ViVo: Visual Vocabulary Construction for Mining Biomedical Images

Arnab Bhattacharya
UC Santa Barbara
arnab@cs.ucsb.edu

Vebjorn Ljosa
UC Santa Barbara
ljosa@cs.ucsb.edu

Jia-Yu Pan
Carnegie Mellon Univ.
jypan@cs.cmu.edu

Mark R. Verardo
UC Santa Barbara
verardo@lifesci.ucsb.edu

Hyungjeong Yang
Chonnam Natl. Univ.
hjyang@chonnam.ac.kr

Christos Faloutsos
Carnegie Mellon Univ.
christos@cs.cmu.edu

Ambuj K. Singh
UC Santa Barbara
ambuj@cs.ucsb.edu

Abstract

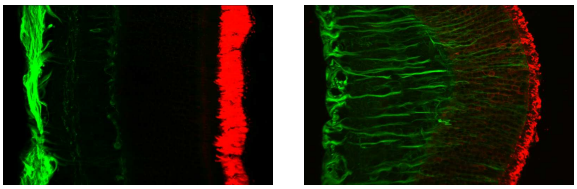
Given a large collection of medical images of several conditions and treatments, how can we succinctly describe the characteristics of each setting? For example, given a large collection of retinal images from several different experimental conditions (normal, detached, reattached, etc.), how can data mining help biologists focus on important regions in the images or on the differences between different experimental conditions?

If the images were text documents, we could find the main terms and concepts for each condition by existing IR methods (e.g., tf/idf and LSI). We propose something analogous, but for the much more challenging case of an image collection: We propose to automatically develop a visual vocabulary *by breaking images into $n \times n$ tiles and deriving key tiles ("ViVos") for each image and condition. We experiment with numerous* domain-independent *ways of extracting features from tiles (color histograms, textures, etc.), and several ways of choosing characteristic tiles (PCA, ICA).*

We perform experiments on two disparate biomedical datasets. The quantitative measure of success is classification accuracy: Our "ViVos" achieve high classification accuracy (up to 83 % for a nine-class problem on feline retinal images). More importantly, qualitatively, our "ViVos" do an excellent job as "visual vocabulary terms": they have biological meaning, as corroborated by domain experts; they help spot characteristic regions of images, exactly like text vocabulary terms do for documents; and they highlight the differences between pairs of images.

1 Introduction

We focus on the problem of summarizing and discovering patterns in large collections of biomedical images. We would like an automated method for processing the images and constructing a visual vocabulary which is capable of describing the semantics of the image content. Particularly, we are interested in questions such as: *"What are the interesting regions in the image for further detailed investigation?"* and *"What changes occur between images from different pairs of classes?"*



(a) Normal (b) 3d

Figure 1. Examples of micrographs of (a) a normal retina and (b) a retina after 3 days of detachment. The retinas were labeled with antibodies to rhodopsin (red) and glial fibrillary acidic protein (GFAP, green). Please see the electronic version of the article for color images.

As a concrete example, consider the images in Figure 1. They depict cross-sections of feline retinas—specifically, showing the distributions of two different proteins—under the experimental conditions "normal" and "3 days of detachment." Even a non-expert human can easily see that each image consists of several vertical layers, despite the fact that the location, texture, and color intensity of the patterns in these layers vary from image to image. A trained biologist can interpret these observations and build hypotheses about the biological processes that cause the differences.

This is exactly the goal of our effort: We want to build a system that will automatically detect and highlight patterns differentiating image classes, after processing hundreds or thousands of pictures (with or without labels and timestamps). The automatic construction of a visual vocabulary of these different patterns is not only important by itself, but also a stepping stone for larger biological goals. Such a system will be of great value to biologists, and could provide valuable functions such as automated classification and supporting various data mining tasks. We illustrate the power

of our proposed method on the following three problems:

Problem 1 *Summarize an image automatically.*

Problem 2 *Identify patterns that distinguish image classes.*

Problem 3 *Highlight interesting regions in an image.*

Biomedical images bring additional, subtle complications: (1) Some images may not be in the canonical orientation, or there may not be a canonical orientation at all. (The latter is the case for one of our datasets, the Chinese hamster ovary dataset.) (2) Even if we align the images as well as possible, the same areas of the images will not always contain the same kind of tissue because of individual variation. (3) Computer vision techniques such as segmentation require domain-specific tuning to model the intricate texture in the images, and it is not known whether these techniques can spot biologically interesting regions. These are subtle, but important issues that our automatic vocabulary creation system has to tackle.

We would like a system that automatically creates a visual vocabulary and achieves the following goals: (1) *Biological interpretations*: The resulting visual terms should have meaning for a domain expert. (2) *Biological process summarization*: The vocabulary should help describe the underlying biological process. (3) *Generality*: It should work on multiple image sets, either color or gray-scale, from different biological domains.

The major contributions of this paper are as follows:

- We introduce the idea of "tiles" for visual term generation, and successfully bypass issues such as image orientation and registration.
- We propose a novel approach to group tiles into visual terms, avoiding subtle problems, like non-Gaussianity, that hurt other clustering and dimensionality reduction methods. We call our automatically extracted visual terms "**ViVos**."

