CELL PATTERN CLASSIFICATION OF INDIRECT IMMUNOFLUORESCENCE IMAGES

Shahab Ensafi

(M.Sc. (Electrical and Computer Engineering))

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2016

DECLARATION

I hereby declare that the this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

.....Shahab.....

SHAHAB ENSAFI

13 July 2016

Acknowledgments

There are many people whom I wish to thank for the help and support they have given me throughout my PhD studies. Foremost, I would like to express my sincere gratitude to my supervisors Prof. Ashraf Kassim, Prof. Tan Chew Lim and Dr. Shijian Lu. I thank them for their patience, care and encouragement that carried me on through all the difficult times, and for their invaluable guidance, insights and suggestions which helped me throughout this work. Besides my advisors, I would like to thank A*STAR SINGA institute which provides the funding and scholarship of my PhD study.

I also take this opportunity to thank all my friend in Singapore, whose presences and fun-loving spirits made this journey all the more memorable. Last but not the least, I would like to thank my family, maman Shahin and Farah, baba Alireza and Elyas, Shaghayegh, Saeid and Shayan for always being there when I needed them most, and for supporting me through all these years. Thanks to my little angel Elsa who made this journey more sweet and pleasant. Most importantly to my amazing wife and best friend Mahsa, your love and patience made this possible and truly the words are inadequate to thank you. This thesis is dedicated to you.

Contents

Summary

Autoimmune Diseases (ADs) develop when the immune system of the body treats some healthy cells as 'foreigners' and attacks them. ADs are among the top ten leading causes of death in children and women in all age groups up to 64 years. Indirect Immunofluorescence (IIF) test is used to capture Human Epithelial Type-2 (HEp-2) cells' images, where the different staining patterns of HEp-2 cells indicate the stage and type of the AD.

Automated classification of Hep-2 cells has attracted much research interest in recent years. Despite the extensive recent work that has been done in this field, there are still many challenges to be overcome. This thesis presents some efficient and practical methodologies that overcome the current limitations of state-of-the-art HEp-2 cells classification methods. The key contributions include:

As the first step of the cell images classification approaches, features are extracted from the image patches. Because of the sparse nature of the image patches, a dictionary learning and sparse coding scheme is then used. A challenging problem of these schemes is the choice of dictionary size, which should neither be too large to increase the complexity, nor too small to increase the reconstruction error. Currently, this parameter is selected manually in the literature. In this thesis, a

non-parametric Bayesian method is proposed to estimate the optimal dictionary size and simultaneously calculate the dictionary.

The high-dimensional dictionaries together with the high-dimensional features not only reduce classification speed significantly, but also add to the computational requirements. The complexities of these above methods make them impractical for realizing real time systems to be used by physicians/clinicians. In this thesis, we propose an adaptive distributed dictionary learning method which divides the dictionary to N sub-dictionaries and build a network where each node is responsible of updating its own sub-dictionary. This method addresses the HEp-2 cell classification problem in a computationally efficient and less memory intensive way compared to the other methods.

Another challenge is the huge number of overlapping image patches and consequently, the need for tuning their sizes and shapes. The number of patches exponentially increases with the image resolution, resulting in a dramatic increase in computational complexity. On the other hand, the amount of information contained in each patch highly depends on the size of the patch and the dataset. To overcome these limitations in this thesis a superpixel approach is proposed. This method can extract the image patches (superpixels) with different sizes and shapes corresponding to the underlying patterns of the image.

In above mentioned approaches, the resulted sparse codes are not necessarily discriminative. Additionally, concatenation of the different input features may increase the redundancy and reconstruction error. To overcome these problems, a feature fusion technique is suggested. The method forces the sparse codes to have the same pattern for each class, resulting in more discriminative feature vectors.

