Texture Classification in Pulmonary CT

Lauge Sørensen   Mehrdad J. Gangeh   Saher B. Shaker

Marleen de Bruijne

Abstract

This chapter presents a comparative study of texture classification in computed tomography images of the human lungs. Popular texture descriptors used in the medical image analysis literature for texture-based emphysema classification are described and evaluated within the same classification framework. Further, it is investigated whether combining the different descriptors is beneficial.

1 Introduction

Computed tomography (CT) images of the human lungs show intensity variations that form certain repeated patterns, also referred to as texture. Some of these variations are due to noise and artifacts, but others are due to properties of the imaged tissue. Abnormal lung tissue can often be distinguished from healthy lung tissue based on its texture. The texture of the abnormal lung tissue depends on the type and severity of the abnormality. Three examples of regions of lung tissue extracted from CT are presented in Figure 1, an example of normal, healthy lung tissue in
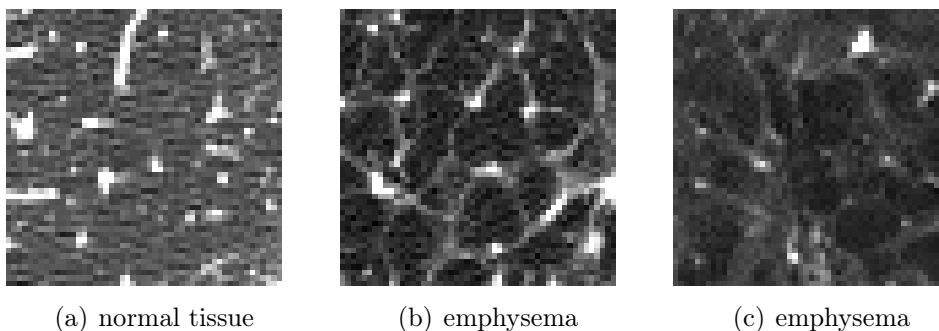

(a) normal tissue   (b) emphysema   (c) emphysema

Figure 1: 51×51 pixel regions, from CT images, with examples of normal and emphysematous lung tissue. The displayed image intensities are linearly scaled between -1024 Hounsfield units (HU) (black) and -300 HU (white).

(a) emphysema 1   (b) emphysema 2

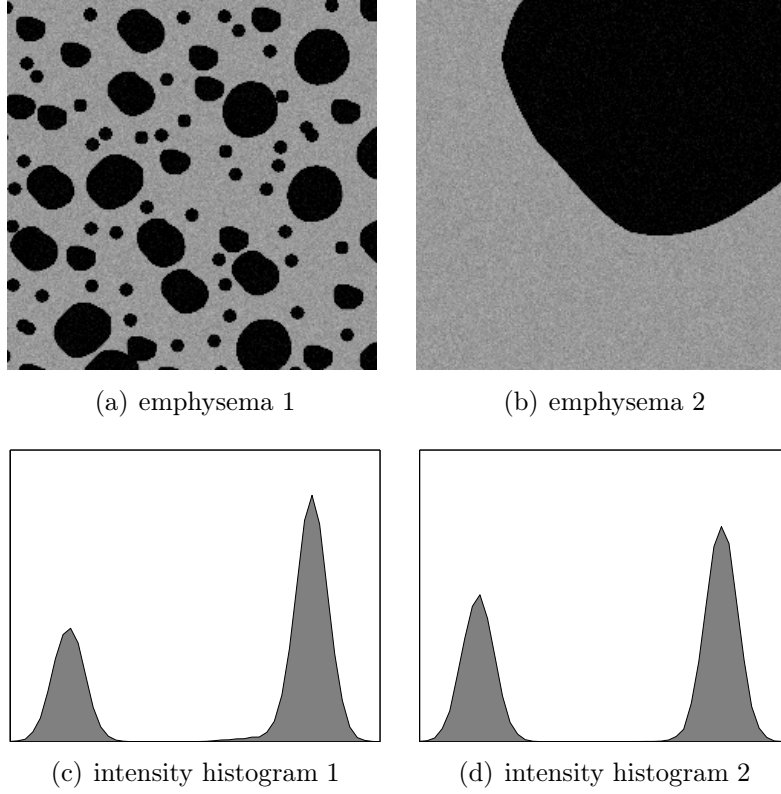(c) intensity histogram 1   (d) intensity histogram 2

Figure 2: Artificial examples of two abnormal patterns with different appearance in CT and the corresponding intensity histograms. Gray is normal tissue and black are air pockets due to missing tissue. The intensity histograms are rather similar for the two cases, even though the two abnormal patterns look very different.

Figure 1(a) and two examples of abnormal lung tissue, in this case emphysema, in Figures 1(b) and 1(c). The normal lung tissue exhibits a noisy, gray pattern with thin, bright structures, which are mainly vessels and airway walls. The abnormal lung tissue is in this case characterized by black blobs, due to missing lung tissue, of varying sizes within the normal appearing tissue.

There exists no generally agreed upon definition of texture, and the definitions provided in the literature often depend on the particular application at hand [15, 37]. When referring to texture in this chapter, we loosely mean: *a distribution of measure(s) of local structure, as measured in a region.* In the extreme case, the measure is the pixel values, and we simply have the histogram of intensity values computed in the region. Purely intensity-based descriptors may, however, not be sufficient to discriminate abnormal lung tissue as illustrated in Figure 2. Even though the textural appearance of the two example patterns is rather different, the intensity histograms look very similar. This chapter focuses on texture classification

in pulmonary CT, and all the considered texture descriptors, except for the purely intensity-based descriptors presented in Section 2.1, use interactions among pixels and should therefore be able to discriminating the two example patterns.

Several popular texture descriptors used in the medical image analysis literature for texture-based classification of abnormal lung tissue, e.g., for discriminating between textures of the types shown in Figure 1, are described and evaluated within the same classification framework. The descriptors are evaluated on a case-study of emphysema classification in CT regions of interest (ROIs). However, the presented descriptors can also be used for other texture analysis tasks as well, including texture segmentation and texture synthesis [15, 37]. Further, a classification system capable of classifying abnormality can serve as a basis for different medical image analysis tasks such as texture-based detection or quantification of abnormality in complete scans [18, 28, 30, 32, 34].

Section 2 describes the different texture descriptors, and Section 3 presents the classification problem, the experimental setup including the general classification system used for all descriptors, as well as the obtained results. Finally, Section 4 concludes the chapter.

2 Texture Descriptors

There exist many texture descriptors in the image processing and pattern recognition literature, see [8, 15, 16, 37], for example. Many of these have also been used for texture-based classification in pulmonary CT [2, 7, 18, 20, 21, 27, 29–34, 38, 39, 45], and in this chapter we focus on the descriptors most often used for this purpose. The following descriptors are considered:

- the intensity histogram (INT). Section 2.1. Used in [2, 18, 20, 21, 27, 29–31, 34, 38, 39, 45];

- local binary patterns (LBPs). Section 2.2. Used in [32–34];

- a Gaussian filter bank (GFB). Section 2.3. Different variants are used in [27, 29–31, 34];

- gray-level co-occurrence matrices (GLCMs). Section 2.4. Used in [2, 18, 20, 21, 34, 38, 39, 45];

- gray-level run-length matrices (GLRLMs). Section 2.5. Used in [2, 18, 20, 21, 34, 38, 39, 45];

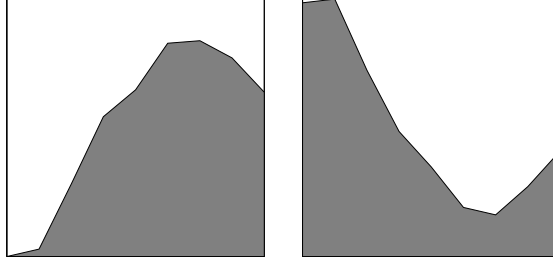- and finally, textons. Section 2.6. Used in [7].

Figure 3: Intensity histograms computed from the regions in Figure 1(a) (left) and Figure 1(b) (right). The histograms are quantized into $N_b = 9$ bins.

We focus on 2-D versions of the texture descriptors since the evaluation is performed on a data set of two-dimensional CT images. However, possible 3-D versions will also be discussed briefly in the end of the chapter in Section 4. Where possible, we consider rotation invariant versions of the descriptors since this property is often enforced in texture-based classification in CT. See [2, 18, 30, 34, 39], for example.

The representations obtained using the different texture descriptors of the central 31×31 pixel parts of the two example regions in Figure 1(a) and Figure 1(b) are shown in the following subsections to illustrate how the different descriptors represent the information in the two regions.