The paper is organized as follows. Section 2 describes related work. In Section 3, we introduce our proposed method for biomedical image classification and pattern discovery. Classification results are presented in Section 4. Experiments illustrating the biological interpretation of ViVos appear in Section 5. Section 6 concludes the paper.

2 Background and Related Work

Biomedical images have become an extremely important dataset for biology and medicine. Automated analysis tools have the potential for changing the way in which biological images are used to answer biological questions, either for high-throughput identification of abnormal samples or for early disease detection [7, 18, 19]. Two specific kinds of biomedical images are studied in this paper: confocal microscopy images of retina and fluorescence microscopy images of Chinese Hamster Ovary (CHO) cells.

The retina contains neurons that respond to light and transmid electrical signals to the brain via the optic nerve. Multiple antibodies are used to localize the expression of specific proteins in retinal cells and layers. The antibodies are visualized by immunohistochemistry, using a confocal microscope. The images can be used to follow a change in the distribution of a specific protein in different experimental conditions, or visualize specific cells across these conditions. Multiple proteins can be visualized in a single image, with each protein represented by a different color.

It is of biological interest to understand how a protein changes expression and how the morphology of a specific cell type changes across different experimental conditions (e.g., an injury such as retinal detachment) or when different treatments are used (e.g., oxygen administration). The ability to discriminate and classify on the basis of patterns (e.g., the intensity of antibody staining and texture produced by this staining) can help identify differences and similarities of various cellular processes.

The second kind of data in our study are fluorescence microscopy images of subcellular structures of CHO cells. These images show the localization of four proteins and the cell DNA within the cellular compartments. This information may be used to determine the functions of expressed proteins, which remains one of the challenges of modern biology [1].

2.1 Visual Vocabulary

A textual vocabulary consists of words that have distinct meanings and serve as building blocks of larger semantic constructs like sentences or paragraphs. To create an equivalent visual vocabulary for images, previous work applied transformation on image pixels to derive tokens that can describe image contents effectively [22, 4]. However, an image usually has tens of thousands of pixels. Due to this high dimensionality, a large number of training images is needed by pixel-based methods to obtain a meaningful vocabulary. This has limited the application of these methods to databases of small images.

One way to deal with this dimensionality curse is to extract a small number of features from image pixels. The vocabulary construction algorithm is then applied to the extracted features to discover descriptive tokens. A feature is usually extracted by filtering and summarizing pixel information. In many applications, these tokens have been shown useful in capturing and conveying image properties, under different names such as "blob," "visterm," "visual keywords," and so on. Examples of applications include object detection [20] and retrieval [21], as well as image classification [22, 4, 14] and captioning [5, 9].

Clustering algorithms or transformation-based methods

are other defenses against the curse of dimensionality. K-means clustering has been applied to image segments [5, 9] and the salient descriptor [21] for vocabulary construction. Examples of transformation-based methods include principal component analysis (PCA) [10, 22, 14] and wavelet transforms [20]. Recently, independent component analysis (ICA) [8] has been used in face recognition [4], yielding facial templates. Like the feature extraction approaches, these methods also have problems with orientation and registration issues, as they rely on global image features.

In this paper, we present a method that discovers a meaningful vocabulary from biomedical images. The proposed method is based on "tiles" of an image, and successfully avoids issues such as registration and dimensionality curse. We use the standard MPEG-7 features *color structure descriptor* (CSD), *color layout descriptor* (CLD) and *homogeneous texture descriptor* (HTD) [16]. The CSD is an n-dimensional color histogram (n is 256, 128, 64, or 32), but it also takes into account the local spatial structure of the color. For each position of a sliding structural element, if a color is present, its corresponding bin is incremented. The CLD is a compact representation of the overall spatial layout of the colors, and uses the discrete cosine transform to extract periodic spatial characteristics in blocks of an image. The HTD characterizes region texture using mean energy and energy deviation of the whole image, both in pixel space and in frequency space (Gabor functions along 6 orientations and 5 scales).

Alternatively, there is work on constructing visual vocabulary [17, 15] with a human in the loop, with the goal of constructing a vocabulary that better captures human perception. Human experts are either asked to identify criteria that they used to classify different images [17], or directly give labels to different patterns [15]. The vocabulary is then generated according to the given criteria and labels. These approaches are supervised, with human feedback as input to the construction algorithms. In contrast, our proposed method presented in this paper is unsupervised: The image labels are used only after the ViVos are constructed, when we evaluate them using classification.

3 Proposed Method for Symbolic Representation of Images

In this section, we introduce our proposed method for transforming images into their symbolic representations. The algorithm is given in Figure 2, and uses the symbols listed in Table 1. The algorithm consists of five steps.

The first step partitions the images into non-overlapping tiles. The optimal tile size depends on the nature of the images. The tiles must be large enough to capture the characteristic textures of the images. On the other hand, they cannot be too large. For instance, in order to recognize the red layer in Figure 1(a), the tile size should not be much larger than the width of the layer. We use a tile size of 64-by-64 pixels, so each retinal image has 8×12 tiles, and each subcellular protein localization image has 8×6 or 8×8 tiles.

