:

$$\sum_{\substack{\upsilon \in \mathbb{Z}_{N_\Phi + N_\Psi}^{\mathcal{X}} \\ \upsilon(x') = n' \\ \upsilon(x'') = n''}} \mathrm{P}_{\Phi \#_\Sigma \Psi}(\upsilon) = \sum_{\substack{\varphi \in \mathbb{Z}_{N_\Phi}^{\mathcal{X}} \\ \psi \in \mathbb{Z}_{N_\Psi}^{\mathcal{X}} \\ \sigma \in \mathbb{Z}_2^{\mathcal{X}} \\ (\varphi \#_\sigma \psi)(x') = n' \\ (\varphi \#_\sigma \psi)(x'') = n''}} \mathrm{P}_\Phi(\varphi) \mathrm{P}_\Psi(\psi) \mathrm{P}_\Sigma(\sigma). \tag{91}$$

Note that if $n = 0, \ldots, N_\Phi - 1$, (31) gives that $(\varphi \#_\sigma \psi)(x) = n$ if and only if $\varphi(x) = n$ and $\sigma(x) = 0$. Meanwhile, if $n = N_\Phi, \ldots, N_\Phi + N_\Psi - 1$, (31) gives that $(\varphi \#_\sigma \psi)(x) = n$ if and only if $\psi(x) = n - N_\Phi$ and $\sigma(x) = 1$. We now use these facts to compute (91) for any $n', n'' \in \mathbb{Z}_{N_\Phi + N_\Psi}$. In the case where $n', n'' = 0, \ldots, N_\Phi - 1$, since Φ and Σ are

two-flat, (91) becomes:

$$\sum_{\substack{\upsilon\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi+N_\Psi}\\ \upsilon(x')=n'\\ \upsilon(x'')=n''}} \mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon) = \sum_{\substack{\varphi\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi}\\ \varphi(x')=n'\\ \varphi(x'')=n''}} \mathrm{P}_\Phi(\varphi) \sum_{\psi\in\mathbb{Z}^{\mathcal{X}}_{N_\Psi}} \mathrm{P}_\Psi(\psi) \sum_{\substack{\sigma\in\mathbb{Z}_2^{\mathcal{X}}\\ \sigma(x')=0\\ \sigma(x'')=0}} \mathrm{P}_\Sigma(\sigma)$$

$$= \lambda_{\Phi,n'}\lambda_{\Phi,n''}\lambda^2_{\Sigma,0}. \tag{92}$$

Meanwhile, when $n' = 0, \cdots, N_\Phi - 1$ and $n'' = N_\Phi, \cdots, N_\Phi + N_\Psi - 1$, since Φ and Ψ are flat and Σ is two-flat, we have that (91) becomes:

$$\sum_{\substack{\upsilon\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi+N_\Psi}\\ \upsilon(x')=n'\\ \upsilon(x'')=n''}} \mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon) = \sum_{\substack{\varphi\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi}\\ \varphi(x')=n'}} \mathrm{P}_\Phi(\varphi) \sum_{\substack{\psi\in\mathbb{Z}^{\mathcal{X}}_{N_\Psi}\\ \psi(x'')=n''-N_\Phi}} \mathrm{P}_\Psi(\psi) \sum_{\substack{\sigma\in\mathbb{Z}_2^{\mathcal{X}}\\ \sigma(x')=0\\ \sigma(x'')=1}} \mathrm{P}_\Sigma(\sigma)$$

$$= \lambda_{\Phi,n'}\lambda_{\Psi,n''-N_\Phi}\lambda_{\Sigma,0}\lambda_{\Sigma,1}. \tag{93}$$