2.1   Intensity Histogram

The intensity histogram (INT) is a simple descriptor of independent pixel information in the image, and either the full histogram or summarizing measures can be used. Several studies use moments of the intensity distribution as features [2, 18, 20, 21, 34, 38, 39, 45] or of the distribution of smoothed intensities [27, 29, 34], while others use the full histogram of (smoothed) intensities [30, 31, 34]. The most commonly used summarizing measures are as follows: the mean intensity, or first moment

$$\mu(I) = \frac{1}{N_p} \sum_{i=1}^{N_p} I(\mathbf{x}_i) \tag{1}$$

where I is an image region, $\mathbf{x}_i$ is pixel within the region, and N_p is the number of pixels in the region; and the standard deviation of the intensity, or square root of the second central moment

$$\sigma(I) = \sqrt{\frac{1}{N_p - 1} \sum_{i=1}^{N_p} (I(\mathbf{x}_i) - \mu(I))^2}. \tag{2}$$

4

(a) $LBP2(\cdot;1,8)^{ri}$ (b) $LBP2(\cdot;1,8)^{riu2}$

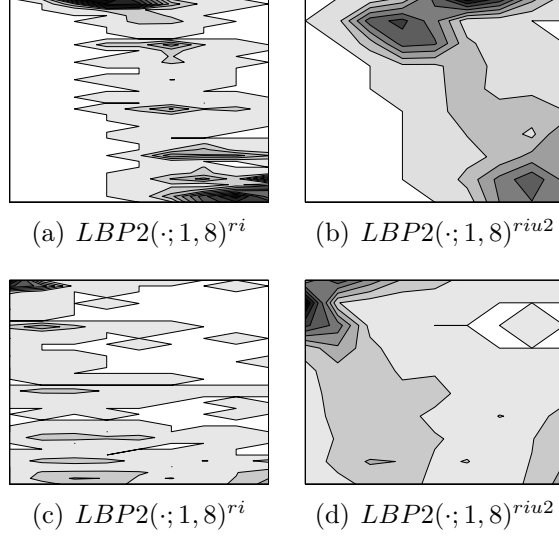(c) $LBP2(\cdot;1,8)^{ri}$ (d) $LBP2(\cdot;1,8)^{riu2}$

Figure 4: Joint LBP and intensity histograms using $R = 1$, $P = 8$, and $N_B = 9$. The histograms in the top row are computed from the region in Figure 1(a) and the histograms in the bottom row are computed from the region in Figure 1(b). Intensity is in the horizontal direction and LBP code is in the vertical direction. Black is high density whereas white is low density.

Further, a measure of asymmetry called skewness, or the third standardized central moment

$$skew(I) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\frac{I(\mathbf{x}_i) - \mu(I)}{\sigma(I)} \right)^3 \tag{3}$$

as well as a measure of peakedness called kurtosis, or the fourth standardized central moment

$$kurtosis(I) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\frac{I(\mathbf{x}_i) - \mu(I)}{\sigma(I)} \right)^4 - 3, \tag{4}$$

are often used.

Both the full intensity histogram as well as the four mentioned moments will be considered in the experiments.

Figure 3 shows intensity histograms computed from the example regions in Figure 1(a) and Figure 1(b). As expected, there is more mass in the low-intensity part of histogram of the emphysema region compared to the histogram computed from the normal lung tissue.

2.2 Local Binary Patterns

Local binary patterns (LBPs) were originally proposed by Ojala *et al.* as a gray-scale invariant measure for characterizing local structure in a 3×3 pixel neighborhood [16]. Later, a more general formulation was proposed that further allowed for multi-resolution analysis, rotation invariance, and vast dimensionality reduction by considering only so-called "uniform" patterns [17].

The LBPs are obtained from an image region I by thresholding samples in a local neighborhood with respect to each center pixel intensity and are given by

$$LBP(\mathbf{x}; R, P) = \sum_{p=0}^{P-1} H(I(\mathbf{x}_p) - I(\mathbf{x}))2^p \tag{5}$$

where $\mathbf{x}$ is the center pixel, $\mathbf{x}_p = [-R\sin(2\pi p/P), R\cos(2\pi p/P)]^T + \mathbf{x}$ are P local samples taken at a radius R around $\mathbf{x}$, and $H(\cdot)$ is the Heaviside function. By choosing a fixed sample position on the circle as the "leading bit", the thresholded samples can be interpreted as bits, and a P bit binary number, the LBP code, can be computed. The LBP codes computed in each pixel in the region are accumulated into a histogram, the LBP histogram, where each unique LBP code corresponds to a particular histogram bin. Different local structures in the region such as edges, corners, spots, etc., are assigned different LBP codes, and the LBP histogram, therefore, captures many structures in one unified framework. The number of samples, P, governs the number of bins in the LBP histogram, e.g., $P = 8$ results in $2^8 = 256$ bins. By varying the radius R and the number of samples P, the structures are measured at different scales, and this way LBP allows for measuring large scale structures without smoothing effects as would for example occur with Gaussian-based filters. As long as the relative ordering among the gray-scale values in the samples does not change, the output of (5) stays the same; hence, LBPs are invariant to any monotonic gray-scale transformation.

Rotation invariance can be achieved by "rotating the circle" until the lowest possible binary number is found

$$LBP^{ri}(\mathbf{x}; R, P) = \min_i(ROR(LBP(\mathbf{x}; R, P), i)) \tag{6}$$

for $i = 0, \ldots, P-1$. $ROR(b, i)$ performs i circular bit-wise right shifts on the P-bit binary number b. (6) assigns the same LBP code to similar structures of different orientations, which also has the effect of reducing the dimensionality of the LBP histogram, e.g., to 36 bins for $P = 8$.

The dimensionality of the LBP histogram can further be reduced by only con-

sidering the so-called "uniform" patterns, which are local structures giving rise to at most two bit changes in the LBP code, and assign the remaining structures a designated "noise" code. This is done in the following way:

$$LBP^{riu2}(\mathbf{x}; R, P) = \begin{cases} \sum_{p=0}^{P-1} H(I(\mathbf{x}_p) - I(\mathbf{x})) & \text{if } U(LBP(\mathbf{x}; R, P)) \leq 2 \\ P + 1 & \text{otherwise,} \end{cases} \quad (7)$$

where

$$U(LBP(\mathbf{x}; R, P)) = |H(I(\mathbf{x}_{P-1}) - I(\mathbf{x})) - H(I(\mathbf{x}_0) - I(\mathbf{x}))| +$$
$$\sum_{p=1}^{P-1} |H(I(\mathbf{x}_p) - I(\mathbf{x})) - H(I(\mathbf{x}_{p-1}) - I(\mathbf{x}))|. \quad (8)$$

(8) counts the number of bit transitions, from 0 to 1 or from 1 to 0, that are encountered when traversing the P thresholded samples on the circle. The resulting LBP histogram contains 10 bins for $P = 8$.

Since LBPs by design are invariant to any monotonic intensity transformation, important discriminative information may be left out when applied in CT. The joint LBP and intensity histogram has been proposed as a solution to this problem [32–34], resulting in a histogram where one dimension measures structure whereas the other dimension measures at which densities the different structures occur. The histogram is obtained by computing the LBP codes in all pixels in the image, according to (6) or (7), and forming the joint histogram between the resulting LBP codes and the intensities in the corresponding center pixels $I(\mathbf{x})$. We denote these operators $LBP2^{ri}$ and $LBP2^{riu2}$, respectively. These are the operators that we will be using in the experiments.

Figure 4 shows joint LBP and intensity histograms computed from the example regions in Figure 1(a) and Figure 1(b) using $LBP2(\cdot; 1, 8)^{ri}$ and $LBP2(\cdot; 1, 8)^{riu2}$. The LBPs, or local structures, generally reside at lower intensities for emphysema compared to healthy tissue, which is to be expected.

2.3 Gaussian Derivative-Based Filter Bank

Filtering an image region using a bank of filters and representing the resulting filter response images using histograms or moments is another commonly employed texture descriptor [40, 47]. We evaluate a multi-scale Gaussian filter bank (GFB) comprising filters based on the Gaussian function and its derivatives [36]. The filtering is done by convolving the image region with the two-dimensional Gaussian
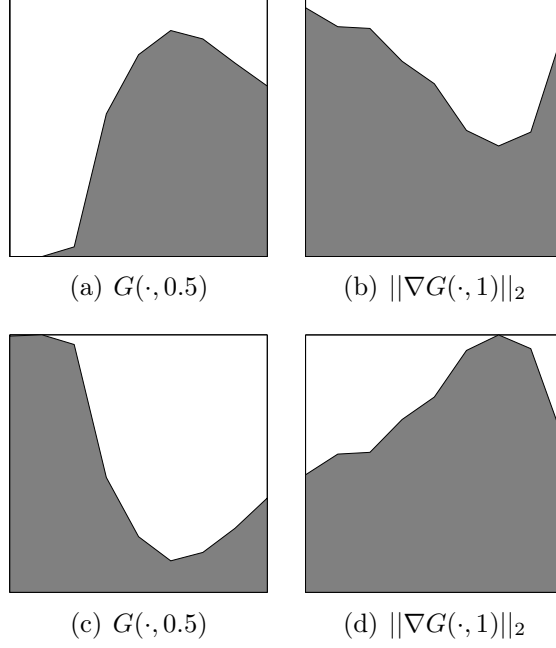
(a) $G(\cdot, 0.5)$ (b) $||\nabla G(\cdot, 1)||_2$

(c) $G(\cdot, 0.5)$ (d) $||\nabla G(\cdot, 1)||_2$

Figure 5: GFB filter response histograms obtained using the filters $G(\cdot, 0.5)$ and $||\nabla G(\cdot, 1)||_2$. The filter responses are quantized into $N_b = 9$ bins. The histograms in the top row are computed from the region in Figure 1(a) and the histograms in the bottom row are computed from the region in Figure 1(b).

function

$$G(\mathbf{x}; \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{||\mathbf{x}||_2^2}{2\sigma^2} \right) \qquad (9)$$

where σ is the standard deviation, or scale, and $|| \cdot ||_2$ is the L_2-norm. Increasing σ results in the region being increasingly blurred, which allows analysis of the image structure at larger scales.