Input: A set of n images $I = \{I_1, \ldots, I_n\}$.
Output: Visual vocabulary (ViVos) $\mathcal{V} = \{v_1, \ldots, v_m\}$.
ViVo-vectors of the n images $\{v(I_1), \ldots, v(I_n)\}$.
Algorithm:
1. Partition each image I_i into s_i non-overlapping tiles
2. For each tile $j \in \{1, \ldots, s_i\}$ in each image I_i, extract $\tilde{\mathbf{t}}_{\mathbf{i,j}}$
3. Generate visual vocabulary $\mathcal{V} = gen\_vv(\cup_{i=1}^{n} \tilde{\mathbf{t}}_{\mathbf{i,j}})$
   Also, compute $\boldsymbol{P}$, the PCA basis for all $\tilde{\mathbf{t}}_{\mathbf{i,j}}$'s.
4. For each tile $j \in \{1, \ldots, s_i\}$ in each image I_i,
   compute the ViVo-vector $v(\tilde{\mathbf{t}}_{\mathbf{i,j}}) = comp\_vivo(\tilde{\mathbf{t}}_{\mathbf{i,j}}, \mathcal{V}, \boldsymbol{P})$
5. For each image I_i, compute the ViVo-vector of I_i:
   $v(I_i) = summarize(\{\tilde{\mathbf{t}}_{\mathbf{i,j}} : j = 1, \ldots, s_i\})$

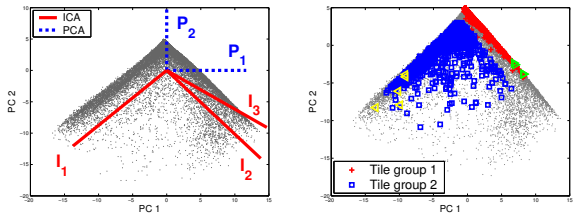
Figure 2. Algorithm for constructing a visual vocabulary from a set of images.

Symbol	Meaning
$\mathcal{V}$	Set of m ViVos: $\mathcal{V}=\{v_1, \ldots, v_m\}$
m'	Number of ICA basis vectors $m' = m/2$
$\tilde{\mathbf{t}}_{\mathbf{i,j}}$	j-th tile (or, tile-vector) of image I_i
$v(\tilde{\mathbf{t}}_{\mathbf{i,j}})$	m-dimensional ViVo-vector of tile $\tilde{\mathbf{t}}_{\mathbf{i,j}}$
$v(I_i)$	m-dimensional ViVo-vector of image I_i
f_k	The k-th element of $v(\tilde{\mathbf{t}}_{\mathbf{i,j}})$
$v_k(I_i)$	The k-th element of $v(I_i)$
$c(I)$	Condition of an image I
$\mathcal{S}_{i,k}$	Set of $\{v_k(I) \forall I, c(I) = c_i\}$ for a condition c_i
$\mathcal{T}(v_i)$	Set of *representative tiles* of ViVo v_i
$\mathcal{R}(c_i)$	Set of *representative ViVos* of condition c_i

Table 1. Symbols used in this paper.

In the second step, a feature vector is extracted from each tile, representing its image content. We have conducted experiments using features such as the color structure descriptor (CSD), color layout descriptor (CLD), and homogeneous texture descriptor (HTD). The vector representing a tile using features of, say CSD, is called a *tile-vector* of the CSD. More details are given in Section 4.

The third step derives a set of symbols from the feature vectors of all the tiles of all the images. In text processing, there is a similar issue of representing documents by topics. The most popular method for finding text topics is latent semantic indexing (LSI) [3], which is based on analysis that resembles PCA. Given a set of data points, LSI/PCA finds a set of orthogonal (basis) vectors that best describe the data distribution with respect to minimized L_2 projection error. Each of these basis vectors is considered a topic in the document set, and can be used to group documents by topics.



(a) PCA vs ICA (b) ViVo tile groups

Figure 3. ViVos and their tile groups. Each point corresponds to a tile. (a) Basis vectors ($\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{I}_1, \boldsymbol{I}_2, \boldsymbol{I}_3$) are scaled for visualization. (b) Two tiles groups are shown here. Representative tiles of the two groups are shown in triangles. (Figures look best in color.)

Our approach is similar: We derive a set of symbols by applying ICA or PCA to the feature vectors. Each basis vector found by ICA or PCA becomes a symbol. We call the symbols ***ViVos*** and the set of symbols a *visual vocabulary*.

Figure 3(a) shows the distribution of the tile-vectors of the CSD, projected in the space spanned by the two PCA basis vectors with the highest eigenvalues. The data distribution displays several characteristic patterns—"arms"—on which points are located. None of the PCA basis vectors (dashed lines anchored at $\langle 0,0 \rangle$: $\boldsymbol{P}_1, \boldsymbol{P}_2$) finds these characteristic arms. On the other hand, if we project the ICA basis vectors onto this space (solid lines: $\boldsymbol{I}_1, \boldsymbol{I}_2, \boldsymbol{I}_3$), they clearly capture the patterns in our data. It is preferable to use the ICA basis vectors as symbols because they represent more precisely the different aspects of the data. We note that only three ICA basis vectors are shown because the rest of them are approximately orthogonal to the space displayed.

