Micro-Structural Tissue Analysis for Automatic Histopathological Image Annotation

GLORIA DÍAZ, AND EDUARDO ROMERO*
Bioingenium Research Group, Faculty of Medicine, National University of Colombia, Bogotá, Colombia

KEY WORDS semantic annotation; histopathology; part-based representation; latent semantic analysis; local descriptors

ABSTRACT This article presents a new approach for extracting high level semantic concepts from digital histopathological images. This strategy provides not only annotation of several biological concepts, but also a coarse location of these concepts. The proposed approach is composed of five main steps: (1) a stain decomposition stage, which separates the contribution of hematoxylin and eosin dyes, (2) a color standardization that corrects color image differences, (3) a part-based representation, which describes the image in terms of the conditional probability of relevant local patches, selected by their stain contributions, (4) a discriminative classification model, which bridges out the found patterns and the biological concepts, (5) a block-based annotation strategy that identifies the multiple biological concepts within an image. A set of 655 skin images, containing 10 biological concepts of skin tissues were used for assessing the proposed approach, obtaining a sensitivity of 84% and a specificity of 67% when annotating images with multiple concepts.

INTRODUCTION

Microscopic pathology is by far the most important para-clinic support for diagnosis, prognosis, and therapy of several diseases. This is the only technique that so far allows examination of both the underlying tissue and cell architectures, not visible under any other medical image modality (Zhu et al., 2006). The art of microscopic examination is a complex combination of different skills that physicians develop along their training periods, during which the know-how is achieved by exposing students to the analysis of as many cases as possible. Overall, actual medical activity is based upon agreement: differences that depend on the cut procedure. This amalgam of factors results in a large group of concepts that are at the very base of a diagnosis. We introduce herein a novel part-based representation approach, which captures the hidden semantic tissue characteristics, based on the analysis of their constituent stain features. Annotation of multiple concepts is approached by a simple block partition strategy that determines the concepts in every image location, using the proposed image representation scheme and discriminative learning models. In addition, a novel color standardization process for correcting the image color differences and reducing the within-concept variability is also presented. The method was evaluated on a very challenging annotation task, i.e., annotation of digital images of skin biopsies.

The article is organized as follows: Section “Previous Work,” presents a summarized review of the previous pathology CADs, consists in building an effective model that extracts information with semantic meaning of the visual microscopical content of the histological image. The problem is that the required level of semantic description is really hard to achieve in such images.

Biological tissues are composed of several colorless complex components that are highlighted by using dyes (Kiernan, 2001). The most common coloration in histopathology is the hematoxylin and eosin (H and E), a well known technique since hundred years ago, which reveals the very complicated structure of cellular and extracellular elements that compose each tissue, complex mixes of low level visual features, namely, shapes adapted to the particular organ function, colors associated to the chemical tissue composition, and orientations that depend on the cut procedure. This amalgam of factors results in a large group of concepts that are at the very base of a diagnosis. We introduce herein a novel part-based representation approach, which captures the hidden semantic tissue characteristics, based on the analysis of their constituent stain features. Annotation of multiple concepts is approached by a simple block partition strategy that determines the concepts in every image location, using the proposed image representation scheme and discriminative learning models. In addition, a novel color standardization process for correcting the image color differences and reducing the within-concept variability is also presented. The method was evaluated on a very challenging annotation task, i.e., annotation of digital images of skin biopsies.

The article is organized as follows: Section “Previous Work,” presents a summarized review of the previous
work, related to the semantic annotation of histopathological images. An overall description of the proposed approach and the evaluation methodology are presented in "Materials and Methods." Experimental results are reported in the next section. Finally, the discussion is carried out.

PREVIOUS WORK

Automatic interpretation of histopathological images is a new research field in the medical image domain and computer vision area (Gurcan et al., 2009) that aims to provide tools that support or assist the pathologist decisions. Research endeavors have been dedicated so far to develop two kinds of tools: (1) Automatic or semiautomatic systems for diagnosis of specific pathologies (Kayser et al., 2002; Llobet et al., 2007; Zhu et al., 2006), attempting to achieve a pre-classification of tissues as malignant or benign (Boucheron et al., 2010; Doyle et al., 2003) or of the severity of a disease grading (Ahammer et al., 2009; Huang et al., 2010; Jondet et al., 2010; Kayser et al., 2008; Kong et al., 2009) and (2) development of content based image retrieval (CBIR) systems (Caicedo et al., 2008; Comaniciu et al., 1999; Naik et al., 2009; Zheng, 2005), defined as systems capable of retrieving the most similar (different) images from the entire collection of images. The decision process is thus supported using the semantic information contained in the database. Both strategies are based on the same framework to construct semantic information, i.e., image dimensionality is reduced to smaller vectors, known as descriptors, and some type of metrics is then set to obtain a notion of distance between these image descriptors.