Letting L_x and L_y denote the first order derivatives of the convolved image region $L = I * G(\mathbf{x}; \sigma)$, and L_{xx}, L_{yy} and L_{xy} denote the second order derivatives, the four base filters in the GFB considered in this chapter are as follows: the Gaussian function (9) itself; gradient magnitude

$$||\nabla G(\mathbf{x}; \sigma)||_2 = \sqrt{L_x^2 + L_y^2}; \qquad (10)$$

the Laplacian of the Gaussian

$$\nabla^2 G(\mathbf{x}; \sigma) = L_{xx} + L_{yy}; \qquad (11)$$

8

and Gaussian curvature

$$K(\mathbf{x}; \sigma) = L_{xx} L_{yy} - L_{xy}^2. \tag{12}$$

These four filters are used in [34] and are common descriptors of local structure invariant to rotation. The Gaussian function itself is included in order to allow the GFB descriptor to take CT density information into account [29, 34].

From a filtered CT image region, a filter response histogram can be computed and used to characterize the texture [30, 34]. Alternatively, moments can be used [28, 29]. We experiment with both approaches in this chapter, and use the four moments defined in Section 2.1, namely, the mean (1), standard deviation (2), skewness (3), and kurtosis (4) of the filter responses computed in the region, as well as the full filter response histograms.

Figure 5 shows GFB filter response histograms computed from the example regions in Figure 1(a) and Figure 1(b) using the two different filters $G(\cdot, 0.5)$ and $||\nabla G(\cdot, 1)||_2$. As expected, there is more mass in the low-intensity part of the histogram computed from the emphysema region compared to the histogram computed from the normal lung tissue, see Figure 5(a) and Figure 5(c). The emphysema region is also characterized by having edges at a different strength, due to transitions from air pockets to lung tissue, compared to the normal lung tissue, see Figure 5(b) and Figure 5(d).

2.4 Gray-Level Co-Occurrence Matrices

Summarizing measures computed from a gray-level co-occurrence matrix (GLCM) are classical texture features [8]. An element $p(i, j; \theta, d)$ of a GLCM computed from an image region I expresses the number of times a pixel $\mathbf{x}$ with gray-level i occurs jointly with a pixel $\mathbf{y}$ with gray-level j in I, where $\mathbf{y}$ is positioned relative to $\mathbf{x}$ according to a displacement vector defined be an angle θ and a distance d. In this chapter, we consider symmetric GLCMs, meaning that $p(i, j; \theta, d) = p(j, i; \theta, d)$. Of the summarizing measures originally proposed in [8], the three most commonly used are: angular second moment or uniformity of energy (UE)

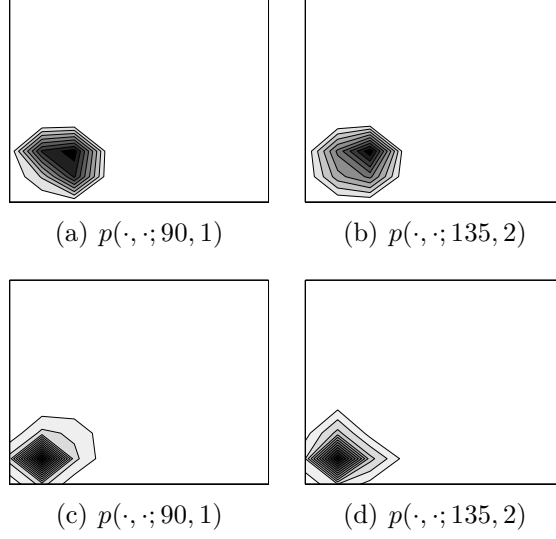$$UE(p; \theta, d) = \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} p(i, j; \theta, d)^2 \tag{13}$$

(a) $p(\cdot, \cdot; 90, 1)$ (b) $p(\cdot, \cdot; 135, 2)$

(c) $p(\cdot, \cdot; 90, 1)$ (d) $p(\cdot, \cdot; 135, 2)$

Figure 6: GLCMs using the two displacements ($\theta = 90°, d = 1$) and ($\theta = 135°, d = 2$). The GLCMs in the top row are computed from the region in Figure 1(a) and the GLCMs in the bottom row are computed from the region in Figure 1(b). The intensities are quantized into $N_b = 9$ bins. Black is high density whereas white is low density.

where N_b is the number of bins that the gray-levels in the region are quantized into; inertia or contrast (C)

$$C(p; \theta, d) = \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} |i - j|^2 p(i, j; \theta, d); \qquad (14)$$

and homogeneity or inverse difference moment (IDM)

$$IDM(p; \theta, d) = \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \frac{p(i, j; \theta, d)}{1 + |i - j|^2}. \qquad (15)$$

These measures are all used in [2, 18, 20, 21, 34, 38, 39, 45]. We further include correlation (CORR)

$$CORR(p; \theta, d) = \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} \frac{(i - \mu_i)(j - \mu_j)p(i, j; \theta, d)}{\sigma_i \sigma_j} \qquad (16)$$

where $\mu_i = \sum_{i=1}^{N_b} \sum_{j=1}^{N_b} ip(i, j; \theta, d)$ and $\sigma_i = \sqrt{\sum_{i=1}^{N_b} \sum_{j=1}^{N_b} p(i, j; \theta, d)(i - \mu_i)^2}$ is the GLCM mean and the GLCM standard deviation, respectively. Note, that since symmetric GLCMs are considered, $\mu_i = \mu_j$ and $\sigma_i = \sigma_j$. $CORR(\theta, d)$ is used in [18, 34, 38, 39, 45].
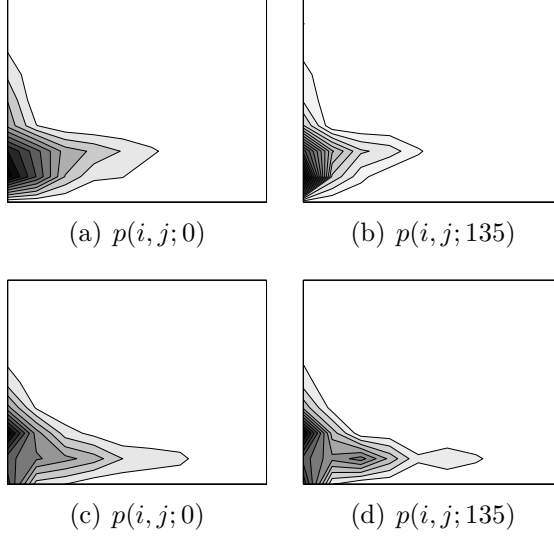
Figure 7: GLRLMs using the two orientations $\theta = 0°$ and $\theta = 135°$. The GLRLMs in the top row are computed from the region in Figure 1(a) and the GLRLMs in the bottom row are computed from the region in Figure 1(b). The intensities are quantized into $N_b = 9$ bins and the maximum run length is $MRL = 10$. Intensity is in the vertical direction and run-length is in the horizontal direction. Black is high density whereas white is low density.

The two parameters θ and d govern the GLCM, and different values of these parameters result in different GLCMs capturing spatial pixel intensity relations in various directions and at various ranges. Some studies average the summarizing measures across different displacements, which brings down the dimensionality of the descriptor and makes it scale and rotation invariant, see [2, 18], for example. We evaluate the common approach of using summarizing measures, but also experiment with using the full GLCMs. Both the summarizing measures and the full GLCMs are averaged across orientation to achieve rotation invariance, but not across distance.

GLCMs for two different displacements computed from the example regions in Figure 1(a) and Figure 1(b) are shown in Figure 6. The GLCMs are more peaked, due to smaller differences between neighboring pixel intensities, and the mass is shifted towards the lower intensities, for the emphysema case compared to the healthy tissue case.