List of Figures

List of Tables

Nomenclature

C	Total number of classes. 107
J	Cost function. 72
$\mathbf{B}$	Binary matrix. 60–62
$\mathbf{D}$	The dictionary matrix with K words. 32, 60, 70
$\mathbf{F}$	The feature matrix. 32, 60, 70, 92
Π	Projection operator. 77
$\mathbb{R}$	Euclidean space. 19, 32
$\mathbf{Z}$	Sparse codes. 32, 60, 62
β	Regularization term. 70
λ	Regularization term. 70
$\mathcal{D}$	Feature vector dimension. 32, 34, 59, 60
$\boldsymbol{\nu}^o$	Optimal solution. 74

Acronyms

AD Autoimmune Disease. 1, 8, 10, 11

ADDL Adaptive Distributed Dictionary Learning. 68, 69, 80–85, 131

ANA Antinuclear Antibody. 9

BoW Bag of Words. 12, 24, 26, 32, 56, 67, 129

CNN Convolutional Neural Network. 16–18

DCT Discrete Cosine Transform. 12, 15

DoG Difference of Gradient. 12

GLCM Grey Level Co-occurrence Matrix. 12

HEp-2 Human Epithelial Type-2. 1, 9–12, 15

HoG Histogram of Gradient. 12, 17, 132

IBP Indian Buffet Process. 59, 60, 62, 130

ICA Independent Component Analysis. 17

IIF Indirect Immunofluorescence. 9, 10

LBP Local Binary Pattern. 2, 11, 12, 15, 132

MCA Mean Class Accuracy. 45, 64, 92, 119

MCMC Markov Chain Monte Carlo. 62

SBoW Sparse Bag of Words. 26, 46, 95

SCS Sparse Coding of Superpixels. 86

SIFT Scale-invariant Feature Transform. 2, 4, 12, 24, 26, 31, 34, 45, 53, 55, 56, 58, 88, 92, 129, 132

SNPB Sparse Non-Parametric Bayesian. 57, 58, 65

SURF Speeded Up Robust Features. 4, 24, 26, 31, 34, 45, 53, 55, 56, 58, 88, 92, 129

SVM Support Vector Machine. 12, 26, 36, 37, 45, 53, 59, 88, 101

VQ Vector Quantization. 21

Chapter 1

Introduction

Autoimmune Diseases (ADs), which arise from abnormal immune responses of the body against normally present substances and tissues, are in the list of top mortality causes according to the American Autoimmune Related Diseases Association (AARDA). Early diagnosis of ADs plays a significant role in its treatment and the demand for methods and procedures for fast, low-cost and repeatable diagnosis has become more and more indispensable [González-Buitrago and González, 2006]. Currently, there is an exponentially increasing demand for AD tests while there is a lack of certified physicians to perform the tests. Another issue that makes the AD diagnosis an even more challenging problem is the repeatability of the test across different physicians. To address these challenges, automatic classification of the Human Epithelial Type-2 (HEp-2) cells has been attracting much research interest in recent years.

Dictionary learning and sparse coding scheme is a main stage of the HEp-2 cell classification (see Chapter 3). In dictionary learning methods, the input images

are divided into small regions called 'patches'. Then the features of the patches are calculated and fed to an intervening procedure to combine them and provide the final feature vectors for classifiers. Despite the extensive recent work that has been done in this field, there are still many fundamental challenges to overcome. For example, fine tuning the parameters (e.g. the number of patches and dictionary dimension) and fusing the input features are open problems, that are highly correlated to the obtained classification performance. Another challenging issue faced is the large amount of image patches with different sizes and shapes which in turns increase the complexity of the dictionary learning process. Addressing these issues, form the basis of the research work presented in this thesis, which are presented in the following sections.

1.1 Motivation

Most works in HEp-2 classification (see Section 2.2 for details) are based on artificial features such as Scale-invariant Feature Transform (SIFT), Local Binary Pattern (LBP) and histograms that are extracted from different image patches. These features have some parameters (such as size and number of patches, number of histogram bins, smoothing parameters, etc.) that are needed to be chosen manually. This can potentially affect the final classification performance and the parameter tuning can be arduous. Moreover, the prior knowledge of the intensity levels (*positive* and *intermediate*) and color data (RGB) of the input images are very useful to the cell classification but largely ignored by previous methods.

The sparse nature of patch-based image classification (see Section 2.2) leads to

perform sparse coding and dictionary learning schemes on HEp-2 cells which are used widely in the literature. However, one critical parameter is the dictionary size and the optimal size depends heavily on the data as well as employing the visual features. A dictionary is called *critically complete* when the dimension of the features and dictionary are close to each other. On the other hand, biologically inspired *over complete* dictionary with dimension much larger than the feature dimension often gives better classification accuracy [Rehn and Sommer, 2007] and is therefore widely adopted. Nevertheless, the high-dimensional dictionaries use the high-dimensional features and codes which often reduce the classification speed significantly. Thus, the objective is to build a dictionary with optimal dimension. However, learning the dictionaries and working with large matrices is a burden of load on CPU and memory of a machine which requires high memory and computational resources. Therefore, splitting the dictionary into sub-dictionaries and learning them in different machines can be beneficial for dictionary learning procedure.