Overall, descriptors are low level features, devised to emulate the diagnosis basis of a particular disease (Gurcan et al., 2009). Provided the large number of disease variations and diagnosis procedures, it results impossible to use a unique descriptor for every possible illness (Sertel et al., 2009). Consequently, several feature descriptors have been proposed, falling into one of the next three categories: object-based, architectural, or global features. Object-based features (Huang et al., 2010; Jondet et al., 2010; Kong et al., 2009), predominantly nuclear and glandular morphometric features, have been used when the searched concepts are associated to specific histological microstructures. Likewise, architectural features characterize a biological tissue by the centroids that generate the graph statistics associated to the spatial arrangement of the microstructures (Doyle et al., 2008; Jondet et al., 2010; Kong et al., 2009; Naik et al., 2009). In both cases, classification is straightforwardly related to the segmentation accuracy of the particular tissue components, a very difficult job in routine histopathological images. Finally, global features, such as texture or color descriptors (Ahammer et al., 2009; Caicedo et al., 2008; Kayser et al., 2008; Oger et al., 2009) have shown to be successful for classifying well differentiated tissues, for instance several degrees of tumoral tissues (Ahammer et al., 2009; Al-Radi, 2010) or benign and malignant breast cancer tissues (Oger et al., 2009). However, these descriptors fail to differentiate complex biological structures that exhibit superimposed features (Caicedo and Izquierdo, 2010). The recognition of semantic concepts that show such mixture of features has been approached through two main strategies: combining low level features, as an attempt to provide the learning model with more discriminative information (Caicedo and Izquierdo, 2010; Lessmann et al., 2007; Sertel et al., 2008), or assembling multiple classifiers, each specialized in separating a small set of concepts (Glotsos et al., 2008; Huang and Lai, 2010). Additionally, the combination of object, texture, and structure based features has been proposed by Kayser and Kayser (2005) and Kayser et al. (2006, 2008) for classifying lung and breast cancer images.

Recently, some methods that extract semantic information by representing visual primitives and latent semantic concepts in the image have been successfully applied in several image classification problems (Andrie et al., 2009; Jiang et al., 2007; Nowak et al., 2006; Yang et al., 2009). These approaches are inspired by the fact that the visual system perceives an object by integrating its constituent parts (Biederman, 1987; Olshausen, 2003). It turns out that the human brain is able to probabilistically infer the presence of learned objects, based on their global geometry, their parts and their structural relations (Biederman, 1987). Ultimately, it has been demonstrated that many algorithms improve their performance by emulating what the visual system naturally does, i.e., recognition of the constituent parts (local patches) which are then used to approaching complex tasks. Several strategies have been reported in the literature for extracting local patches, including dense regular image partitions (Jurie and Triggs, 2005; Winn et al., 2005), coarse segmented regions (Chen et al., 2006), and local keypoint detectors (Nowak et al., 2006). In general, dense grid samplers have shown to be more effective for classification tasks (Jurie and Triggs, 2005; Nowak et al., 2006). However, their efficiency has been questioned because they need a large number of local patches for representing each image (Nowak et al., 2006). The use of conventional part extraction approaches has been explored in histological (Caicedo and Izquierdo, 2010) and histopathological images (Caicedo et al., 2009; Cruz-Roa et al., 2009; Mehta et al., 2009), but with no specific interest on searching visual and structural features.

Overall, histopathological images are composed of randomly arranged structures, bounded together by the connective tissue. Cells are organized in structures that perform a particular function, whereas the rest, the stroma, is the support tissue that contains the connective tissue and the cells that produce it. Therefore, they are always annotated with more than two concepts: the biological structures and the stroma. Different investigations for approaching such multiclass problem have been reported in the literature. In histopathological images, Caicedo et al. (2008) and Cruz-Roa et al. (2009) proposed to train a bank of learning models, each model specialized in deciding whether or not an image contained one of the searched concepts. These approaches report a very poor annotation performance (a sensitivity below 0.23, and an effectiveness below 0.25). The main drawback with these approaches is that they base their analysis on a global frequentist strategy, either using global (Caicedo et al., 2008) or local (Cruz-Roa et al., 2009) features, thereby missing small or single concepts. In contrast, Tang et al. proposed a block-based strategy for extracting different
levels of semantic concepts from images of the gastrointestinal tract (GI; Tang et al., 2003). Instead of attempting to extract a unique concept from the whole image, they tried to describe each component inside using three classification levels: coarse, semi-fine, and fine. The coarse level uses a three-layer neural network that constructs normalized color and gray-level histograms for classifying blocks of 64 × 64 pixels into one of 10 categories, corresponding to the different general tissues (mucosa, submucosa, extern muscularis, and serosa) and their histological junctions. The semi-fine detector introduces a texture measurement using Gabor features, for distinguishing more than 70 categories from the gastrointestinal tractus, corresponding to specific organ tissues and their different histological junctions. Finally, specialized fine detectors are constructed using morphological priors. The main advantage with the latter strategy is that it allows not only detection of multiple concepts, but also identification of their particular location. The drawback of this approach is that it requires extra work to obtain the training subimages for every possible concept. In other applications, object-oriented approaches such as rough segmentation (Athanasiadis et al., 2007; Deruyver et al., 2009) or window sliding (Felzenszwalb et al., 2010), have been proposed. These methods try to detect complex objects in the image, a very difficult task in histological images since object boundaries are really difficult to establish, even for experts.

MATERIALS AND METHODS

Method Overview

An overview of the proposed approach is illustrated in Figure 1. Two main phases are well identified: learning (up) and test (down). Arrows represent the flow of data between the processing steps. In the learning phase, represented by the continuous line, a set of cropped images is used for training a classification model, under the restriction that each cropped image contains mainly a single concept. Provided that the feature space was dominated by the contribution of two main semantic tissue components, i.e., the basophilic (stained by the hematoxylin) and eosinophilic (stained by the eosin) microstructures, the image processing starts by separating them, under the restriction that each cropped image contains mainly a single concept. Provided that the feature space was dominated by the contribution of two main semantic tissue components, i.e., the basophilic (stained by the hematoxylin) and eosinophilic (stained by the eosin) microstructures, the image processing starts by separating them, under the restriction that any contribution should be always non negative. Afterward, local patches are extracted as the regions around each basophilic structure detected on the hematoxylin component. A feature vector describing the distribution of both hematoxylin and eosin dyes is then computed.
for any local patch. These feature vectors are clustered for defining a visual vocabulary of patch prototypes, which is used for characterizing each image as the conditional probability distribution of a set of hidden semantic topics. These topics correspond to simple occurrences of a statistical model whose dynamics is behind our observations: image patches with semantic meaning. Finally, the hidden topic probability distribution of the training set is used for learning a classification model, which groups up these topics into concepts, previously defined by an expert pathologist.