Relating Figure 3 to our algorithm in Figure 2, each point is a $\tilde{\mathbf{t}}_{\mathbf{i,j}}$ in step 2 of the algorithm. Function *gen_vv()* in step 3 computes the visual vocabulary which is defined according to the set of the ICA basis vectors. Intuitively, an ICA basis vector defines two ViVos, one along the positive direction of the vector, another along the negative direction.

Formally, let $\boldsymbol{T_0}$ be a t-by-d matrix, where t is the number of tiles from all training images, and d is the number of features extracted from each tile. Each row of $\boldsymbol{T_0}$ corresponds to a tile-vector $\tilde{\mathbf{t}}_{\mathbf{i,j}}$, with the overall mean subtracted. Suppose we want to generate m ViVos. We first reduce the dimensionality of $\boldsymbol{T_0}$ from d to $m' = m/2$, using PCA, yielding a t-by-m' matrix $\boldsymbol{T}$. Next, ICA is applied in order to decompose $\boldsymbol{T}$ into two matrices $\boldsymbol{H}_{[t \times m']}$ and $\boldsymbol{B}_{[m' \times m']}$ such that $\boldsymbol{T} = \boldsymbol{HB}$. The rows of $\boldsymbol{B}$ are the ICA basis vectors (solid lines in Figure 3(a)). Considering the positive and negative directions of each basis vector, the m' ICA basis vectors would define $m = 2m'$ ViVos, which are the outputs of the function *gen_vv()*.

How do we determine the number of ViVos? We follow the rule of thumb, and make $m' = m/2$ be the dimensionality which preserves 95 % spread/energy of the distribution.

With the ViVos ready, we can use them to represent an image. We first represent each d-dim tile-vector in terms of ViVos by projecting a tile-vector to the m'-dim PCA space and then to the m'-dim ICA space. The positive and negative projection coefficients are then considered separately, yielding the $2m'$-dim ViVo-vector of a tile. This done by *comp_vivo()* in the fourth step of the algorithm in Figure 2. The $m = 2m'$ coefficients in the ViVo-vector of a tile also indicate the contributions of each of the m ViVos to the tile.

In the fifth and final step, each image is expressed as a combination of its (reformulated) tiles. We do this by simply adding up the ViVo-vectors of the tiles in an image. This yields a good description of the entire image because ICA produces ViVos that do not "interfere" with each other. That is, ICA makes the columns of $\boldsymbol{H}$ (coefficients of the basis vectors, equivalently, contribution of each ViVo to the image content) as independent as possible [8]. Definition 1 summarizes the outputs of our proposed method.

Definition 1 (ViVo and ViVo-vector) *A ViVo is defined by either the positive or the negative direction of an ICA basis vector, and represents a characteristic pattern in image tiles. The ViVo-vector of a tile* $\tilde{\mathbf{t}}_{\mathbf{i,j}}$ *is a vector* $v(\tilde{\mathbf{t}}_{\mathbf{i,j}}) = f_1, \ldots, f_m]$, *where* f_i *indicates the contibutions of the i-th ViVo in describing the tile. The ViVo-vector of an image is defined as the sum of the ViVo-vectors of all its tiles.*

Representative tiles of a ViVo. A ViVo corresponds to a direction defined by a basis vector, and is not exactly equal to any of the original tiles. In order to visualize a ViVo, we represent it by a tile that strongly expresses the characteristics of that ViVo.

We first group tiles that are majorly located along the same ViVo direction together as a "tile group". Formally, let the ViVo-vector of a tile $\tilde{\mathbf{t}}_{\mathbf{i,j}}$ be $v(\tilde{\mathbf{t}}_{\mathbf{i,j}})$=$[f_1, \ldots, f_m]$. We say that the tile $\tilde{\mathbf{t}}_{\mathbf{i,j}}$ *belongs to* ViVo v_k, if the element with largest magnitude is f_k, i.e., $k = \arg\max_{k'} |f_{k'}|$. The *tile group* of a ViVo v_k is the set of tiles that belong to v_k. Figure 3(b) visualizes the tile groups of two ViVos on the 2-D plane defined by the PCA basis vectors ($\boldsymbol{P}_1, \boldsymbol{P}_2$).

The *representative tiles* of a ViVo v_k, $\mathcal{T}(v_k)$, are then selected from its tile group (essentially the tiles at the "tip" of the tile group). The top 5 representative tiles of the two ViVos in Figure 3(b) are shown in light triangles. The top representative tile of ViVo v_k has the maximum $|c_k|$ value among all tiles in v_k's tile group. In Section 5.1, we show the representative tiles of our ViVos and discuss their biological interpretation.

4 Quantitative Evaluation: Classification

The experiments in this section evaluate the combinations of image features and ViVo generation methods for ViVo construction. In these experiments, our goal is to find the

Feature	Dim.	Accuracy	Std. dev.
Original CSD	512	0.838	0.044
14 ViVos from CSD	14	0.832	0.042
12 ViVos from CSD	12	0.826	0.038
Original CLD	24	0.346	0.049
24 ViVos from CLD	24	0.634	0.023
Original HTD	124	0.758	0.048
12 ViVos from HTD	12	0.782	0.019

Table 2. Classification accuracies for combinations of feature and ViVo set size. All ViVo sets reported here are based on ICA.

best representation of the images in the symbolic space and ensure that classification accuracies obtained using these symbols are close to the best accuracy that we could obtain with the raw features.