2.5   Gray-Level Run-Length Matrices

Summarizing measures computed from a gray-level run-length matrix (GLRLM) is another set of classical texture features [8]. An element $p(i, j; \theta)$ of a GLRLM computed from an image region I expresses the number of times j consecutive pixels

of the same gray-level i, in a given direction θ, occur in I. The five most commonly used summarizing measures of GLRLMs are: short run emphasis (SRE)

$$SRE(p;\theta, MRL) = \frac{1}{N_r} \sum_{i=1}^{N_b} \sum_{j=1}^{MRL} \frac{p(i,j;\theta)}{j^2} \tag{17}$$

where N_b is the number of bins that the gray-levels in the region are quantized into, MRL is the maximum run-length considered, and $N_r = \sum_{i=1}^{N_b} \sum_{j=1}^{MRL} p(i,j;\theta)$ is the total number of runs in the image; long run emphasis (LRE)

$$LRE(p;\theta, MRL) = \frac{1}{N_r} \sum_{i=1}^{N_b} \sum_{j=1}^{MRL} p(i,j;\theta)j^2; \tag{18}$$

gray-level nonuniformity (GLN)

$$GLN(p;\theta, MRL) = \frac{1}{N_r} \sum_{i=1}^{N_b} \left(\sum_{j=1}^{MRL} p(i,j;\theta) \right)^2; \tag{19}$$

run-length nonuniformity (RLN)

$$RLN(p;\theta, MRL) = \frac{1}{N_r} \sum_{j=1}^{MRL} \left(\sum_{i=1}^{N_b} p(i,j;\theta) \right)^2; \tag{20}$$

and run percentage (RP)

$$RP(p;\theta, MRL) = \frac{1}{N_p} \sum_{i=1}^{N_b} \sum_{j=1}^{MRL} p(i,j;\theta) \tag{21}$$

where N_p is the number of pixels in the region. These five measures are used in [2, 18, 20, 21, 34, 38, 39, 45]. The direction of the runs, θ, and the number of bins, N_b, that the intensities are quantized into, are the only parameters for the computation of the GLRLM. The maximum run-length considered, MRL, is an extra parameter for the computation of the summarizing measures, and MRL effectively truncates the GLRLM in the run-length direction. Also note that N_r is influenced by MRL, since this is the total number of runs in the truncated GLRLM. Most studies average the summarizing measures across orientations, which decreases the dimensionality of the descriptor and makes it rotation invariant. We evaluate the common approach of using summarizing measures, but also experiment with using the full GLRLMs. Both the summarizing measures and the full GLRLMs are averaged across orientation.

(a) $TS = 3, k = 30$ (b) $TS = 3, k = 20$

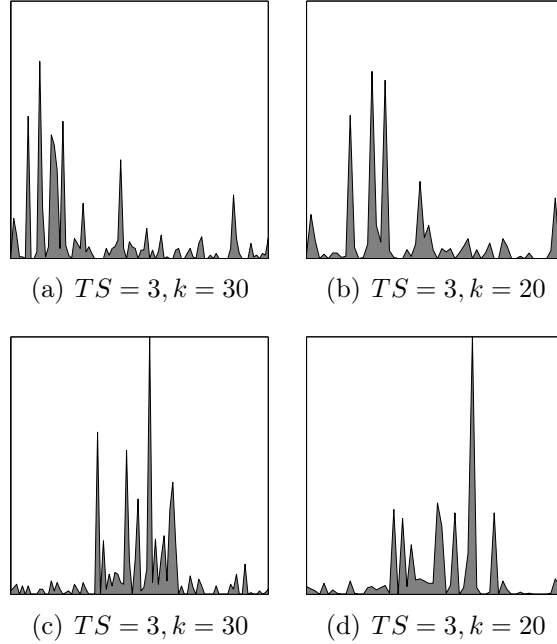(c) $TS = 3, k = 30$ (d) $TS = 3, k = 20$

Figure 8: Texton histograms using $TS = 3$ in conjunction with $k = 30$ or $k = 20$. The Texton histograms in the top row are computed from the region in Figure 1(a) and the histograms in the bottom row are computed from the region in Figure 1(b).

GLRLMs computed from the example regions in Figure 1(a) and Figure 1(b) are shown in Figure 7. The emphysema case contains longer runs at a lower intensity, which is related to areas with missing lung tissue, compared to the healthy case.

2.6   Textons

Textons were first introduced by Julesz in his influential paper [10] as fundamental primitives capable of representing texture. Based on Julesz' theory, Leung and Malik developed a complete texture classification system [13]. They defined 2-D textons as the cluster centers in filter bank responses, which made it possible to generate textons from the images automatically as the prototypes representing the source textures. These textons formed a codebook from which a texton histogram could be constructed for each image using a similarity measure. This work was further improved by Schmid [23, 24], Cula and Dana [3], and Varma and Zisserman [41–43]. The texton histogram computed from an image is a fully data-driven texture descriptor. This is different from the other descriptors considered in this chapter, which are based on pre-specified rules for which only the associated parameters are optimized on the training data.

There are three main representations associated with the texton-based approach

in the literature, filter banks [3, 13, 42], raw pixel representation [43], and Markov random field representation, where the probability of the central pixel value is conditioned on the pixel values in the neighboring pixels [43]. However, irrespective of the representation used to describe local image information, the descriptor construction consists of two stages: construction of a texton codebook and computation of a texton frequency histogram.

To construct the texton codebook, small-sized local patches of size $TS \times TS$ are randomly extracted from each image region in the training set and converted to an appropriate representation. The patches are subsequently aggregated over all images in a class and clustered using a clustering algorithm. The cluster centers obtained form a codebook of textons representing the class of textures. The size of the final codebook thus depends on the number of cluster centers N_c as well as the number of classes. For example, for a three-class problem with 30 cluster centers per class, $3 \times 30 = 90$ textons are generated in the codebook. The texton codebook is used to compute texton histograms, both for the training data and the test data.

The texton frequency histogram is defined as a histogram of the same size as the size of the codebook where each bin is associated with one of the learned textons in the codebook. A histogram is computed from an image region by extracting patches of the same size as the textons, one centered on each pixel in the region. Each extracted patch is converted to the same representation as is used in the texton codebook and compared to all textons in the codebook using a similarity measure. The bin corresponding to the closest texton is updated with a count. The final histogram is normalized.

In this chapter, the textons are based on a raw pixel representation, and the codebook thus consists of cluster centers in a feature space of dimension $TS \times TS$ where each dimension corresponds to the intensity in a certain pixel. The texton frequency histogram is therefore variant to both rotation and scale. Clustering is performed using the k-means algorithm [9], and the Euclidean distance in this space is used as similarity measure between the small patches, both in k-means and in the texton histogram formation.

Figure 8 shows texton histograms from the example regions in Figure 1(a) and Figure 1(b). It is not straightforward to interpret the texton histograms, since the ordering of the bins is not meaningful. Each bin merely corresponds to a certain texton in the codebook and neighboring bins need not be similar. For $N_c = 10$, the left-most 10 bins correspond to cluster centers computed from the healthy samples in the training set, the center 10 bins correspond to one type of emphysema, and the right-most 10 bins correspond to another type of emphysema. As can be seen, the

histograms computed from the healthy region contain most mass in the left-most part, and the histograms computed from the emphysema region contain most mass in the central part, which is the correct emphysema type.

3 Evaluation

The experiments are performed in Matlab using in-house developed code as well as publicly available implementations of some of the descriptors. LBPs are computed using `lbp.m`[1] and `getmapping.m`[2]. GLCMs are computed using version 6.2 of Mathworks's own Image Processing Toolbox for Matlab. GLRLMs are computed using the Gray Level Run Length Matrix Toolbox[3].

3.1 A Case-Study: Classification of Emphysema in CT

The different texture descriptors are evaluated and compared on the classification of different types of emphysema in CT regions of interest (ROIs). Emphysema is one of the main components of chronic obstructive pulmonary disease (COPD), the fourth leading cause of morbidity and mortality in the United States alone, which is predicted to rise from its ranking in 2000 as the 12th most prevalent disease worldwide to the 5th, and from the 6th most common cause of death worldwide to the 3rd, by 2020 [1, 22]. Automated emphysema classification may lead to better sub-typing, diagnosis, and understanding of the development of emphysema, but more importantly, to a better assessment of COPD.

Emphysema lesions, or bullae, are visible in CT images as areas of abnormally low attenuation values close to that of air, and are often characterized according to their subtype and extent. Subtyping is performed according to the morphology, or textural appearance, of the emphysema patterns, and can according to a common definition be classified into three subtypes, or patterns, [44]. These subtypes are the following: centrilobular emphysema (CLE), defined as multiple small low-attenuation areas; paraseptal emphysema (PSE), defined as multiple low-attenuation areas in a single layer along the pleura often surrounded by interlobular septa that is visible as thin white walls; and panlobular emphysema (PLE), defined as a low-attenuation lung with fewer and smaller pulmonary vessels. Examples of emphysema, more specifically CLE, as well as normal tissue (NT), are shown in

---

[1]version 0.3.2 obtained from http://www.ee.oulu.fi/mvg/page/lbp_matlab

[2]version 0.1.1 obtained from http://www.ee.oulu.fi/mvg/page/lbp_matlab

[3]version 1.0 obtained from http://www.mathworks.com/matlabcentral/fileexchange/17482-gray-level-run-length-matrix-toolbox

Figure 1.