Once the training phase is achieved, images are annotated using a block-based classification strategy. The input image is split into a set of overlapped blocks. Each block is then characterized and classified using the combination of the part-based representation and the classification model. Finally, each overlapped area is annotated as the most frequent concept found in the blocks intersecting that area.

**Stain Decomposition and Normalization**

This stage estimates each dye contribution to the final stained tissue, aiming to define an objective parameter to compare micro-structures. The main challenge of this task consists in dealing with the colocalization problem, a very common phenomenon produced by the chemical reactions of the target proteins with more than one specific dye, or by the co-existence of structures that also respond to various dyes (Cataldo et al., 2010). The color unmixing process herein implemented was followed by normalization of each of the found stain components, allowing to compare feature descriptors between local patches from different tissues.

**Color Unmixing by Non-Negative Matrix Factorization.** Several approaches have been proposed to separate stained tissues. Most of them classify pixels using a partition of a particular color space (RGB, HSI, CIELab, or CMYK), usually generated or even learned from characteristic regions of the image. Although these techniques have reported promising results when separating a single color contribution from a stained tissue, they ignore that an observed color is very likely a complex combination of many dyes. Instead the procedure herein used performs the color unmixing based on the stain-specific RGB absorption, and hence this method can separate overlapped stains.

Even the more complex amalgam of colors, observed in actual microscopical images, can be described as a linear combination of simpler elements, not necessarily independent, but with a maximal statistical independence [Barlow’s principle of redundancy reduction (Barlow, 1961)]. In the present problem, the observed color comes from the two dyes applied to the sample, described by Eq. (1) (Ruijrok and Johnston, 2001).

\[ I_{t} = V_{S_{t}} + e_{t} \]  

where \( I_{t} \) denotes a \( 3 \times 1 \) vector of the \( r, g, b \) color components, the \( 2 \times 1 \) stain color-base, \( V \) is the \( 3 \times 2 \) matrix containing the unknown mixing coefficients that correspond to the contribution of the each dye, and \( e_{t} \) a \( 3 \times 1 \) vector, representing an additive noise. Provided that every observed color is in a \( m \times 3 \) matrix \( I \) (where rows correspond to the number of single colors in the image and columns to a RGB component) and assuming that noise can be neglected, the problem of dye unmixing can be formulated as a linear factorization problem:

\[ I \approx VS \]  

There exist several methods for solving such linear factorization, \( VS, \) of \( I, \) among others, Principal Component analysis (PCA), independent component analysis (ICA), and non-negative matrix factorization (NMF). Rabinovich et al. have shown that NMF outperforms ICA factorization, under the restrictions that there are two stain sources (\( S \)) and their contributions are forced to be non-negative (Rabinovich et al., 2003). In other words, the searched solution should expose the sources with maximal statistical independence, even though these sources are linearly dependent. In this work, the NMF approach finds non-negative matrices \( V \) and \( S, \) at minimizing the Frobenius norm between the observed data and their respective factorizations. For doing so, we use the projected gradient bound-constrained optimization method (Lin, 2007), which has better convergence properties than the standard multiplicative actualization rules used by Rabinovich et al. (2003).

Before factorization, images were converted to their corresponding optical density (OD) units by taking the negative log of the RGB sample, with each component normalized to \([0,1]\). For stability reasons, the colors that represent pixels with little stain were not considered (\( OD \leq 0.1 \)). Consequently, for avoiding a boundary effect on the reconstructed images, after the stain color-basis were defined, we linearly unmixed the unconsidered data (\( OD \leq 0.1 \)) using the transpose of the Moore–Penrose pseudoinverse of \( S \) (Penrose, 1995).

\[ S = \text{pinv}(V) \times I \]  

Figure 2 illustrates the unmixing process of typical histopathological skin images. Original images are shown in the first row, second and third rows show the resulting hematoxylin and eosin contributions, respectively.

**Stain Normalization.** Many factors may jeopardize any histological preparation, the environment illumination conditions, the relative dye quantity, or the remaining film inhomogeneities produced by slide storage and handling. These variability sources contribute in many different ways and scales to the observed differences of the same microstructural concept, for instance the basophilic organelle appear as a large pale of different blues, as observed in Figure 2. A rigorous control of this chain of events results then essential for obtaining slides with a proper final quality, a precondition difficult to meet in actual clinical scenarios. This problem has been bypassed by transforming images to grayscale (Ahammer et al., 2009; Huang and Lai, 2010) or to another color space that provides high discrimination (Doyle et al., 2010; Pham et al., 2007; Yang et al., 2005). However, when the image representation schemes highly depend on color and intensity information, color standardization must be applied, attempting to reduce the influence of the mentioned factors when normalizing the stain components to a
gold standard. The color linear normalization proposed by Reinhard et al. (2001) has been used in previous works (Wang et al., 2007, 2009). However, they apply normalization to the whole image, resulting in an inappropriate color transfer when there exists an imbalance of the stain components (for instance, images mainly composed of connective tissue). This drawback was herein overcome by separately normalizing each stain component. This normalization process is guided by the statistical properties of the two stain distributions, similar to what Reinhard et al. proposed, but instead of using a perceptual iab color model, we introduced the unmixed stains distributions, whereby the mean and standard deviations are separately equalized for each of the two stains, as suggested by Reinhard et al. (2001).

Figure 3 illustrates the normalization process (same images of Fig. 2), first and second rows display the normalization of the hematoxylin and eosin contributions, whereas the third row shows the resulting RGB image as a linear combination of the first two. Notice that images stained with the same dye look pretty similar, that is to say, about the same red and blue spectral values.