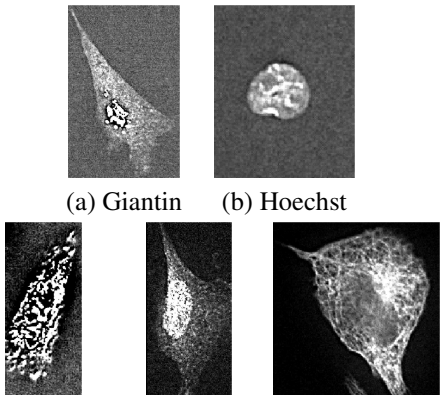
Biologists have chosen experimental conditions which correspond to different stages of the biological process. Thus, a combination that successfully classifies images is also likely to be a good choice for other analyses, such as the qualitative analyses described in Section 5, where we investigate the ability of the visual vocabulary to reveal biologically meaningful patterns.

Classification experiments were performed on two datasets: one dataset of 433 retinal micrographs, and another dataset of 327 fluorescence micrographs showing subcellular localization patterns of proteins in CHO cells. In the following, we refer to the datasets by their cardinality: the 433 dataset and the 327 dataset.

4.1 Classification of Retinal Images

The 433 dataset contains retinal images from the UCSB BioImage database (http://bioimage.ucsb.edu/), which contains images of retinas detached for either 1 day (label `1d`), 3 days (`3d`), 7 days (`7d`), 28 days (`28d`), or 3 months (`3m`). There are also images of retinas after treatment, such as reattached for 3 days after 1 hour of detachment (`1h3dr`), reattached for 28 days after 3 days of detachment (`3d28dr`), or exposed to 70 % oxygen for 6 days after 1 day of detachment (`1d6dO2`), and images of control tissues (`n`) [6, 13, 12].

We experimented extensively with different features and vocabulary sizes. Features are extracted separately for the red and green channels and then concatenated. The channels show the staining by two antibodies: anti-rod opsin (red) and anti-GFAP (green). The number of ViVos should be small, as large vocabularies contain redundant terms and become difficult for domain experts to interpret. Preserving 95 % of the energy resulted in 14, 24, and 12 ViVos for CSD, CLD, and HTD, respectively. The classification accuracies, reported in Table 2, are from 5-fold cross-validation



(a) Giantin (b) Hoechst (c) LAMP2 (d) NOP4 (e) Tubulin

Figure 4. Examples from the dataset of 327 fluorescence micrographs of subcellular protein localization patterns. The images have been enhanced in order to look better in print.

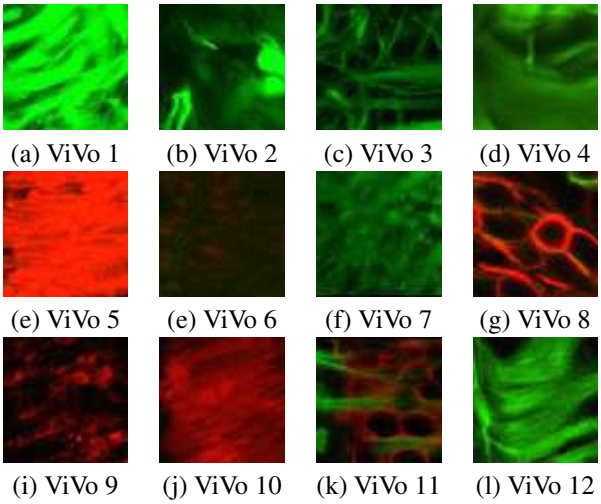
using SVM [2] with linear kernels. SVM with polynomial kernels and a k-NN ($k = 1$, 3, or 5) classifier produced results that were not significantly different. ViVos from CSD perform significantly better than ViVos from CLD ($p < 0.0001$) and also significantly better than ViVos from HTD ($p = 0.0492$). Further, manual inspection of HTD ViVos did not reveal better biological interpretations.

Two of the 14 CSD ViVos were removed because none of the images had high coefficients for them. Those two ViVos had no interesting biological interpretation either. As expected, removing these two ViVos (using only 12 ViVos) resulted in insignificantly ($p = 0.8187$) smaller classification accuracy compared to the 14 CSD ViVos (Table 2). The difference from the original CSD features is also insignificant ($p = 0.6567$). We therefore choose to use the 12 CSD ViVos as our visual vocabulary.

4.2 Classification of Subcellular Protein Localization Images

In order to assess the generality of our visual vocabulary approach, we also applied our method to classify 327 fluorescence microscopy images of subcellular protein localization patterns [1]. Example micrographs depicting the cell DNA and four protein types are shown in Figure 4. We partitioned the data set into training and test sets in the same way as Boland et al. [1].