Current quantitative analysis of emphysema from CT relies only on the intensity information in individual pixels [25] and is not able to distinguish between the two abnormal patterns in Figure 2. This includes measures such as the relative area, also called emphysema index, and the percentile density [25]. In recent years, various studies have looked into using richer texture descriptors for improving classification of emphysema [2, 7, 18, 20, 21, 29–34, 38, 39, 45].

3.2 Data

The data comes from an exploratory study carried out at the Department of Respiratory Medicine, Gentofte University Hospital [26] and consist of CT images of the thorax acquired using General Electric equipment (LightSpeed QX/i; GE Medical Systems, Milwaukee, WI, USA) with four detector rows. A total of 75 high resolution CT (HRCT) slices were acquired by scanning 25 subjects in the upper, middle, and lower lung. The CT scanning was performed using the following parameters: in-plane resolution 0.78×0.78 mm, 1.25 mm slice thickness, tube voltage 140 kV, and tube current 200 mAs. The slices were reconstructed using an edge enhancing (bone) algorithm.

An experienced chest radiologist and a CT experienced pulmonologist each assessed the leading pattern, either NT, CLE, or PSE, in each of the 75 slices, and consensus was obtained in cases of disagreement. This resulted in 26 NT slices, 21 CLE slices, and 28 PSE slices. 168 non-overlapping 31×31 pixel ROIs were subsequently manually annotated in the HRCT slices by indicating the center position of the ROIs. The following three classes were considered: NT (59 observations), CLE (50 observations), and PSE (59 observations). The 168 labeled CT ROIs is the data set used throughout the evaluation, and this is also the data set used in [34].

3.3 Classification Setup

All texture descriptors are applied using the same classification setup, which consists of a supervised classifier, the k nearest neighbor classifier (kNN) [5, 9] where k is optimized on the training data. When a descriptor comprises several features or several feature histograms, an optimal subset is found by feature selection using the training data.

The kNN classifier is a flexible density-based classifier that makes no assumptions about the distribution of the data samples. Both unimodal and multimodal class distributions are automatically handled, as well as multiple classes as is the

case in the experiments carried out in this chapter. Further, both a vector space representation of the objects as well as a distance representation of the objects can easily be handled by kNN. This is practical in the experiments conducted, since two representations are considered for each texture descriptor, where possible: directly using the distances computed between full histograms, e.g., distances between full GLCMs, and classification in a feature vector space where the objects are represented by summarizing measures computed from the histograms, e.g., using (13), (14), (15), and (16) computed from the GLCMs.

An ROI $\mathbf{x}$ is classified by estimating the posterior probabilities $P(C_i|\mathbf{x})$ of the m different classes C_i, $i = 1, \ldots, m$, given that the input object is $\mathbf{x}$, using the moderated kNN estimate [11]

$$P(C|\mathbf{x}) = \frac{k_c(\mathbf{x}) + 1}{k + m} \tag{22}$$

where $k_c(\mathbf{x})$ is the number of neighbors of $\mathbf{x}$ belonging to class C among the k nearest neighbors. The parameter k determines how many prototypes are used in the estimation of $P(C|\mathbf{x})$, and k effectively determines the smoothness of the decision boundary of the classifier. A larger k leads to a more smooth decision boundary. The moderated kNN posterior estimate [11] is used since classifier combination [9] for combining the output of different descriptors will be investigated later in the chapter.

The kNN estimate in (22) produces a soft classification output. A hard classification, i.e., a specific labeling of objects is done by maximum a posteriori (MAP) classification [9], where an input object $\mathbf{x}$ is assigned to class C_i if

$$P(C_i|\mathbf{x}) > P(C_j|\mathbf{x}) \text{ for all } j \neq i. \tag{23}$$

The NT class is selected in case of a draw.

The distance between the objects used in kNN is computed differently depending on the representation.

When using texture descriptors based on summarizing measures, the "traditional" approach to pattern recognition is taken [6], and the kNN classifier is applied in a feature vector space where each ROI is a point, positioned according to its corresponding feature values. We use the L_2-norm between the feature vectors as object distances in the kNN classifier

$$d_m(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{24}$$

17

where ROI $\mathbf{x}$ is represented by a vector of n values $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$.

When a texture descriptor based on one or several histograms is used, classification can be based on direct distance computations between the histograms. This can be seen as a "featureless" representation where object distances are used directly in classification [6]. The L_1-norm is used as histogram distance measure, which is equivalent to the histogram intersection, a traditional histogram distance measure in computer vision [35]. When the texture descriptor is based on several histograms, the distances between the individual histograms are combined by summation, giving rise to the following combined histogram distance measure

$$d_h(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N_f} \sum_{j=1}^{N_b} |f_i(\mathbf{x})_j - f_i(\mathbf{y})_j| \qquad (25)$$

where $f_i(\cdot)$ is the ith histogram generating function, $f_i(\cdot)_j$ is the j'th bin of the i'th histogram, N_f is the number of histograms in the texture descriptor, and N_b is the number of histogram bins.

In cases where the descriptor consists of a set Y of several summarizing measures or histograms, these are optimized on the training data by feature selection. Here the goal is to find a subset $X \subseteq Y$ that is optimal according to some objective function. In this work, SFS [9], a greedy combinatorial optimization search strategy, is used with the classification accuracy of the kNN classifier using the current representation X^i as objective function. One starts with an empty set $X^0 = \emptyset$ and iteratively evaluates each of the candidate measures or histograms x in conjunction with the existing subset $x \cup X^{i-1}$, $x \in Y \setminus X^{i-1}$ according to the objective function. The best performing combination of current subset X^{i-1} and candidate feature $\hat{x}$ is retained producing $X^i = \hat{x} \cup X^{i-1}$. This procedure is iterated until adding a new feature starts increasing the classification error.

3.4   Training and Parameter Selection

There are several parameters to set in the classification system, some of which are fixed while the remaining are optimized on the training data. An ROI size of 31×31 pixels is used in all experiments for all texture descriptors. The number of histogram bins used when quantizing intensity in LBPs, GLCMs, GLRLMs, and INT, or when quantizing filter responses in the GFB is fixed according to the following rule: $N_b = \lfloor \sqrt[3]{N_p} \rfloor$ where N_p is the number of pixels in the ROI. In all experiments, we have $N_b = \lfloor \sqrt[3]{31^2} \rfloor = 9$ bins. In INT, GFB, and for the intensity part of LBPs, we employ an adaptive binning principle similar to that of [16]; the

Table 1: The texture descriptors together with their corresponding parameters, and the parameter values considered. The parameters are: (R, P) (LBP radius and number of samples pair, different pairs are considered for rotation invariant and uniform LBP indicated by ^{ri} and ^{riu2}.), σ (GFB scale), θ (orientation in GLCMs and GLRLMs), d (distance in GLCMs), MRL (maximum run length in GLRLMs), N_c (number of cluster centers per class in in textons), TS × TS (texton size), N_b (number of histogram bins, used both for quantizing image gray-levels in LBPs, GLCMs, GLRLMs, and INT, and for quantizing filter responses in the GFB). 'Number of features' is the number of candidate histograms or summarizing measures in the descriptor. An optimal subset of these is found using SFS. 'Feature dimension' is the dimensionality of one candidate "feature" considered in SFS. For histograms, this is the number of bins. For summarizing measures, this is one. When both full histograms and summarizing measures are considered for a particular descriptor, this is indicated by $^{h/m}$, and the corresponding number of features and feature dimension(s) are shown for both representations separated by a forward slash.

Descriptor	Parameters	Number of features	Feature dimension
LBP	$(R, P)^{ri} = \{(1, 8), (2, 8), (3, 8)\}$ $(R, P)^{riu2} = \{(1, 8), (2, 16), (3, 24), (4, 24), (5, 24)\}$ $N_b = 9$	8	$\{36 \times 9, 10 \times 9, 18 \times 9, 26 \times 9\}$
$GFB^{h/m}$	$\sigma = \{0.5, 1, 2, 4\}$ $N_b = 9$	16 / 16 × 4	9 / 1
$GLCM^{h/m}$	$\theta = \{0, 45, 90, 135\}$ (averaged) $d = \{1, 2, 3, 4, 5\}$ $N_b = 9$	5 / 5 × 4	9 × 9 / 1
$GLRLM^{h/m}$	$\theta = \{0, 45, 90, 135\}$ (averaged) $MRL = 10$ $N_b = 9$	1 / 5	9 × 10 / 1
$INT^{h/m}$	$N_b = 9$	1 / 4	9 / 1
$TEXTONS$	$N_c = \{10, 20, 30, 40, 50\}$ TS × TS = $\{3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9\}$	5 × 4	$\{30, 60, 90, 120, 150\}$

total feature distribution across all ROIs in the training set is made approximately uniform. Consequently, densely populated areas in feature space are quantized with a high resolution while sparse areas are quantized with a low resolution. In INT and GFB, the summarizing measures are computed from the original CT intensities and filter responses, respectively. In GLCMs and GLRLMs, the intensities are traditionally quantized using a fixed bin width, and we therefore use this scheme for these descriptors. For TEXTONS no quantization of the intensities is performed. All histograms are normalized to sum to one, and all summarizing measures are standardized to unit variance.

k in the kNN classifier is optimized using cross-validation, where the following k's are considered: $k = 1, 3, \ldots, \lfloor \sqrt{n} \rfloor$ where n is the number of prototypes in the kNN classifier and the upper limit on k is depicted by the usual square root rule [11].