Image Representation

Each image, at the training stage (sub-image at the test stage), is represented by the probability distribution of the hidden topics, inferred from its constituent parts, using a probabilistic Latent Semantic Analysis (pLSA) model (Hofmann, 2001). This representation attempts to capture the semantic concepts in the image. The probability distribution of the hidden topics is given by the co-occurrence of smaller parts, whereas concepts are obtained when somehow these topics are combined. In this section the process of generating such representation is introduced.

Representing Images as a Bag of Visual Primitives. Overall, a linear representation model of the whole image hardly captures the complete richness of the image structures, because they are constructed under non-linear rules. However, if the image is analyzed at the level of small local patches, it is possible that non linearity can be linearly approximated (Olshausen, 2003). Image semantic can thus be captured when expressing the global contents as a combination of its atomic components, for instance using a Bag-of-Features (BoF). The BoF representation is generated when selecting a set of local image patches (LiPs), which are then characterized in a feature space. Instances of this feature space are grouped up using a conventional clustering algorithm, e.g., k-means, with k fixed. The image is finally represented as a k-bin histogram of visual primitives, i.e., the quantized visual patches.

The success of the BoF strategy lies upon finding the semantic image parts, i.e., how to select the LiPs and how to characterize them (Nowak et al., 2006). Several approaches have been proposed for selecting LiPs, including the use of multiscale keypoint detectors (Laplacian of Gaussian, Harris-affine, Scale Invariant Points, etc.) and many densely sampling strategies (regular grids, random sampling, etc.). Such selectors have proven their effectiveness in matching and classifying natural scenes (Jiang et al., 2007; Nowak et al., 2006; Yang et al., 2009), but they were not designed at all for capturing the relevant patches of histopathology images. In the present work, these simple semantic units are captured when detecting and describing the basophilic structures, typically cells, within the histopathological image.

Extracting Local Patches (LiPs) Based on Stain Image Features. LiPs corresponding to image cells were determined as regions surrounding the detected cell
nuclei (typically, $24 \times 24$ pixels), easily segmented from the hematoxylin dye contribution. For doing so, it was applied a simple Otsu threshold, followed by a morphological gray-scale opening filter with a $2 \times 2$ squared structuring element. Clumped or superimposed nuclei were split using a morphological gray-scale closing filter on areas larger than the largest found nucleus. Figure 4 shows some examples of detected points using this strategy, which will be noted as stainP hereafter. 

Describing the Stain Features of Detected Visual Patches. Once regions were determined, the next step was the characterization of these regions by projecting the raw data into a feature space. The aim of this phase was to set up small regions, attempting to figure out micro-structures with biological meaning. From a pathologist point of view, visual identification of these micro-structures lies upon cytoplasm and nuclei variations, specifically stain, texture, and structure. These differences can be fairly captured by some local feature descriptors such as the scale-invariant feature transform (SIFT) and the discrete cosine transform (DCT) descriptors, which have been used to effectively describe this kind of variations in small regions (Bosch et al., 2006; Kamiya et al., 2009).

- **SIFT descriptors.** The SIFT descriptor has shown to be a very robust and reliable representation of the local variations around an image point (Kamiya et al., 2009). This descriptor characterizes local image structures by using a set of gradient orientation histograms. In this work, the local stain variations were described by the conventional SIFT descriptor proposed by Lowe, i.e., LiPs were split into $4 \times 4$ blocks and eight orientation histograms were computed for each block (Lowe, 2004), resulting in a 256-dimensional feature vector given by concatenating of the SIFT descriptors, obtained from both hematoxylin and eosin contributions.

- **DCT Coefficients.** The DCT based descriptors has shown to be one of the most accurate, compact, and robust methods for the extraction of local texture features in the spatial frequency domain (Chaddha, 1994). The DCT descriptor was herein obtained by applying the DCT to each stain component image and storing the 21 lower frequency coefficients.
Constructing the Visual Vocabulary. Once the sets of feature vectors are computed, the space of characteristics is partitioned into \( k \) clusters and a visual vocabulary is obtained, i.e., a reduced set of visual primitives defined by the cluster centers. The clustering algorithm herein implemented was the incremental \( k \)-means, which minimizes the intra-cluster variance given by the summed squared error (Bermejo and Cabestany, 2002). This algorithm starts with a set of \( k \) randomly selected instances as cluster centers, which are iteratively updated using a learning function that adjusts each cluster center to the closer training LiP, such that the error between the training LiP and the new cluster center is gradually reduced. Finally, the visual vocabulary is used for generating a \( k \)-bin histogram that represents the frequency of each visual primitive within the image.

A new image histogram is generated by associating each corresponding LiP feature vector to the closest visual primitive, using an euclidean distance.

Representing Images as a Distribution of Hidden Topics. So far, images are represented with local feature descriptors, the histogram of visual primitives. This simple description of visual primitive frequencies ignores the co-occurrence among primitives. This issue results fundamental for figuring out the primitives. This simple description of visual primitive frequencies ignores the co-occurrence among primitives. This simple description of visual primitive frequencies ignores the co-occurrence among primitives. Therefore, the distribution functions \( P(z_i | I_i) \) and \( P(w_i | z_i) \) can be learned from the likelihood of the observed data as:

\[
L = \prod_{i=1}^{M} \prod_{j=1}^{N} P(w_i | I_j) P(z_i | I_j) \tag{5}
\]

where \( N \) is the number of images, \( M \) is the number of visual primitives in the vocabulary (given by the number of clusters), \( n(w_i | I_j) \) is the number of occurrences of a visual primitive \( w_i \) in the image \( I_j \) and \( P(w_i | I_j) \) is given by Eq. (4). Best parameters are found using the expectation maximization (EM) algorithm, which iteratively estimates the posterior probabilities of the latent variables \( P(z_k | I_i, w_j) \) (expectation step) and optimizes the \( P(w_j | z_k) \) and \( P(z_k | I_i) \) estimations (maximization step), until convergence.