Similarly, for $n' = N_\Phi, \cdots, N_\Phi + N_\Psi - 1$ and $n'' = 0, \cdots, N_\Phi - 1$, (91) becomes:

$$\sum_{\substack{\upsilon\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi+N_\Psi}\\ \upsilon(x')=n'\\ \upsilon(x'')=n''}} \mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon) = \sum_{\substack{\varphi\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi}\\ \varphi(x'')=n''}} \mathrm{P}_\Phi(\varphi) \sum_{\substack{\psi\in\mathbb{Z}^{\mathcal{X}}_{N_\Psi}\\ \psi(x')=n'-N_\Phi}} \mathrm{P}_\Psi(\psi) \sum_{\substack{\sigma\in\mathbb{Z}_2^{\mathcal{X}}\\ \sigma(x')=1\\ \sigma(x'')=0}} \mathrm{P}_\Sigma(\sigma)$$

$$= \lambda_{\Phi,n''}\lambda_{\Psi,n'-N_\Phi}\lambda_{\Sigma,0}\lambda_{\Sigma,1}. \tag{94}$$

Finally, for $n', n'' = N_\Phi, \cdots, N_\Phi + N_\Psi - 1$, since Ψ and Σ are two-flat, (91) becomes:

$$\sum_{\substack{\upsilon\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi+N_\Psi}\\ \upsilon(x')=n'\\ \upsilon(x'')=n''}} \mathrm{P}_{\Phi\#_\Sigma\Psi}(\upsilon) = \sum_{\varphi\in\mathbb{Z}^{\mathcal{X}}_{N_\Phi}} \mathrm{P}_\Phi(\varphi) \sum_{\substack{\psi\in\mathbb{Z}^{\mathcal{X}}_{N_\Psi}\\ \psi(x')=n'-N_\Phi\\ \psi(x'')=n''-N_\Phi}} \mathrm{P}_\Psi(\psi) \sum_{\substack{\sigma\in\mathbb{Z}_2^{\mathcal{X}}\\ \sigma(x')=1\\ \sigma(x'')=1}} \mathrm{P}_\Sigma(\sigma)$$

$$= \lambda_{\Psi,n'-N_\Phi}\lambda_{\Psi,n''-N_\Phi}\lambda^2_{\Sigma,1}. \tag{95}$$

Thus, for any $x', x'' \in \mathcal{X}$ such that $x' \neq x''$, we have (92), (93), (94), or (95) meaning

$\Phi \#_{\Sigma} \Psi$ is two-flat (90), as claimed. $\square$

Having demonstrated the potential usefulness of local histograms in the analysis of images composed of distinct texture types, we now turn to using them in a segmentation-and-classification algorithm.

VI. A Proof-of-Concept Classification Algorithm

In this chapter, we discuss a preliminary segmentation-and-classification algorithm inspired by Theorem 4 in which local histograms are decomposed using principal component analysis (PCA). In particular, the work presented in this dissertation 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 Figure 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. 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 Figure 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.

Up to this point, we have presented a theoretical justification for using local histograms to segment and label images like Figure 1(a). In this chapter, we discuss a preliminary segmentation-and-classification algorithm inspired by Theorem 4. 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 4 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 [3]; these facts are not reprinted here.

From the point of view of our motivating application, the significance of Theorems 4 and 11 is that they give 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 4, it is reasonable to *demix*—decompose into convex combinations—the local histograms of that type into a set of more basic distributions. Indeed, it is reasonable to expect a local histogram computed over a region of cartilage (Figure 3(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 pseudovascular tissue (Figure 3(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. 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 algorithm we present in this chapter 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 Figure 10(a) or (d). For any tissue type $k = 1, \ldots, K$, 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. We note here that the computation of the variance of the local histogram transform, which was given in the previous chapter, demonstrated the importance of the scale of the local histogram window. In our algorithm, we use a Gaussian window and have chosen the scale experimentally to have a standard deviation of $16\sqrt{2}$. 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 (Figure 10(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 4, 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:

$$h_{k;m} \approx \sum_{n=1}^{N_k} \lambda_{k;m,n} g_{k;n}, \quad \forall m = 1, \ldots, M_k. \tag{96}$$