Table 1 lists the texture descriptors together with the descriptor specific parameter values that are considered in the experiments. The texture descriptors described in Sections 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6, which we term INT, LBP, GFB, $GLCM$, $GLRLM$, and $TEXTONS$, respectively, are applied. Where applicable, full histograms as well as summarizing measures are used, indicated by ^{h} and ^{m}. LBPs and textons are applied only as histograms. In $GLCM$ and $GLRLM$, both the full matrices and the summarizing measures are averaged across the four orientations $\theta = \{0, 45, 90, 135\}$ as is often done in the literature. However, we do not average across distances in $GLCM$. In LBP, we consider $LBP2^{ri}$ and $LBP2^{riu2}$ with different pairs of parameters (R, P).

The performance is estimated using leave-one-out error estimation at subject level. In each of the 25 leave-out trials, all ROIs from the test subject are held out and the classification system using the different texture descriptors is trained on the remaining data. The remaining subjects are separated into a training set and a validation set. In this separation, balanced class distributions are ensured by placing half the subjects from each class in the training set and the rest in the validation set. k in the kNN classifier is optimized using cross-validation with the ROIs in the training set as prototypes, and the classification accuracy on the validation set is used as performance criterion. SFS is applied within this cross-validation procedure to find an optimal subset in the texture descriptor, either a set of histograms or of summarizing measures, for each value of k considered. Subsequently, the ROIs in the test set are classified with the training set ROIs as prototypes using the optimal k, and the optimal histogram or measure subset obtained using SFS. Note, that INT^h and $GLRLM^h$ only consist of one histogram, and SFS is not applied in these

20

Table 2: Results of the leave-one-out error estimation at subject level. CA: classification accuracy. The left part shows the results of using full histograms, and the right part shows the results of using summarizing measures.

Descriptor	CA	Descriptor	CA
LBP	92.9	GFB^m	75.6
GFB^h	95.2	$GLCM^m$	92.9
$GLCM^h$	92.3	$GLRLM^m$	86.9
$GLRLM^h$	89.9	INT^m	78.0
INT^h	91.1		
$TEXTONS$	92.3		

cases.

3.5 Classification Results

The texture descriptors are evaluated by MAP classification accuracy (CA), and the performance of the different descriptors is reported in Table 2, and the p-values of comparing the MAP classification outputs of the different descriptors are reported in Table 3. We also inspect the CA as a function of the size of the training set. This is done by varying the number of subjects in the training set, and for each size, do 10 repeated leave-one-subject-out error estimations with a randomly drawn training set of that particular size. As previous, the classes are balanced in the data splits, and the same randomly drawn training sets are used for all descriptors. The whole learning framework is applied in the same way as described previously, including optimization of k and feature selection on the training set, for each of the 10 repeated leave-one-out trials. The average curves $\pm$ the standard deviation across the 10 trials are shown in Figure 9.

3.6 Selected Parameters

The number of times a particular k is selected in the leave-one-out estimation is reported in Table 4 for the different descriptors. Further, the histograms or measures that are most often selected in SFS in the leave-one-out estimation are reported in Table 5 for the different descriptors. Note, that this does not reflect which "features" are in the same subsets in the different leave-out trials, only how often a "feature" is selected independently. Nothing is reported for INT^h and $GLRLM^h$, as these descriptors only comprise one histogram.

Table 3: p-values of comparing the MAP classification outputs of the different texture descriptors using McNemar's test [4]. Significant p-values, i.e., $p < 0.05$, are highlighted in bold font.

	LBP	GFB^h	$GLCM^h$	$GLRLM^h$	INT^h	$TEXTONS$	GFB^m	$GLCM^m$	$GLRLM^m$	INT^m
LBP	-	-	-	-	-	-	-	-	-	-
GFB^h	0.386	-	-	-	-	-	-	-	-	-
$GLCM^h$	1.000	0.332	-	-	-	-	-	-	-	-
$GLRLM^h$	0.359	0.066	0.502	-	-	-	-	-	-	-
INT^h	0.546	0.146	0.773	0.831	-	-	-	-	-	-
$TEXTONS$	1.000	0.228	0.773	0.522	0.752	-	-	-	-	-
GFB^m	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	-	-	-	-
$GLCM^m$	0.814	0.423	1.000	0.404	0.646	1.000	$\mathbf{< 10^{-4}}$	-	-	-
$GLRLM^m$	0.078	**0.006**	0.124	0.458	0.265	0.124	**0.002**	**0.044**	-	-
INT^m	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	**0.002**	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	0.596	$\mathbf{< 10^{-4}}$	**0.021**	-

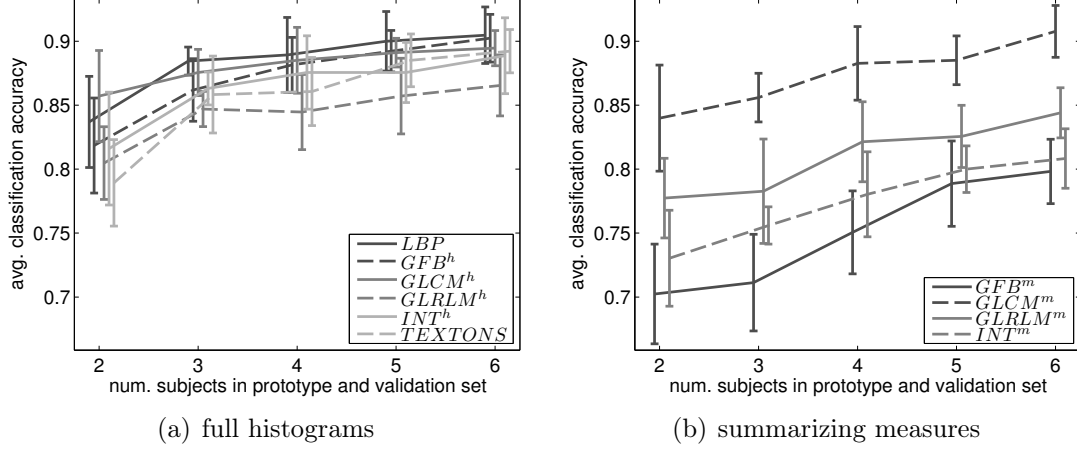(a) full histograms (b) summarizing measures

Figure 9: Learning curves for the different texture descriptors. The average $\pm$ the standard deviation of the CA from 10 random trials is reported.

Table 4: Optimized ks in the kNN classifier for the different descriptors in the leave-one-out experiments at subject level. 100% means that the value was selected in all 25 leave-out trials.

Descriptor	Value of k in kNN
LBP	1 (100%)
GFB^h	1 (80%) 3 (12%) 5 (8%)
$GLCM^h$	1 (92%) 5 (4%) 7 (4%)
$GLRLM^h$	5 (72%) 1 (28%)
INT^h	1 (68%) 3 (20%) 5 (8%) 7 (4%)
$TEXTONS$	1 (100%)
GFB^m	9 (44%) 1 (40%) 3 (12%) 5 (4%)
$GLCM^m$	1 (68%) 3 (24%) 9 (8%)
$GLRLM^m$	1 (72%) 3 (20%) 5 (4%) 9 (4%)
INT^m	1 (44%) 5 (24%) 3 (16%) 7 (8%) 9 (8%)

Table 5: The four most frequently selected histograms or summarizing measures in the SFS procedure for the different texture descriptors. Individual occurrences are reported, but the histograms / measures are usually part of a larger subset. 100% means that the histogram or measure was selected in all 25 leave-out trials.