We note that although these images are very different from the retinal images, the combination of CSD and ICA still classifies 84 % of the images correctly. The 1-NN classifier achieves 100 % accuracy on 3 classes: Giantin, Hoechst, and NOP4. The training images of class LAMP2 in the data set have size 512-by-512, which is different from that of the others, 512-by-382. Due to this discrepancy, class LAMP2 is classified at 83 %, and around half of Tubulin images are classified as LAMP2.



(a) ViVo 1 (b) ViVo 2 (c) ViVo 3 (d) ViVo 4

(e) ViVo 5 (e) ViVo 6 (f) ViVo 7 (g) ViVo 8

(i) ViVo 9 (j) ViVo 10 (k) ViVo 11 (l) ViVo 12

Figure 5. Our visual vocabulary. The vocabulary is automatically constructed from a set of images. Please see the electronic version of the article for color images.

To summarize, our classification experiments show that the symbolic ViVo representation captures well the contents of microscopy images of two different kinds. Thus, we are confident that the method is applicable to a wider range of biomedical images.

5 Qualitative Evaluation: Data Mining Using ViVos

Deriving a visual vocabulary for image content description opens up many exciting data mining applications. In this section, we describe our proposed methods for answering the three problems we introduced in Section 1. We first discuss the biological interpretation of the ViVos in Section 5.1 and show that the proposed method correctly summarizes a biomedical image automatically (Problem 1). An automated method for spotting differential patterns between classes is introduced in Section 5.2 (Problem 2). Several observations on the class-distinguishing patterns are also discussed. Finally, in Section 5.3, we describe a method to automatically highlight interesting regions in an image (Problem 3).

5.1 Biological Interpretation of ViVos

The representative tiles of ViVos 2, 3, 4, 7, and 12 shown in Figure 5 demonstrate the hypertrophy of Müller cells. These ViVos correctly discriminate various morphological changes of Müller cells. The green patterns in these representative tiles is due to staining produced by immunohistochemistry with an antibody to GFAP, a protein found in glial cells (including Müller cells). Our visual vocabulary also captures the normal expression of GFAP in the inner retina, represented by ViVo 1. The Müller cells have been shown to hypertrophy following experimental retinal detachment. Understanding how they hypertrophy and change morphology is important in understanding how these cells can ultimately form glial scars, which can inhibit a recovery of the nervous system from injury.

Also, our ViVos correctly place tiles into different groups, according to the different anti-rod opsin staining which may due to functional consequences following injury. In an uninjured retina, anti-rod opsin (shown in red) stains the outer segments of the rod photoreceptors, which are responsible for converting light into an electrical signal and are vital to vision. ViVos 5 and 10 show a typical staining pattern for an uninjured retina, where healthy outer segments are stained. However, following detachment or other injury to the retina, outer segment degeneration can occur (ViVo 9). Another consequence of retinal detachment can be a re-distribution of rod opsin from the outer segments of these cells to the cell bodies (ViVo 8).

As described above, both the re-distribution of rod opsin and the Müller cell hypertrophy are consequences of retinal detachment. It is of interest to understand how these processes are related. ViVo 11 captures the situation when the two processes co-occur. Being able to sample a large number of images that have these processes spatially overlapping will be important to understanding their relationship. ViVo 6 is rod photoreceptor cell bodies with only background labeling.

5.2 Finding Most Discriminative ViVos

We are interested in identifying ViVos that show differences between different retinal experimental conditions, including treatments. Let images $\{I_1, \ldots, I_n\}$ be the training images of condition c_i. Suppose that our analysis in Section 3 suggests that m ViVos should be used. Following the algorithm outlined in Figure 2, we can represent an image I as an m-dimensional ViVo-vector $v(I)$. The k-th element of a ViVo-vector, $v_k(I)$, gives the expression level of ViVo v_k in the image I. Let $\mathcal{S}_{ik}$=$\{v_k(I_1), \ldots, v_k(I_n)\}$ be a set that contains the k-th elements of all image ViVo-vectors in condition c_i.

To determine if a ViVo v_k is a discriminative ViVo for two conditions c_i and c_j, we perform an analysis of variance (ANOVA) test, followed by a multiple comparison [11]. If the 95% confidence intervals of the *true* means of $\mathcal{S}_{ik}$ and $\mathcal{S}_{jk}$ do not intersect, then the means are not significantly different, and we say that ViVo v_k discriminates conditions c_i and c_j, i.e., v_k is a discriminative ViVo for c_i and c_j. The separation between $\mathcal{S}_{ik}$ and $\mathcal{S}_{jk}$ indicates the "discriminating power" of ViVo v_k.