Unfortunately, computing the $g_{k;n}$'s that minimize the approximation error in (96) 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. PCA is a technique from linear algebra that takes a set of correlated local histograms, such as those computed at pixels of the same tissue type, and converts the set into distributions that are uncorrelated; these

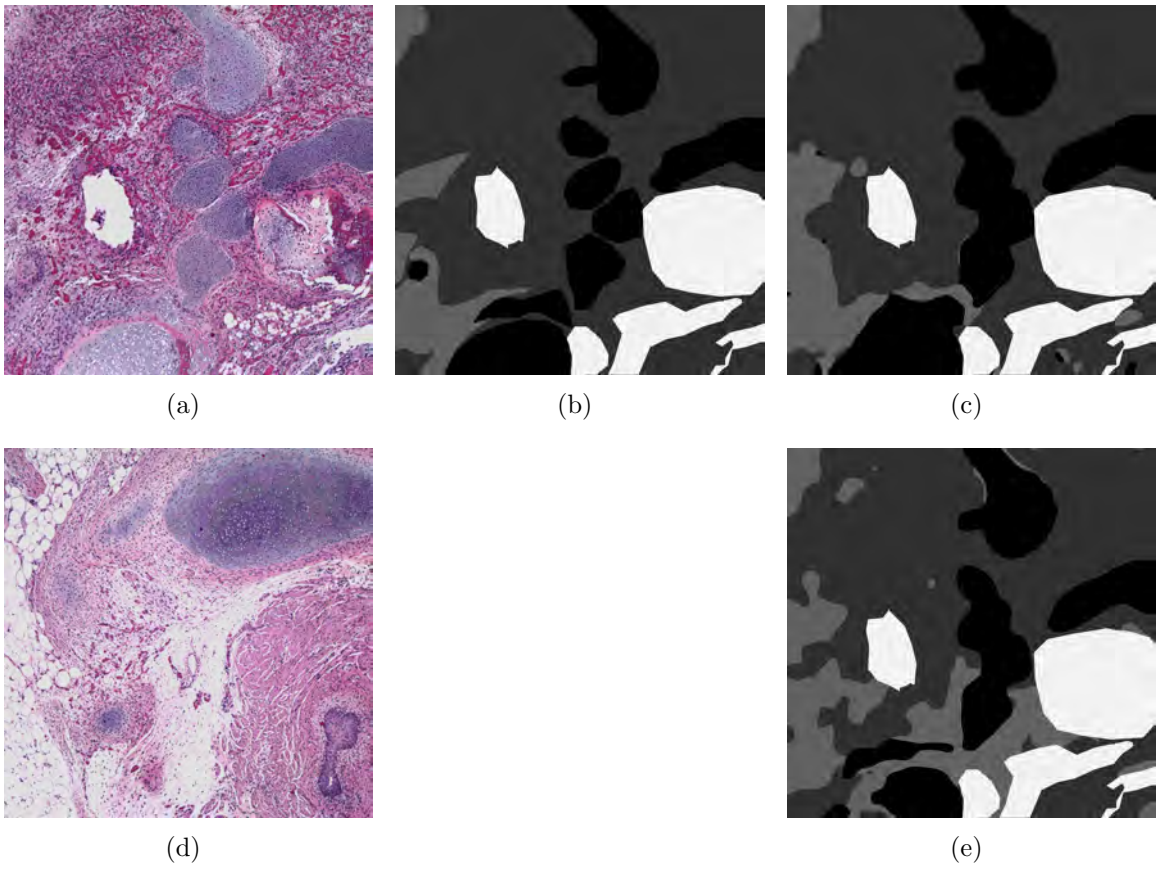


Figure 10. An example of using PCA of local histograms to perform segmentation and classification of the image given in (a), which is a 3-bit quantized version of Figure 1(a). A manually segmented and labeled version of (a) is shown in (b) where black represents cartilage, light gray represents connective tissue, dark gray represents pseudovascular tissue, and white represents other tissues that have been ignored in this proof-of-concept experimentation. That is, the (ignored) white pixels in (c) and (e) are copied directly from (b), while the black and gray pixels are the result of our classification algorithm. Using (a) as both the training and testing image in a PCA-based classification scheme (99), 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).

distributions are called the principal components. The first principal component has the highest possible variance overall while the other components have the highest possible variance with respect to the restriction that they must also be orthogonal to each other.