Descriptor	Features			
LBP	$LBP2^{ri}(\cdot;1,8)$ (64%)	$LBP2^{riu2}(\cdot;4,24)$ (36%)	$LBP2^{ri}(\cdot;2,8)$ (28%)	$LBP2^{riu2}(\cdot;1,8)$ (28%)
GFB^h	$G(\cdot,0.5)$ (100%)	$||\nabla G(\cdot,1)||_2$ (76%)	$||\nabla G(\cdot,0.5)||_2$ (40%)	$K(\cdot,0.5)$ (16%)
$GLCM^h$	$p(\cdot,\cdot;2)$ (76%)	$p(\cdot,\cdot;4)$ (12%)	$p(\cdot,\cdot;3)$ (8%)	$p(\cdot,\cdot;5)$ (4%)
$TEXTONS$	$\{TS=3,N_c=30\}$ (48%)	$\{TS=3,N_c=20\}$ (44%)	$\{TS=5,N_c=30\}$ (40%)	$\{TS=5,N_c=20\}$ (24%)
GFB^m	$\mu(G(\cdot,0.5))$ (68%)	$\mu(G(\cdot,1))$ (44%)	$\mu(G(\cdot,4))$ (44%)	$skew(G(\cdot,0.5))$ (32%)
$GLCM^m$	$CORR(\cdot;1)$ (52%)	$UE(\cdot;1)$ (44%)	$CORR(\cdot;3)$ (44%)	$C(\cdot;1)$ (28%)
$GLRLM^m$	$RLN(\cdot)$ (100%)	$RP(\cdot)$ (96%)	$LRE(\cdot)$ (80%)	$GLN(\cdot)$ (72%)
INT^m	$\mu(\cdot)$ (100%)	$skew(\cdot)$ (60%)	$\sigma(\cdot)$ (36%)	$kurtosis(\cdot)$ (8%)

Table 6: Results of classifier combination. CA: classification accuracy.

Descriptor set	Combiner	CA
histograms	majority voting	94.6
	mean	94.6
	product	94.6
summarizing measures	majority voting	86.3
	mean	87.5
	product	86.9
histograms + summarizing measures	majority voting	95.2
	mean	95.8
	product	95.8

3.7 Combining Information

Several different texture descriptors have been evaluated, giving rise to several different classifiers, one for each descriptor. These different classifiers may capture different, but complementary, information in the ROIs. This motivates the idea of using classifier combination [9] for combining the information captured by the different descriptors into an overall classification. The outputs of the different classifiers are combined using a fixed combination rule, or combiner. This classifier combination scheme, where all the individual classifiers are invoked independently, is the most commonly used in the literature [9]. Three combiners are considered [9, 12]: the majority voting rule, the mean rule, and the product rule. In majority voting, MAP classification (23) is applied prior to voting. The mean and the product rule are both applied directly on the posteriors. Three setups are considered, one using only the histogram-based descriptors, i.e., LBP, GFB^h, $GLCM^h$, $GLRLM^h$, INT^h, and $TEXTONS$; one using only the summarizing measure-based descriptors, i.e, GFB^m, $GLCM^m$, $GLRLM^m$, INT^m; and one using all descriptors. Table 6 lists the resulting classification accuracies. Combining descriptors slightly increases performance in two cases, the product rule and the majority voting rule applied to all descriptors achieves CA = 95.8%, which means that one more sample is correctly classified than with the best performing individual descriptor, GFB^h that achieved CA = 95.2%. However, these differences are not significant.

We also experimented with an alternative way of combining information from different descriptors by considering all histograms or measures collectively in the classification system. In this case, all descriptors are candidates in the feature selection. Using full histograms achieved CA = 92.9 and using summarizing measures achieved CA = 86.9. Further, a combination of $GLCM^{h/m}$, $GLRLM^{h/m}$, and $INT^{h/m}$ was experimented with. This represents a descriptor that is used in sev-

eral studies in the literature [2, 18, 20, 21, 38, 45]. This resulted in CA = 93.5 and CA = 87.5, for histograms and summarizing measures, respectively.

4   Discussion and Conclusion

This chapter reviewed several popular texture descriptors in the literature for tissue classification in pulmonary CT, with a focus on emphysema classification.

Most of the evaluated descriptors performed well. The considered data set is small and there were therefore few significant differences between the descriptors in general and none between that descriptors that performed well. The results suggest for using full histogram approaches, i.e., LPB^h, GFB^h, $GLCM^h$, and $TEXTONS$ performed similar and slightly better than $GLRLM$ and INT while significantly better than using moments of distributions as was used in GFB^m and INT^m. No descriptor stood out as the single best one. The GFB using full filter response histograms, GFB^h, achieved the highest accuracy, CA = 95.2%, see Table 2, however, this was not significantly different from any of the competing full histogram descriptors, see Table 3. The data used here is that same as is used in [34], but the results are slightly different due to differences in the classification setup. In [34], the ROI size was also optimized on the training data, for example. However, the GFB using full histograms and LBPs both performed well and were not significantly different according to the results in this chapter, and this was also the result in [34].

Two descriptors that are still commonly used, GFB^m and INT^m, performed significantly worse than all the remaining descriptors , see Table 2 and Table 3. The two descriptors were not significantly different from each other, which is explained by the fact that the four most often selected summarizing measures in GFB^m are based on blurred intensities, see Table 5, i.e., the two descriptors capture similar information.

Summarizing measures, which are heavily used in the literature, did not work well for all descriptors. The results of the conducted experiment show that using moments instead of full histograms for INT and GFB discards important discriminative information for the investigated classification task, as indicated by the significant degradation in performance when comparing INT^h to INT^m and GFB^h to GFB^m, see Table 2 and Table 3. For the classical texture descriptors, GLCMs and GLRLMs, the common approach of using summarizing measures, instead of full matrices, works well.

The data set used in the experiment is small, and as can be seen from the

learning curves in Figure 9, all descriptors could benefit from more training data, i.e., non of the curves converge. Some descriptors perform well already at small data sizes and these descriptors are: LBP, $GLCM^h$, and $GLCM^m$. However, descriptors such as GFB^h and $TEXTONS$ approach a good performance as the data size increases, and may prove even better for more data than is available in this experiment. Allowing many options in the feature selection, such as in GFB where many filters a considered, also increases the demand on the training data size in order to avoid overtraining. This may also be reflected in the learning curves.

The descriptors should also be judged based on other criteria than classification performance. One criterion is the complexity of the descriptor. Histogram-based descriptors are generally more rich in information at the expense of increased complexity, while summarizing measures reduce the dimensionality, and hence the complexity. The crucial point here is whether important discriminative information is lost in this process, and this was clearly the case for some descriptors in the conducted experiments. The complexity of the different descriptors can be seen in Table 1 where the dimensionality of the individual components of the descriptors, i.e., number of bins in the individual histograms for the histogram-based descriptors, is listed in the fourth column. In general, LBP is the most complex descriptor followed by $TEXTONS$. This should also be judged in the light of how many and which "features" are selected in SFS, and as can be seen in Table 5, $LBP2^{ri}(\cdot; 1, 8)$ that has 36×9 bins is frequently selected.

The data considered in this chapter is two-dimensional, and hence the focus was on 2-D descriptors. However, since volumetric CT images are becoming state-of-the-art, it is important to consider whether the descriptors are applicable in 3-D. INT and GFB can be directly, and have already been [30, 31], extended to 3-D without any increase in complexity. Note, however, that more samples are available for estimating histograms and that the number of histogram bins therefore can be increased. GLCMs and GLRLMs have also already been extended and applied in 3-D [45]. In this case, the number of orientations, and thereby matrices, increases. However, when averaging across orientations the complexity stays the same. As for GFB and INT, the intensity can be quantized into more bins due to more available samples for histogram estimation. It is not straightforward how to extend LBPs to 3-D. Two approximate extensions of LBP to 3-D are presented in [46], with the specific application being temporal texture data in computer vision. The first approach forms a helical path in the temporal direction. This idea could be applied in volumetric CT by, e.g., forming helical paths in various directions and combining the resulting LBPs. The second approach in [46] computes 2-D LBPs in

three orthogonal planes and concatenates these. Other studies consider samples on a sphere centered on the voxel of interest [14, 19]. In these cases there is no natural ordering of the samples as is the case when considering a circle. The complexity increases in all approaches. The texton approach considered in this chapter uses a raw pixel representation, and increasing the dimensionality to 3-D has the effect of increasing the feature space where both the clustering and the texton codebook similarities are computed. This needs not increase the complexity of the descriptor, i.e., the number of cluster centers, but it puts increasing demand on more training data in order to reliably estimate cluster centers.

It may seem somewhat surprising that the combination of $GLCM^m$, $GLRLM^m$, and INT^m, often used in the literature [2, 18, 20, 21, 38, 45], performed worse than $GLCM^m$ alone. This is because SFS is a greedy combinatorial search procedure that can get stuck in local optima, in this case due to elements from the descriptors INT^m or $GLRLM^m$ being selected early in the optimization process. Care should therefore be taken when using a feature selection scheme. The same argument holds for the results obtained when performing feature selection on the combined set of all descriptors, which also performed worse than several of the individual descriptors. Combining the texture descriptors by classifier combination in a parallel scheme using a fixed combiner, improved results slightly, with CA = 95.8, when all descriptors, both histograms and summarizing measures were considered in conjunction with the product or the majority voting rule, see Table 6. However, this was not significantly better than the best performing individual descriptor, GFB^h. We expect classifier combination to be more beneficial in problems with more complicated and diverse texture classes than the ones encountered when classifying emphysema in CT.