The other constraint of patch-based image classification technique is the high computational cost due to a huge amount of overlapped image patches to be processed and the tedious parameter tuning (for patch size, scanning step size, etc.) for optimal cell classification performance. Therefore, the question is how the patch sizes, shapes and positions can be found automatically such that decrease the number of patches in one hand and increase the final classification accuracy in other hand.

The majority of existing dictionary learning methods, can handle only single source of data but fusion of information from different sensor modalities can be

more robust to single sensor failure. In the sparse coding method, the feature fusion is imposed by concatenating all of features in one vector. The dimension of this vector is high and suffers from curse-of-dimensionality while it does not even contain the valuable information of correlations between feature types. Multiple features can be combined using joint sparsity priors which makes them suitable for reconstructing samples that originate from different sources. Therefore, combining different features should lead to better classification results.

Chapter 2 presents a more detailed review of the current methods outlined in this section.

1.2 Main Contributions

This thesis presents some new methods and solutions which address some of the limitations of current and related approaches to the problem of HEp-2 cell image classification as outlined briefly in Section 1.1. The overview of the proposed methods is shown in Fig. 1.1. As evident in Fig. 1.1, a cell extraction method is proposed to obtain the cells from the specimen images following the feature extraction, different dictionary learning procedures and pooling strategies which are discussed in corresponding chapters of the thesis. Here is a summary of the main contributions, further details of which are presented in the indicated chapters:

- A HEp-2 cell image classification technique that exploits the sparse coding of the visual features together with the Bag of Words model (SBoW) is proposed (see Chapter 3). In particular, Speeded Up Robust Features (SURF) and SIFT features are specially integrated to work in a complemen-

tary fashion, helping to greatly improve the cell classification accuracy. To provide final feature vector, a hierarchical max-pooling method is proposed that aggregates the local sparse codes in different layers. Furthermore, the correlation between the iteration of the dictionary learning and the cell classification performance is investigated and the optimal iteration is identified with superior cell classification accuracy.

- The dimension of the dictionary in SBoW model is essentially important where a non-parametric Bayesian model is proposed to learn the optimal dimension with little human intervention. In particular, the 'non-parametric' here means that the dimension of the dictionary can intuitively extend to infinity at the beginning of the learning process and leads to the correct and most efficient value at the end (see Chapter 4).

- An adaptive distributed dictionary learning is proposed (see Chapter 5) which addresses the HEp-2 cell classification problem in a computationally efficient and less memory intensive way compared to the other methods where, the dictionary matrix and the coding vector are partitioned into N blocks/nodes and each block is associated with a sub-dictionary and a sub-vector. Each node is connected to a number of neighboring nodes sharing their information to update the sub-dictionaries. Essentially, we propose to combine the information of neighboring nodes in an adaptive way which enables the nodes to learn about the usefulness of the information received from their neighbors which helps the nodes to ignore misleading information.

- A novel superpixel based HEp-2 cell classification technique based on the

sparse coding scheme is proposed (see Chapter 6). To the best of our knowledge, this is the first work that uses superpixels to guide the selection of the right image patches to contain more 'informative' features. Additionally, 'extended superpixels' are designed by dilating the boundary of each superpixel to capture more discriminative gradient information across the boundaries of the HEp-2 cell.

- Features fusion, rather than solely concatenating them, and building a multi-modal dictionary are investigated (see Chapter 7). To utilize information fusion between feature modalities, an algorithm is designed in which sparse codes of each sample from all modalities share the same sparsity pattern. The contribution of this work is two-fold. First, we propose a new framework for multi-modal fusion at the feature level. Second, we impose an additional constraint on consistency of sparse coefficients among different modalities of the same class.

Some concluding remarks and suggestions for further extending this research are presented in Chapter 8.