- **E-step:** compute the posterior probabilities of the latent variables.

\[
P(z_k | I_i, w_j) = \frac{P(w_j | z_k) P(z_k | I_i)}{\sum_{l=1}^{K} P(w_j | z_l) P(z_l | I_i)} \tag{6}
\]

- **M-step:** maximize the expected complete data log-likelihood.

\[
P(w_j | z_k) = \frac{\sum_{m=1}^{N} n(I_i, w_j) P(z_k | I_i, w_j)}{\sum_{m=1}^{M} n(I_i, w_m) P(z_k | I_i, w_m)} \tag{7}
\]

\[
P(z_k | I_i) = \frac{\sum_{j=1}^{M} n(I_i, w_j) P(z_k | I_i, w_j)}{n(I_i)} \tag{8}
\]

Inferring the Topic Distribution of a New Image. The conditional probability distribution of a new image \( P(z | I_{new}) \) can be inferred by maximizing the likelihood of the image \( I_{new} \) with respect to the conditional probability distribution \( P(w_j | z_k) \) learned from the training data. For
Classification Model of Single Concept Images

After the pdf $P(z | I)$ is computed, the set of probability values of each topic within the image forms a feature vector given by:

$$d_i = \{P(z_1 | I_i), P(z_2 | I_i), \ldots, P(z_k | I_i)\}$$  \hspace{1cm} (9)

These vectors are then used as the input for training a discriminative learning model, a bank of support vector machines (SVMs) that learns whether a particular concept is present in the image. The SVM strategy maps the input vectors into a high dimensional feature space, induced by a kernel function. In the feature space, the learning algorithm produces a hyperplane that optimally separates two classes, maximizing a convex quadratic form. This optimal hyperplane corresponds to the maximal distance to the closest training data and is represented as a linear combination of training points, called the support vectors. Although SVM is a linear discriminator, non-linear discrimination can be achieved using an appropriate kernel. In the present investigation we used the radial base function (RBF) kernel because of its well known performance in image classification tasks. The multi-class classification problem is approached using a bank of these binary RBF classifiers. A single classifier learns the conditional probability of one class given another, ignoring the instances that belong to the remaining classes. At the end, the bank of classifiers estimates the probability of a concept as the largest average of conditional probability of one class given another, ignoring the instances that belong to the remaining classes. The image semantics is now hidden in a set of 29 blocks, 12 from the original partition and nine (eight) from the vertical (horizontal) splits. Each of these 29 blocks is associated to a single semantic concept, set by the learning model. A further redistribution of concepts is achieved by evaluating the overlap influence. For doing so, the concept associated to a particular image region is set by a rule which takes into account the labels of the intersecting blocks. The rule uses a simple majority vote weighted by the confidence interval. This approach makes that concepts are spatially located, a semantic advantage to characterize more complex entities.

Model Evaluation

The proposed approach was evaluated by classifying normal and pathological skin tissues. The performances of the three main methodological issues were separately assessed: the sampling strategy, the feature descriptor, and the annotation task. The sampling strategy was evaluated and compared with two baseline sampling techniques. The stain-based descriptor was also compared with two baseline feature descriptors. Finally, we evaluated the performance of the block-based strategy for annotating multiple concept images.

Histopathological Material and Image Acquisition. Slides provided by the pathology department of the National University of Colombia were used for evaluation. Skin biopsies of patients diagnosed with different types of basal cell carcinoma (BCC) were fixed in formalin, embedded in paraffin, and stained with hematoxylin–eosin. Images were digitized using a standard microscope (eclipse E600 Nikon), coupled to a color digital camera (DXM1200 Nikon), controlled by Akt-1 V. 2.62 software from Nikon Corporation. Each image was acquired at 10× objective, corresponding to 0.67μ².

This carcinoma was herein selected as our use case because of its particular richness in pathological concepts, i.e., characteristic arrangements of cells, surrounded by several combinations of normal–abnormal epithelial and connective tissues, also found in many other pathologies (McGee et al., 1992). This pathology is highly variable as a whole but also regarding the number of associated concepts. With the available biological material, 10 concepts were established, including normal biological structures (Hair follicles, sebaceous glands, eccrine glands, collagen, and epidermis), different kinds of BCC (nodular, morpheiform,
micro-nodular, and cystic) and the inflammatory infiltration, a particular pathological finding characterized by large macrophage and lymphocytic presence within the stroma. Figure 7 shows examples of the 10 concepts, notice their morphological differences, whereas Figure 8 illustrates the large visual variability of two single concepts, i.e., the hair follicles (first row) and the nodular BCC (second row).

**Data Set Composition.** The evaluation set was composed of 655 digital images. The entire image set was randomly divided into training (40%), test (40%), and validation (20%) sets. Square subimages were manually cropped from the training and test image sets, aiming to locate single concepts. In general, there is no typical size for these concepts, even for anatomical structures such as the hair follicle, because of the large intrinsic variability, namely, the particular organ, the type of biopsy, the cut orientation and the fixation procedure. Therefore, single concepts were estimated as an average value of a set of regular regions that the pathologist marked as containing a single biological concept, i.e., square subimages of 300 × 300 pixels. A total of 775 training and 757 test regions were finally obtained and annotated by an expert, with a single one among the ten possible concepts. The concept distribution of these sets is presented in Table 1.