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 $\overline{h}_k$:

$$H_k(:, m) = h_{k;m} - \overline{h}_k \qquad \text{where} \qquad \overline{h}_k = \frac{1}{M_k} \sum_{m=1}^{M_k} h_{k;m}. \tag{97}$$

We then compute the singular value decompositions $H_k = U_k \Sigma_k V_k^{\mathrm{T}}$ 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 (96) is replaced by:

$$h_{k;m} \approx \overline{h}_k + \sum_{n=1}^{N_k} \langle h_{k;m} - \overline{h}_k, u_{k;n} \rangle u_{k;n}, \quad \forall m = 1, \ldots, M_k. \tag{98}$$

The classical theory of PCA states that the approximation error in (98) is optimally small in the sense that these specific $u_{k;n}$'s span the particular N_k-dimensional subspace of $\mathbb{R}^{\mathcal{Y}}$ whose orthogonal projection operator P_k minimizes the total squared-error $\sum_{m=1}^{M_k} \|h_{k;m} - \overline{h}_k - P_k(h_{k;m} - \overline{h}_k)\|^2$. The vectors $\overline{h}_k$ 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 $h_x : \mathcal{Y} \to \mathbb{R}$, $h_x(y) = (\mathrm{LH}_w f)(x, y)$ 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 $\overline{h}_k + \mathrm{span}\{u_{k;n}\}_{n=1}^{N_k}$ is nearest to h_x. Specifically,

we let:

$$
\begin{aligned}
k(x) &= \operatorname*{argmin}_{k=1,\dots,K} \left\| h_x - \overline{h}_k - \sum_{n=1}^{N_k} \langle h_x - \overline{h}_k, u_{k;n} \rangle u_{k;n} \right\|^2 \\
&= \operatorname*{argmin}_{k=1,\dots,K} \left(\|h_x - \overline{h}_k\|^2 - \sum_{n=1}^{N_k} |\langle h_x - \overline{h}_k, u_{k;n} \rangle|^2 \right) \\
&= \operatorname*{argmin}_{k=1,\dots,K} \left(\sum_{y \in \mathcal{Y}} |h_x(y) - \overline{h}_k(y)|^2 - \sum_{n=1}^{N_k} \left| \sum_{y \in \mathcal{Y}} \big(h_x(y) - \overline{h}_k(y) \big) u_{k;n}(y) \right|^2 \right). \qquad (99)
\end{aligned}
$$

In implementation, we compute the summations over $\mathcal{Y}$ in (99) 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 $\mathrm{LH}_w f$. By Theorem 1, 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 [33]. 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 Figure 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 2, 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 Figure 1(a) is given in Figure 10(a); for the sake of readability, a 3-bit quanitized 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 (99), running on standard desktop hardware. 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 ($\mathcal{X} = \mathbb{Z}_{600}^2$).

Two runs of this classification algorithm are depicted in Figure 10. In the first run, we train the classifier on the 3-bit 1200×1200 red-blue image given in Figure 10(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 Figure 10(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 $\overline{h}_k$ as well as the dominant left-singular vectors of H_k (97). 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 Figure 10(a) using the decision rule (99), resulting in per-pixel labels such as the ones given in Figure 10(c) for $N = 4$. Comparing Figure 10(c) and the ground truth of Figure 10(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	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 tell us that when using a single principal component, those points labeled as cartilage by a medical expert in Figure 10(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 Figure 10(d), making use of its ground truth labels (not pictured). We then apply the principal components obtained from Figure 10(d) to generate labels (Figure 10(e)) for Figure 10(a) using the decision rule (99). 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	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 4: 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.