Figure 6 shows the conditions as boxes and the discriminative ViVos on edges connecting pairs of conditions that are of biological interest. ViVos 6 and 8 discriminate `n` from `1d` and `1d` from `3d`. The two ViVos represent rod photoreceptor cell bodies with only background labeling

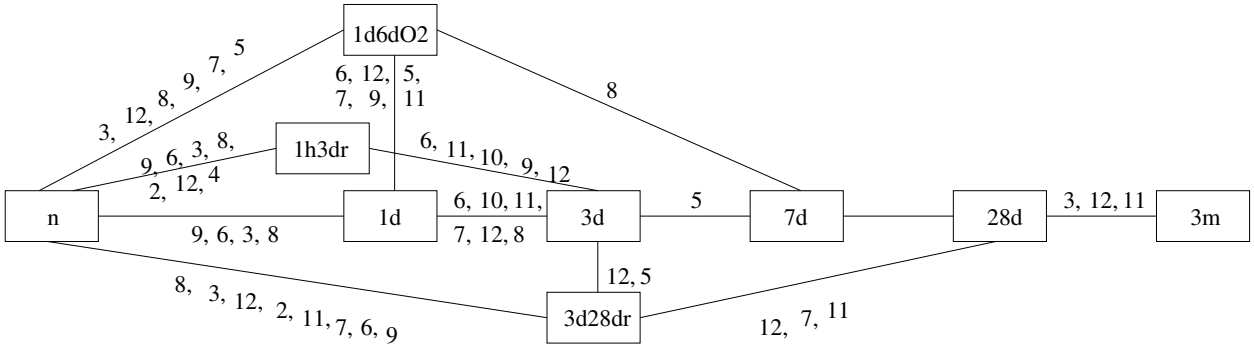


Figure 6. Pairs of conditions and the corresponding discriminative ViVos. There is an edge in the graph for each pair of conditions that is important from a biological point of view. The numbers on each edge indicate the ViVos that contribute the most to the differences between the conditions connected by that edge. The ViVos are specified in the order of their discriminating power.

and with redistribution of rod opsin, respectively, indicating that the redistribution of rod opsin is an important effect in the short-term detachment. Note also that ViVo 8 distinguishes `1d6dO2` from `7d`. This suggests that there are cellular changes associated with this oxygen treatment, and the ViVo technique can be used for this type of comparison.

The ViVos that represent Müller cell hypertrophy (ViVo 2, 3, 4, 7, and 12) discriminate `n` from all other conditions. We note that ViVo 1, which represents GFAP labeling in the inner retina in both control (`n`) and detached conditions, is present in all conditions, and therefore cannot discriminate any of the pairs in Figure 6. In addition, several ViVos discriminate between `3d28dr` and `28d`, and `1h3dr` and `3d`, suggesting cellular effects of the surgical procedure. Interestingly, there are no ViVos that discriminate between `7d` and `28d` detachments, suggesting that the effects of long-term detachment have occurred by 7 days.

Although these observations are generated automatically by an unsupervised tool, they correspond to observations and biological theory of the underlying cellular processes.

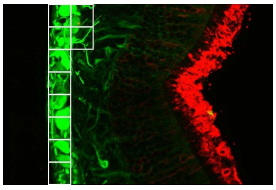
5.3 Highlighting Interesting Regions by ViVos

In this section, we propose a method to find class-relevant ViVos and then use this method to highlight interesting regions in images of a particular class.

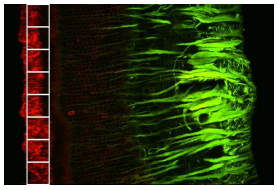
In order to determine which condition a ViVo belongs to, we examine its representative tiles and determine the most popular condition among them (majority voting). We define the condition of a tile to be that of the image from which it was extracted, i.e., $c(\tilde{\mathbf{t}}_{\mathbf{i,j}}) = c(I_i)$. Intuitively, for a ViVo, the more its representative tiles are present in images of a condition, the more relevant the ViVo is to that condition.

Formally, the set $\mathcal{R}(c_k)$ of representative ViVos of a condition c_k is defined as

$$\mathcal{R}(c_k) = \left\{ v_r : \sum_{t \in \mathcal{T}(v_r)} I(c(t) = c_k) > \sum_{t \in \mathcal{T}(v_r)} I(c(t) = c_q), \forall c_q \neq c_k \right\},$$



(a) ViVo 1 highlighted



(b) ViVo 10 highlighted

Figure 7. Two examples of images with ViVo-annotations (highlighting) added. (a) GFAP-labeling in the inner retina (28d); (b) rod photoreceptor recovered as a result of reattachment treatment (3d28dr).

where t is a tile, and $I(p)$ is an indicator function that is 1 if the predicate p is true, and 0 otherwise. The representative ViVos of a condition c_k can be used to annotate images of that particular condition in order to highlight the regions with potential biological interpretations.

Figure 7(a) shows an annotated image of a retina detached for 28 days. The GFAP labeling in the inner retina is highlighted by ViVo 1 (see Figure 5(a)).

Figure 7(b) shows an annotated image of a retina detached for 3 days and then reattached for 28 days. The annotation algorithm highlighted the outer segments of the rod photoreceptors with ViVo 10 (see Figure 5(j)). As pointed out in Section 5.1, ViVo 10 represents healthy outer segments. In the retina depicted in Figure 7(b), the outer segments have indeed recovered from the degeneration caused by detachment. This recovery of outer segments has previously been observed [6], and confirms that ViVos can recognize image regions that are consistent with previous biological interpretations.

6 Conclusion

Mining biomedical images is an important problem because of the availability of high-throughput imaging, the applicability to medicine and health care, and the ability of images

to reveal spatio-temporal information not readily available in other data sources such as genomic sequences, protein structures and microarrays.