The validation set was composed of 131 images, annotated by an expert. These images may contain more than one concept and were used for the evaluation of the block-based annotation strategy.
Experimental Setup. The proposed strategy was two-fold evaluated: first, an exhaustive parameter search was performed on the test image set, each image labeled with a unique concept. For the RBF kernel, the complexity parameter $\gamma$ was varied from 0.1 to 2.0, with increment steps of 0.1, and the regularization parameter $C$ (complexity) was varied between 1 and 10, with increment steps of 0.2. On the other hand, the size of the visual vocabulary and the number of hidden topics were also evaluated. The former was varied from 100 to 500, with increment steps of 100, whereas the latter was varied from 10 to 50, with increments of 10. Best parameter values were stored and used to evaluate the annotation performance in a set of images, containing more than one concept.

Base-Line Techniques. Sampling and feature description strategies used for generating the visual vocabulary, were compared against baseline techniques. Provided that any image is first mapped to a vocabulary, were compared against baseline techniques used for generating the visual description strategies used for generating the visual vocabulary, were compared against baseline techniques.

Sampling: The proposed sampler was compared to two state-of-the-art strategies: a SIFT point detector (siftP; Lowe, 2004) and a dense grid partition (gridP; Caicedo et al., 2009; Cruz-Roa et al., 2009). The former finds relevant regions around local keypoints whereas the latter captures every local feature in the image, typically using a 8 × 8 partition window.

Feature description: Three local descriptors were used for comparison, two describing intensity characteristics, the SIFT and DCT descriptors (computed on the intensity channel), and a third one, the DCT coefficients of each of the RGB channels (rgbDCT) concatenated into a single feature.

Performance Measurement. Although the performance of any classification task is commonly quantified in terms of its predictive accuracy, this measure results inadequate in problems with unbalanced class distribution, because the contribution of each class to the entire accuracy rate is a function of its cardinality, case in which misclassification of the minority class instances has insignificant impact on the performance. In the present work, the classification of each biological concept was independently assessed by applying the “one against all” strategy, which consists in generating multiple two-class evaluations, each reporting the performance of distinguishing a single concept (positive samples) from the rest (negative samples) and thereby assessing multiple class unbalanced problems. As an alternative, some authors have proposed the use of statistical metrics, attempting to give preference to the correct classification of either positive or negative samples. Therefore, the classification task, i.e., the identification of the biological concepts in the single-concept images, was assessed using the effectiveness F-measure, given by Eq. (10).

$$F_\beta = \frac{TPR \times PPV}{\beta \times PPV + (1 - \beta) \times TPR}$$

with,

$$PPV = \frac{TP}{TP + FP} \text{ and } TPR = \frac{TP}{TP + FN}$$

where TP stands for the true positives, TN for the true negatives, FN for the false negatives, FP for the false positives, PPV for the precision or positive predictive value, and TPR for the sensitivity or true positive rate. Additionally, the $\beta$ coefficient ($0 < \beta < 1$) was set to 0.4 for facing the unbalanced class distribution generated by the one-against-all strategy. So, the parameter search was addressed to favor the sensitivity, aiming that misclassification of the minority class instances had a larger impact on the performance (Daskalaki et al., 2006).

On the other hand, the performance of the annotation tasks, i.e., identification of the concepts present in an image with multiple biological concepts, was also assessed, independently for each concept and quantified in terms of sensitivity (TPR), specificity (TNR), and accuracy as:

$$TNR = \frac{TN}{FP + TN} \quad \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

EXPERIMENTAL RESULTS

Sampler Strategy Assessment

Classification effectiveness of the proposed sampler (stainP) was compared to three baseline techniques (siftP, gridP). The presented results were obtained using 300 clusters, a complexity $C$ set to 2.0 and a regularization parameter $\gamma$ set to 1.0. These parameters reported a stable trade-off between classification performance and model complexity. Results are two-fold reported, effectiveness, and efficiency. The effectiveness measure ($F_\beta$) is presented in Figure 9 for the three feature descriptors in the three left panels: SIFT (upper panel), DCT (mid panel), and rgbDCT (lower panel). Each individual graph plots the effectiveness measure ($F_\beta$) for the evaluated sampling strategies in the y-axis against the number of hidden topics in the x-axis. A good descriptor should be able of capturing most statistical properties of the original data set, while spending a minimal possible computational cost. Therefore, we also evaluated the efficiency provided by any of the three descriptors, herein defined as the ratio between the effectiveness and the average number of
patches used for representing the image. This efficiency was plotted in the right panels for the three descriptors: SIFT (upper panel), DCT (mid panel) and rgbDCT (lower panel). Each individual graph plots the efficiency measure in the y-axis against the number of hidden topics in the x-axis.

Overall, the best and steadier performance is always observed for the gridP sampler with any of the used descriptors (left panels). The proposed sampling approach is slightly smaller, showing also a high and stable performance. In contrast, the siftP detector performance is variable and shows a small value (<0.3) when the sift descriptor was used (upper left panel).

Lower left panel shows the evaluated descriptors that used color information, case in which differences among the sampling strategies decreased, i.e., although the gridP sampler performance was still high, the other samplers improved their performance. Descriptors played a main role in the final performance, much more important than the sampling strategy. This statement is illustrated when evaluating the SIFT feature vector (upper left panel), which showed the lower performance for any of the sampling strategies.

Right panels in Figure 9 plot the efficiency reported for each of the pairs samplers–descriptors, depicted in

---

Fig. 9. Classification performances for different sampling strategies against the number of hidden topics. Left plots correspond to the effectiveness measure, whereas efficiency is depicted in right panels as follows: SIFT (upper panel), DCT (middle panel), and rgbDCT (lower panel) descriptors. In all cases, the proposed approach (continuous line) shows the best tradeoff between effectiveness and efficiency.
the corresponding left panels. The results demonstrate that the best trade-off between classification accuracy and computational cost is our proposed strategy. The plot presents the number of patches for an image of 300 × 300 pixels: whereas the gridP sampler required a constant number of 1444 patches, the other two approaches required a variable number of patches, which was a function of the information contained in the image, i.e., the same image was represented with an average of only 189 (our proposal) and 385 (SIFT point detector) patches, calculated from the test and training sets.