VII. Conclusions

In this dissertation, we provided a rigorous mathematical analysis of local histograms to be used in the segmentation and classification of histology images. In particular, we found a nice relationship between local histograms and images that are composites of several distinct textures: local histograms distribute, on average, over flat occlusion models. Additionally, we showed that the distance of our local histograms from the average is highly dependent on the scale of the local histogram window. We also characterized all transforms which satisfy the three key properties of local histograms thus showing that local histograms are essentially the only transforms that distribute over random composites, on average. These results demonstrate the usefulness of local histograms in analyzing textures commonly found in histology images and give credence to certain types of segmentation-and-classification algorithms. We demonstrated the real-world potential of these ideas by using them to segment and classify tissues in a histology image.

We now discuss several ideas for future work. The first idea arises from the fact that storing local histograms of an image requires a significant amount of memory. Suppose we want to compare the local histogram against a library of ground truth histograms and assign a label based on the closest match in the library. Not only does computing the local histograms require a significant amount of memory, but the size of our library is limited because of memory. In order to reduce this cost, one could study sparse representations of local histograms, that is, see when they can be written as combinations of a few essential fundamental distributions. One could also investigate how classification accuracy suffers, if at all, as a result of such sparse approximation. Second, we note that up to this point, we have focused on local histograms in the case where the images in question are discrete. By extending this theory to the continuous case, one could apply local histograms not just to the

pixel values themselves, but also to real-valued quantities derived from them, such as image orientations. In particular, one could study how the local histogram transform interacts with *multiscale moment transforms* and incorporate orientation information into the segmentation-and-classification algorithm. Such geometrical context is vital when color alone is insufficient for proper classification. One could also exploit the theory of frames and filter banks to generalize the local histogram transform in a way that allows it to incorporate both high and low spatial frequency information. Next, one could use information theory metrics to compare the information content of the RGB and RB images. One could study how the distinct red-blue pairs are related to the possible green values and study when the green channel may be disregarded in the classification scheme. Finally, one could also apply this algorithm to a broader class of images, such as aerial and satellite images. Indeed, being able to automatically classify the terrain in such images has applications in image registration and visual navigation systems.

Bibliography

[1] Bhagavatula, Ramamurthy. *Towards Automatic Identification and Delineation of Tissues and Pathologies in H&E Stained Images*. PhD Thesis. Carnegie Mellon University, Pittsburgh, PA (September 2011).

[2] Bhagavatula, Ramamurthy, Matthew Fickus, John A. Ozolek, Carlos A. Castro, and Jelena 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," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 1041–1044 (April 2010).

[3] Bhagavatula, Ramamurthy, Melody L. Massar, Matthew Fickus, Carlos A. Castro, John A. Ozolek, Jelena Kovačević. "Automated Identification of Tissues for Digital Pathology," in preparation, to be submitted to: *IEEE Transactions on Image Processing*.

[4] Bordenave, Charles, Yann Gousseau, and Francois Roueff. "The Dead Leaves Model: A General Tessellation Modeling Occlusion," *Advances in Applied Probability*, 38: 31–46 (2006).

[5] Chebira, Amina, John A. Ozolek, Carlos A. Castro, William G. Jenkinson, Mukta Gore, Ramamurthy Bhagavatula, Irina Khaimovich, Shauna E. Ormon, Christopher S. Navara, Meena Sukh-wani, Kyle E. Orwig, Ahmi Ben-Yehudah, Gerald Schatten, Gustavo K. Rohde, and Jelena Kovačević. "Multiresolution Identification of Germ Layer Components in Teratomas Derived from Human and Nonhuman Primate Embryonic Stem Cells," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 979–982 (May 2008).

[6] Cui, Yan, Nils Hasler, Thorsten Thormählen, and Hans-Peter Seidel. "Scale Invariant Feature Transform with Irregular Orientation Histogram Binning," *Proceedings of the 6th International Conference on Image Analysis and Recognition*, 5627: 258–267 (2009).