We focus on the problem of describing a collection of biomedical images succinctly (Problem 1). Our main contribution is to propose an automatic, domain-independent method to derive meaningful, characteristic tiles (*ViVos*), leading to a *visual vocabulary* (Section 3). We apply our technique to a collection of retinal images and validate it by showing that the resulting ViVos correspond to biological concepts (Section 5.1).

Using ViVos, we propose two new data mining techniques. The first (Section 5.2) mines a large collection of images for patterns that distinguish one class from another (Problem 2). The second technique (Section 5.3) automatically highlights important parts of an image that might otherwise go unnoticed in a large image collection (Problem 3).

The conclusions are as follows:

- *Biological Significance:* The terms of our visual vocabulary correspond to concepts biologists use when describing images and biological processes.
- *Quantitative Evaluation:* Our ViVo-tiles are successful in classifying images, with accuracies of 80 % and above. This gives us confidence that the proposed visual vocabulary captures the essential contents of biomedical images.
- *Generality:* We successfully applied our technique to two diverse classes of images: localization of different proteins in the retina, and subcellular localization of proteins in cells. We believe it will be applicable to other biomedical images, such as X-ray images, MRI images, and electron micrographs.
- *Biological Process Summarization:* Data mining techniques can use the visual vocabulary to describe the essential differences between classes. These techniques are unsupervised, and allow biologists to screen large image databases for interesting patterns.

Acknowledgements. We would like to thank Geoffrey P. Lewis from the laboratory of Steven K. Fisher at UCSB and Robert F. Murphy from CMU for providing the retinal micrographs and subcellular protein images, respectively. This work was supported in part by NSF grants no. IIS-0205224, INT-0318547, SENSOR-0329549, EF-0331657, IIS-0326322, EIA-0080134, DGE-0221715, and ITR-0331697; by PITA and partnership between CMU, Leigh Univ. and DCED; by donations from Intel and NTT; and a gift from Northrop-Grumman Corporation.

References

[1] M. V. Boland, M. K. Markey, and R. F. Murphy. Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. *Cytometry*, 3(33):366–375, 1998.

[2] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[3] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci. Technol.*, 41(6):391–407, 1990.

[4] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge. Recognizing faces with PCA and ICA. *Comp. Vis. and Image Understanding*, (91):115–137, 2003.

[5] P. Duygulu, K. Barnard, N. Freitas, and D. A. Forsyth. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In *Proc. ECCV*, volume 4, pages 97–112, 2002.

[6] S. K. Fisher, G. P. Lewis, K. A. Linberg, and M. R. Verardo. Cellular remodeling in mammalian retina: Results from studies of experimental retinal detachment. *Progress in Retinal and Eye Research*, 24:395–431, 2005.

[7] Y. Hu and R. F. Murphy. Automated interpretation of subcellular patterns from immunofluorescence microscopy. *Journal of Immunological Methods*, 290:93–105, 2004.

[8] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley and Sons, 2001.

[9] J. Jeon and R. Manmatha. Using maximum entropy for automatic image annotation. In *Proc. CIVR*, pages 24–32, 2004.

[10] I. T. Jolliffe. *Principal Component Analysis*. Springer, 2002.

[11] A. J. Klockars and G. Sax. *Multiple Comparisons*. Number 07-061 in Sage Univ. Paper series on Quantitative Applications in the Social Sciences. Sage Publications, Inc., 1986.

[12] G. Lewis, K. Talaga, K. Linberg, R. Avery, and S. Fisher. The efficacy of delayed oxygen therapy in the treatment of experimental retinal detachment. *Am. J. Ophthalmol.*, 137(6):1085–1095, June 2004.

[13] G. P. Lewis, C. S. Sethi, K. A. Linberg, D. G. Charteris, and S. K. Fisher. Experimental retinal detachment: A new perspective. *Mol. Neurobiol.*, 28(2):159–175, Oct. 2003.

[14] J.-H. Lim. Categorizing visual contents by matching visual "keywords". In *Proc. VISUAL*, pages 367–374, 1999.

[15] W.-Y. Ma and B. S. Manjunath. A texture thesaurus for browsing large aerial photographs. *Journal of the American Society for Information Science*, 49(7):633–648, 1998.

[16] B. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG-7*. Wiley, 2002.

[17] A. Mojsilović, J. Kovačević, J. Hu, R. J. Safranek, and S. K. Ganapathy. Matching and retrieval based on the vocabulary and grammar of color patterns. *IEEE Trans. Image Proc.*, 9(1):38–54, 2000.

[18] R. F. Murphy. Automated interpretation of protein subcellular location patterns: Implications for early cancer detection and assessment. *Annals N.Y. Acad. Sci.*, 1020:124–131, 2004.

[19] R. F. Murphy, M. Velliste, and G. Porreca. Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. *Journal of VLSI Signal Processing*, 35:311–321, 2003.

[20] C. P. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Proceedings of the Sixth International Conference on Computer Vision (ICCV'98)*, volume 2, pages 555–562, January 4-7 1998.

[21] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proc. ICCV*, volume 2, pages 1470–1477, 2003.

[22] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–96, 1991.