Figure 10 shows the plot of the number of patches used to represent the visual content of 100 randomly selected images. Again, the gridP sampler uses a constant number of patches, whereas the other two require a variable number. Note that the SIFT point detector almost always uses a larger number of regions.

**Feature Descriptor Assessment**

The second evaluated factor was the effect of using stain based visual features as LiP descriptors for the same classification task. The performance obtained with SIFT and DCT descriptors was compared with the standard intensity SIFT and rgbDCT descriptors. Figure 11 shows the plots of the $F_B$ measures for the three sampling strategies: gridP, siftP, and stainP (from top to down). Each of these graphs show the performance measure (y axis) against the number of hidden topics (x axis) for the four feature descriptors. In these experiments, the number of clusters, complexity, and regularization parameters were also set to 300, 2, and 1, respectively.

In the three panels it is systematically observed that the use of the proposed stain-based descriptors improves the classification performance. The stain-based SIFT descriptor obtained a higher performance, when comparing to the baseline intensity SIFT descriptor, which reports the poorer performance among the three samplers. The stainSift performance was nevertheless lower than the obtained with the stainDCT descriptor (continuous line), which outperforms every descriptor in every assessed sampler.

The best performance was obtained with the combination gridP sampler—stainDCT descriptor (second panel). Interestingly, there is a large performance improvement when combining this descriptor with any other sampler, up to reducing differences from 6.5 to 1.5%, in average. Likewise, the use of stain-based features highly improved the proposed sampler performance, comparable to what was reported for the combination gridP and rgbDCT, which showed an effectiveness of 0.69 in Section “Sampler Strategy Assessment.”
Annotating Images with Multiple Concepts

The block-based annotation strategy estimated the visual concepts contained in the validation set. The optimized learning model, resulting of representing the training images with the proposed approach, was used to classify each overlapped image block. Afterward, a weighted voting strategy was applied for selecting a list of the definitive concepts per image.

Figure 12 illustrates the process of annotating an image of 1024 × 768 pixels, using the block partition described in Section “Annotation of Multiple Concept Images.” Therefore, images were partitioned into 256 × 256 blocks of pixels, which had similar size to those used for training the single-concept classification model, thereby attempting that they contain a unique biological concept. The two upper panels and the left lower one show the same image, partitioned by following the protocol defined, whereas the right lower panel displays the final annotation. In this particular example there are multiple errors which are corrected when the final voting strategy decides. For instance, in the upper left panel, the image is partitioned in 12 blocks and the right block of the upper row was erroneously classified as a hair follicle (HF). Similarly, in the down left panel, the image is partitioned in eight blocks and the four block of the second row was erroneously annotated as sebaceous glands (SG). Observe that the area defined by these blocks intersect different blocks in the other partitions, that is to say, in the upper right and lower left panels. Notice also that the classification label is different and that the voting strategy corrects these misclassifications. The lower right panel shows that those labels were correctly assigned: a hair follicle (HF) in a great percentage of this image, a block of epidermis (EP) correctly found at the upper right corner, the normal collagen sections (CO) and the irregular distribution of inflammatory infiltration (II), again correctly labeled.

Table 2 shows the performance per concept in the validation set. The proposed method was evaluated using accuracy, sensitivity, and specificity measurements. In average, the method reports an accuracy of 71%, a sensitivity of 84%, and a 67% of specificity. Sen-
Semantic concepts are then drawn from the image using a part-based latent semantic analysis of the set of patches. A complementary block-based partition splits the image into regions containing a single concept so that a coarse concept location is achieved. The proposed model was assessed in a set of 655 skin tissue images, containing 10 different biological concepts: normal biological structures (hair follicles, sebaceous glands, eccrine glands, collagen, and epidermis), and different kinds of pathological tissues (four types of BCC and the inflammatory infiltration). The strategy was compared to state-of-the-art strategies (Caicedo et al., 2009; Caicedo and Izquierdo, 2010), improving the sensitivity from a reported 15% to a 84%.

A main goal of the present investigation was to develop automatic strategies for useful annotation of actual histopathological images, i.e., a variable distribution of very complex concepts and cluttered background, as illustrated in Figures 7 and 8. In this work these challenges were approached by applying a particular part-based representation approach. Instead of using classical keypoint detectors such as the SIFT points or dense sampling, a relevant patch detector was proposed, which determines the hematoxylin stain contribution for detecting the nuclei that are used as the generator of semantics associated to basic information units that describe how cells are organized. Annotation of multiple concepts was herein successfully performed using a block-based partition strategy, evaluated on a very complex pattern recognition task. Although a straight comparison was not possible since the evaluation data sets were different from those used in the literature, the obtained results largely outperform what has been so far reported in similar tasks. Specifically, the proposed method exhibits average specificity and sensitivity of 67% and 84%, for the annotation task, respectively. Likewise, the precision of the annotated regions was visually evaluated by an expert, resulting in an average precision of 49% and an average effectiveness of 0.6. The observed sensitivity and specificity values are comparable to what is reported in the literature regarding inter-observer variability, i.e., specificity around 70% and sensitivity around 90% for most of the assessed concepts (Brochez et al., 2002), indicating that the proposed approach maybe used in real clinical and training scenarios. Systematic errors were reported because of the similarity among tissues, for instance some small regions of the epidermis and hair follicles were erroneously marked as any of the BCCs, or a carcinoma was confused with another type, or sebaceous and eccrine glands were mixed up. These misclassifications were observed in small regions and affected very little the reported sensitivity which was always high (>89% in almost cases), whereas its influence on the reported precision was more important. These systematic errors could be reduced by introducing spatial or structural contextual information, chosen for the specific classification tasks. For instance, in this kind of images a prior knowledge about the relative location of the epidermis could be used for correcting misclassifications of this concept in unexpected regions.