[7] Dalal, Navneet and Bill Triggs. "Histograms of Oriented Gradients for Human Detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1: 886–893 (2005).

[8] Fickus, Matthew and Dustin G. Mixon. "Isotropic Moments over Integer Lattices," *Applied and Computational Harmonic Analysis*, 26: 77–96 (2009).

[9] Alemán-Flores, Miguel and Luis Álvarez-León. "Texture Classification through Multiscale Orientation Histogram Analysis," *Lecture Notes in Computer Science*, 2695: 479–493 (2003).

[10] Freeman, William T. and Michael Roth. "Orientation Histograms for Hand Gesture Recognition," *Proceedings of the IEEE International Workshop on Automatic Face and Gesture Recognition*, 296–301 (1995).

[11] Ghita, Ovidiu, Paul F. Whelan, and Dana E. Ilea. "Multi-resolution Texture Classification Based on Local Image Orientation," *Proceedings of the 5th International Conference on Image Analysis and Recognition, Lecture Notes in Computer Science*, 5112: 688–696 (2008).

[12] van Ginneken, Bram and Bart M. ter Haar Romeny. "Applications of Locally Orderless Images," *Journal of Visual Communication and Image Representation*, 11: 196–208 (2000).

[13] Griffin, Lewis D. "Scale-Imprecision Space," *Image and Vision Computing*, 15: 369–398 (1997).

[14] Hadjidemetriou, Efstathios, Michael D. Grossberg, and Shree K. Nayar. "Multiresolution Histograms and Their Use for Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26: 831–847 (July 2004).

[15] Ilea, Dana E., Ovidiu Ghita, and Paul F. Whelan. "Evaluation of Local Orientation for Texture Classification," *Proceedings of the 3rd International Conference on Computer Vision Theory and Applications*, 1: 357–364 (2008).

[16] Kass, Michael and Justin Solomon. "Smoothed Local Histogram Filters," *ACM Transactions on Graphics*, 29: 100/1–10 (July 2010).

[17] Kober, Vitaly, Tomasz Cichocki, Marcin Gedziorowski, and Tomasz Szoplik. "Optical - Digital Method of Local Histogram Calculation by Threshold Decomposition," *Applied Optics*, 32: 692–698 (February 1993).

[18] Koenderink, Jan J. and Andrea J. van Doorn. "The Structure of Locally Orderless Images," *International Journal of Computer Vision*, 31: 159–168 (1999).

[19] Koenderink, Jan J. and Andrea J. van Doorn. "Blur and Disorder," *Journal of Visual Communication and Image Representation*, 11: 237–244 (2000).

[20] Lee, Ann B. and David Mumford. "An Occlusion Model Generating Scale-Invariant Images," *Proceedings of the IEEE Workshop on Statistical and Computational Theories of Vision*, (1999).

[21] Lee, Ann B., David Mumford, and Jinggang Huang. "Occlusion Models for Natural Images: A Statistical Study of a Scale-Invariant Dead Leaves Model," *International Journal of Computer Vision*, 41: 35–59 (2001).

[22] Liu, Xiuwen and DeLiang Wang. "Texture Classification Using Spectral Histograms," *IEEE Transactions on Image Processing*, 12: 661–670 (June 2003).

[23] Liu, Xiuwen and DeLiang Wang. "Image and Texture Segmentation Using Local Spectral Histograms," *IEEE Transactions on Image Processing*, 15: 3066–3077 (October 2006).

[24] Liu, Xiuwen, DeLiang L. Wang, and Anuj Srivastava. "Image Segmentation Using Local Spectral Histograms," *Proceedings of the International Conference on Image Processing*, (2001).

[25] Massar, Melody L. *Time-Frequency Analysis of Terahertz Radar Signals for Rapid Heart and Breath Rate Detection*. MS Thesis. Air Force Institute of Technology, Wright-Patterson Air Force Base, OH (June 2008).