In conclusion, several popular texture descriptors in the medical image analysis literature for texture-based emphysema classification in pulmonary CT were reviewed and evaluated within the same classification framework. The evaluation was performed on an emphysema CT ROI classification task. No descriptor was significantly better than all other descriptors. In general, full histogram representations performed better than moments of histograms, and intensity and run length features alone seemed to perform less good than the more complex texture descriptors. There were indications that combining the classification output of several classifiers, each based on a different texture descriptor, is beneficial. However, the increase in performance was not significant for the considered classification task.

28

References

[1] P. J. Barnes. Chronic obstructive pulmonary disease. *The New England Journal of Medicine*, 343(4):269–280, 2000.

[2] F. Chabat, G.-Z. Yang, and D. M. Hansell. Obstructive lung diseases: texture classification for differentiation at CT. *Radiology*, 228(3):871–877, 2003.

[3] O. G. Cula and K. J. Dana. 3D texture recognition using bidirectional feature histograms. *International Journal of Computer Vision*, 59(1):33–60, 2004.

[4] T. G. Dietterich. Approximate statistical test for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923, 1998.

[5] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.

[6] R. P. W. Duin, D. de Ridder, and D. M. J. Tax. Featureless pattern classification. *Kybernetika*, 34(4):399–404, 1998.

[7] M. J. Gangeh, L. Sørensen, S. B. Shaker, M. S. Kamel, M. de Bruijne, and M. Loog. A texton-based approach for the classification of lung parenchyma in CT images. In T. Jiang, N. Navab, J. P. W. Pluim, and M. A. Viergever, editors, *Medical Image Computing and Computer Assisted Intervention*, volume 6363 of *Lecture Notes in Computer Science*, pages 596–603. Springer, 2010.

[8] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

[9] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: a review. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[10] B. Juelsz. Textons, the elements of texture-perception, and their interactions. *Nature*, 290(5802):91–97, 1981.

[11] J. Kittler and F. M. Alkoot. Moderating k-NN classifiers. *Pattern Analysis and Applications*, 5(3):326–332, 2002.

[12] Ludmila I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.

[13] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[14] S. Liao and A. C. S. Chung. Non-rigid image registration with uniform spherical structure patterns. In J. L. Prince, D. L. Pham, and K. J. Myers, editors,

Information Processing in Medical Imaging, volume 5636 of *Lecture Notes in Computer Science*, pages 163–175. Springer, 2009.

[15] M. Mirmehdiand, X. Xie, and J. Suri, editors. *Handbook of Texture Analysis.* Imperial College Press, 2008.

[16] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[17] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[18] Y. S. Park, J. B. Seo, N. Kim, E. J. Chae, Y. M. Oh, S. D. Lee, Y. Lee, and S.-H. Kang. Texture-based quantification of pulmonary emphysema on high-resolution computed tomography: Comparison with density-based quantification and correlation with pulmonary function test. *Investigative Radiology*, 43(6):395–402, 2008.

[19] L. Paulhac, P. Makris, and Y.-J. Ramel. Comparison between 2D and 3D local binary pattern methods for characterisation of three-dimensional textures. In A. C. Campilho and M. S. Kamel, editors, *Image Analysis and Recognition*, volume 5112 of *Lecture Notes in Computer Science*, pages 670–679. Springer, 2008.

[20] M. Prasad, A. Sowmya, and I. Koch. Designing relevant features for continuous data sets using ICA. *International Journal of Computational Intelligence and Applications*, 7(4):447–468, 2008.

[21] M. Prasad, A. Sowmya, and P. Wilson. Multi-level classification of emphysema in HRCT lung images. *Pattern Analysis and Applications*, 12(1):9–20, 2009.

[22] K. F. Rabe, S. Hurd, A. Anzueto, P. J. Barnes, S. A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. van Weel, J. Zielinski, and Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *American Journal of Respiratory and Critical Care Medicine*, 176(6):532–555, 2007.

[23] C. Schmid. Constructing models for content-based image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 39–45. IEEE Computer Society Press, 2001.

[24] C. Schmid. Weakly supervised learning of visual models and its application to content-based retrieval. *International Journal of Computer Vision*, 56(1-2, Sp. Iss. SI):7–16, 2004.

[25] S. B. Shaker, A. Dirksen, K. S. Bach, and J. Mortensen. Imaging in chronic obstructive pulmonary disease. *COPD*, 4(2):143–161, 2007.

[26] S. B. Shaker, K. A. von Wachenfeldt, S. Larsson, I. Mile, S. Persdotter, M. Dahlbäck, P. Broberg, B. Stoel, K. S. Bach, M. Hestad, T. E. Fehniger, and A. Dirksen. Identification of patients with chronic obstructive pulmonary disease (COPD) by measurement of plasma biomarkers. *The Clinical Respiratory Journal*, 2(1):17–25, 2008.

[27] I. C. Sluimer, M. Prokop, I. Hartmann, and B. van Ginneken. Automated classification of hyperlucency, fibrosis, ground glass, solid and focal lesions in high resolution CT of the lung. *Medical Physics*, 33(7):2610–2620, 2006.

[28] I. C. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Transactions on Medical Imaging*, 25(4):385–405, 2006.

[29] I. C. Sluimer, P. F. van Waes, M. A. Viergever, and B. van Ginneken. Computer-aided diagnosis in high resolution CT of the lungs. *Medical Physics*, 30(12):3081–3090, 2003.

[30] L. Sørensen, P. Lo, H. Ashraf, J. Sporring, M. Nielsen, and M. de Bruijne. Learning COPD sensitive filters in pulmonary CT. In G.-Z. Yang, D. J. Hawkes, D. Rueckert, J. A. Noble, and C. J. Taylor, editors, *Medical Image Computing and Computer Assisted Intervention*, volume 5761 of *Lecture Notes in Computer Science*, pages 699–706. Springer, 2009.

[31] L. Sørensen, M. Loog, P. Lo, H. Ashraf, A. Dirksen, R. P. W. Duin, and M. de Bruijne. Image dissimilarity-based quantification of lung disease from CT. In T. Jiang, N. Navab, J. P. W. Pluim, and M. A. Viergever, editors, *Medical Image Computing and Computer Assisted Intervention*, volume 6361 of *Lecture Notes in Computer Science*. Springer, 2010.

[32] L. Sørensen, S. B. Shaker, and M. de Bruijne. Texture based emphysema quantification in lung CT. In M. Brown, M. de Bruijne, B. van Ginneken, A. Kiraly, J.-M. Kuhnigk, C. Lorenz, K. Mori, and J. Reinhardt, editors, *Proc. of The First International Workshop on Pulmonary Image Analysis*, pages 5–14, 2008.

[33] L Sørensen, S. B. Shaker, and M de Bruijne. Texture classification in lung CT using local binary patterns. In D. N. Metaxas, L. Axel, G. Fichtinger, and

G. Székely, editors, *Medical Image Computing and Computer Assisted Intervention*, volume 5241 of *Lecture Notes in Computer Science*, pages 934–941. Springer, 2008.

[34] L. Sørensen, S. B. Shaker, and M. de Bruijne. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Transactions on Medical Imaging*, 29(2):559–569, 2010.

[35] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[36] B. M. ter Haar Romeny. *Gaussian Scale-Space Theory*, chapter Applications of scale-space theory, pages 3–19. Kluwer Academic Publishers, 1997.

[37] M. Tuceryan and A. K. Jain. *The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*, chapter Texture Analysis, pages 207–248. World Scientific Publishing, 1998.

[38] R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W. Hunninghake, and G. McLennan. Computer recognition of regional lung disease patterns. *American Journal of Respiratory and Critical Care Medicine*, 160(2):648–654, 1999.

[39] R. Uppaluri, T. Mitsa, M. Sonka, E. A. Hoffman, and G. McLennan. Quantification of pulmonary emphysema from lung computed tomography images. *American Journal of Respiratory and Critical Care Medicine*, 156(1):248–254, 1997.

[40] Bram van Ginneken and Bart M. ter Haar Romeny. Multi-scale texture classification from generalized locally orderless images. *Pattern Recognition*, 36(4):899–911, 2003.

[41] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 691–698. IEEE Computer Society, 2003.

[42] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1-2):61–81, 2005.

[43] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 31(11):2032–2047, 2009.

[44] W. R. Webb, N. L. Müller, and D. P. Naidich. *High-Resolution CT of the Lung*. Lippincott Williams & Wilkins, third edition edition, 2001.

[45] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman. MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. *IEEE Transactions on Medical Imaging*, 25(4):464–475, 2006.

[46] G. Zhao and M. Pietikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 29(6):915–928, 2007.

[47] Song C. Zhu, Yingnian Wu, and David Mumford. Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. *International Journal of Computer Vision*, 27(2):107–126, 1998.