In the absence of standardized histological and capturing procedures, histopathological images exhibited several variations regarding their color and intensity.

**TABLE 2. Block based annotation performance for every searched concept**

<table>
<thead>
<tr>
<th>Label</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hair follicles</td>
<td>55</td>
<td>58</td>
<td>54</td>
</tr>
<tr>
<td>Collagen</td>
<td>83</td>
<td>89</td>
<td>78</td>
</tr>
<tr>
<td>Nodular BCC</td>
<td>63</td>
<td>97</td>
<td>51</td>
</tr>
<tr>
<td>Morpheiform BCC</td>
<td>62</td>
<td>93</td>
<td>58</td>
</tr>
<tr>
<td>Micro-nodular BCC</td>
<td>72</td>
<td>67</td>
<td>73</td>
</tr>
<tr>
<td>Cystic BCC</td>
<td>81</td>
<td>90</td>
<td>81</td>
</tr>
<tr>
<td>Epidermis</td>
<td>83</td>
<td>90</td>
<td>81</td>
</tr>
<tr>
<td>Sebaceous glands</td>
<td>69</td>
<td>94</td>
<td>60</td>
</tr>
<tr>
<td>Eccrine glands</td>
<td>76</td>
<td>95</td>
<td>72</td>
</tr>
<tr>
<td>Inflammatory infiltration</td>
<td>66</td>
<td>69</td>
<td>60</td>
</tr>
<tr>
<td>Average</td>
<td>71</td>
<td>84</td>
<td>67</td>
</tr>
</tbody>
</table>
properties. This drawback was overcome using the well known color transfer approach proposed by Reinhard et al. (2001), but normalizing separately each stain component. The color normalization effect was herein evaluated by comparing the \(F_p\)-performance on the whole experimental set (results not shown) for the classification task. The color standardized method outperformed the non standardized in a 5.2\% with the rgbDCT, in a 14.1\% with the stainSIFT and in a 29.1\% with the stainDCT descriptors. Interestingly, the best performance was obtained when the stain contribution was used, an expected result since the normalization process attempted to preserve the biological meaning of colors at local patches, i.e., the color decomposition process results crucial since semantics lies upon the separation of particular tissues.

Our fundamental hypothesis was that the stain decomposition process allowed to determine the basic information units. The construction of a new representation space was accomplished using a conventional non-negative matrix factorization (NMF) approach. This approach assumed that the observed color comes from two sources, the two dyes. However, this assumption can fail when artifacts or other pigments are present. This drawback can be overcome by introducing a term of data fidelity in the non-negative matrix normalization formulation (Li et al., 2007). The stain decomposition, followed by a stain normalization step, allowed a reliable detection of nuclei cells, which were herein used as relevant keypoints, by applying a simple threshold on the hematoxylin component. As the representation approach proposed here did not require accurate cell segmentation, endeavors were not focused towards splitting the clumped cells. However, these cases were easily solved using simple mathematical morphology tools.

Once relevant local patches were found, efforts were dedicated to model the complex contextual relationships of the minimal semantic units (local patches). The conditional analysis carried out by the pLSA model was used to capture the dependencies of these local patches, allowing to describe the image in terms of middle semantic concepts, the hidden topics. Although this model can be used as a classifier by setting the category of every instance with the higher probability topic, we found that this strategy showed poor classification performance (results not shown). It should be strengthen out that the conditional probability of these local patches varies for different instances of the same concept, because of the large variability of biological structures. Provided that the number of possible background-structure combinations is finite, it was possible to learn this semantics using a bank of discriminative pair-wise classifiers that sets a unique concept associated to an image block.

Before annotating images with multiple concepts, learning models were trained and tested with images, mainly containing a single concept. Performance was evaluated in terms of effectiveness and efficiency measures. Effectiveness combines in a single measure the well known precision and recall (sensitivity) measures. Herein, we give more importance to the sensitivity, allowing some accuracy loss, whereas the ratio between the effectiveness and the average number of patches used for representing the image, gives an idea of the method efficiency. Performance results were compared to state-of-the-art techniques. Two sampling strategies (SIFT point and dense regular grid) and two descriptors (DCT coefficients and SIFT descriptor) were evaluated. Results indicated that: (1) the low frequency information, captured by the DCT coefficients, was a best descriptor of the differences between the local patches. (2) Unlike to what has been reported for global descriptors, color is a very important feature in local part based representations of histopathological images. (3) Description based on the stain contributions improves the classification performance for every sampler, with largest improvements reported to the proposed sampler. (4) The proposed sampler is able to identify the most informative local regions in the classification task, i.e., for very similar performance, the number of local patches used by the grid sampling strategy is large (1444 per 300 x 300 pixel image), whereas our approach barely uses an average of 191, much lower than the former. A side gain is of course that the proposed method could be used in actual scenarios of large image data bases or virtual slides.

The proposed stain-based representation model allowed to identify the local relevant image regions in terms of a set of semantic concepts. The good performance shown in our experiments suggests that this approach can be applied to any other type of histopathological image, that is to say that new concepts may be included in the training phase and the proposed model would learn these new patterns. Likewise, the method could be easily extended to new stains by associating the particular color to either the nucleus or the cytoplasm because in essence every dye in pathology is devised to color them.

REFERENCES


G. DÍAZ AND E. ROMERO