[26] Massar, Melody L., Ramamurthy Bhagavatula, Matthew Fickus, and Jelena Kovačević. "Local Histograms for Classifying H&E Stained Tissues," *Proceedings of the 26th Southern Biomedical Engineering Conference*, 348–352 (2010).

[27] Massar, Melody L., Ramamurthy Bhagavatula, Matthew Fickus, and Jelena Kovačević. "Local Histograms and Image Occlusion Models," submitted to: *Applied and Computational Harmonic Analysis*.

[28] Massar, Melody L., Ramamurthy Bhagavatula, John A. Ozolek, Carlos A. Castro, Matthew Fickus, Jelena Kovačević. "A Domain-Knowledge-Inspired Mathematical Framework for the Description and Classification of H&E Stained Histopathology Images," *Proceedings of SPIE*, 8138: 81380U/1–7 (2011).

[29] Massar, Melody L. and Matthew Fickus. "Local Histograms and Texture Classification," in preparation, to be submitted to: *Journal of Fourier Analysis and Applications*.

[30] Massar, Melody L., Matthew Fickus, Erik Bryan, Douglas T. Petkie, and Andrew J. Terzuoli, Jr. "Fast Computation of Spectral Centroids," *Advances in Computational Mathematics*, 35: 83–97 (2011).

[31] Mumford, David and Basilis Gidas. "Stochastic Models for Generic Images," *Quarterly of Applied Mathematics*, 59: 85–111 (2000).

[32] Ni, Kangyu, Xavier Bresson, Tony Chan, and Selim Esedoglu. "Local Histogram Based Segmentation using the Wasserstein Distance," *International Journal of Computer Vision*, 97–111 (2009).

[33] Srinivasa, Gowri, Matthew Fickus, Yusong Guo, Adam D. Linstedt, Jelena Kovačević. "Active Mask Segmentation of Fluorescence Microscope Images," *IEEE Transactions on Image Processing*, 18: 1817–1829 (August 2009).

[34] Szoplik, Tomasz and Marcin Gedziorowski. "Optical - Digital Processors for Morphological and Rank Order Filtering," *Optics & Laser Technology*, 28: 73–82 (1996).

[35] Tomasi, Carlo and Roberto Manduchi. "Bilateral Filtering for Gray and Color Images," *Proceedings of the IEEE International Conference on Computer Vision*, 839–846 (January 1998).

[36] van de Weijer, Joost and Rein van den Boomgaard. "Local Mode Filtering," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2: 428–433 (December 2001).

[37] Ying, Zhengrong and David Castañon. "Statistical Model for Occluded Object Recognition," *International Conference on Information Intelligence and Systems*, 324–327 (1999).

[38] Ying, Zhengrong and David Castañon. "Partially Occluded Object Recognition Using Statistical Models," *International Journal of Computer Vision*, 49: 57–78 (2002).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
22-03-2012	PhD Dissertation	Jun 2008 - Mar 2012

4. TITLE AND SUBTITLE

Local Histograms for Per-Pixel Classification

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Massar, Melody L.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/DAM/ENC/12-02

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution A. Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We introduce a rigorous mathematical theory for the analysis of local histograms, and study how they interact with textures that can be modeled as occlusions of simpler components. We first show how local histograms can be computed as a system of convolutions and discuss some basic local histogram properties. We then introduce a probabilistic, occlusion-based model for textures and formally demonstrate that local histogram transforms are natural tools for analyzing the textures produced by our model. Next, we characterize all nonlinear transforms which satisfy the three key properties of local histograms and consider the appropriateness of local histogram features in the automated classification of textures commonly encountered in histological images. We discuss how local histogram transforms can be used to produce numerical features that, when fed into mainstream classification schemes, mimic the baser aspects of a pathologist's thought process.

15. SUBJECT TERMS

Local Histogram, Occlusion, Texture, Classification, Segmentation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Matthew Fickus
U	U	U	UU	103	19b. TELEPHONE NUMBER *(Include area code)* (937) 255-3636 x 4513 matthew.fickus@afit.edu

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18