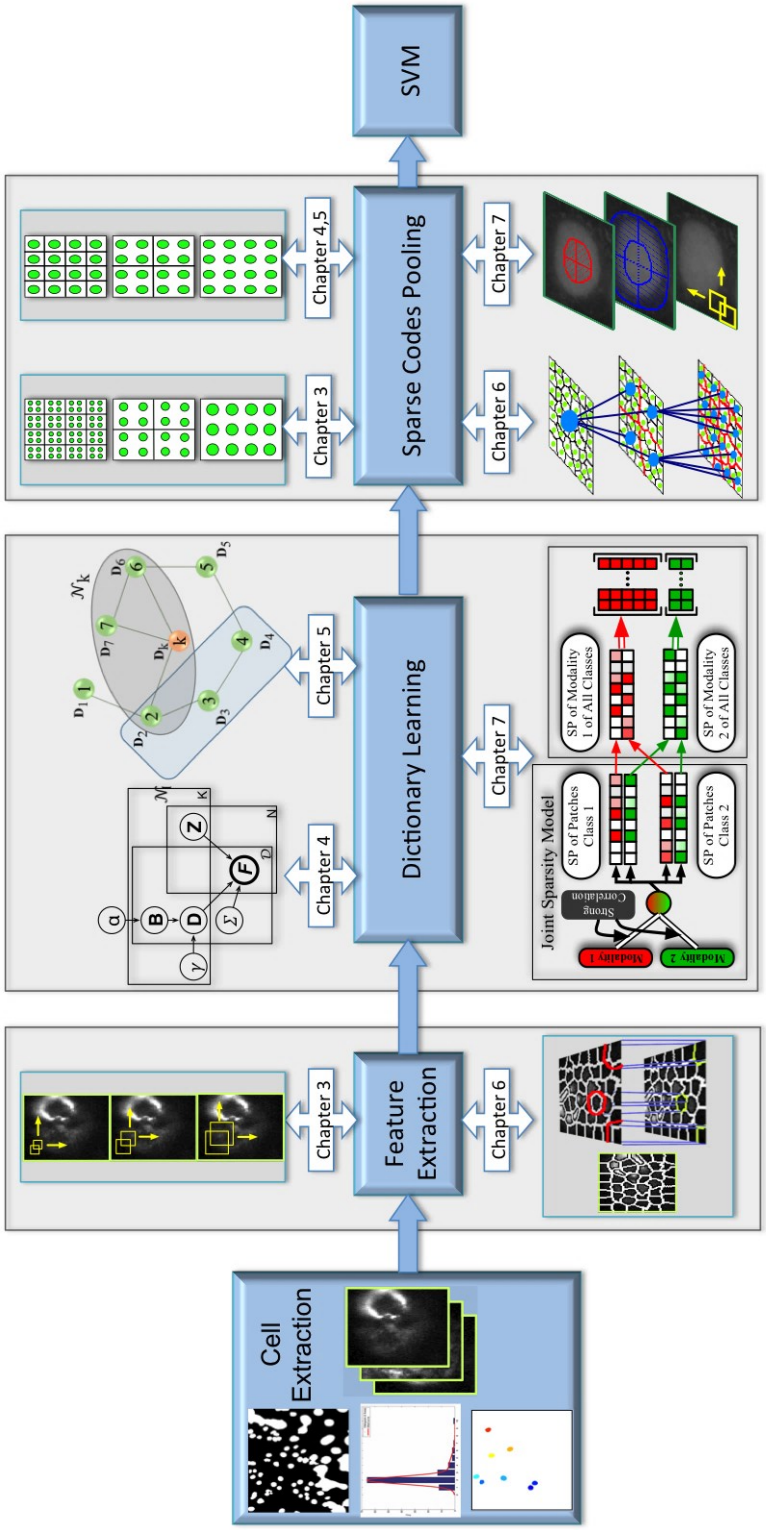


Figure 1.1: The overview of the proposed methods and cantoning chapters in this thesis.

Chapter 2

Background and Literature Review

The basic element of the body's immune system is a 'Y' shape protein named 'antibody', which is produced by the plasma cells. The main role of antibodies is to identify and mark the molecules of harmful agents, called 'antigens'. Antigens are foreign substances from the environment, such as chemicals, bacteria, viruses, or pollen. In particular, the antibody uses its Y-shape tips to bind to the antigen and tags it for neutralization by the other parts of the immune system. Sometimes, the antibody may even neutralize its target directly; for instance by blocking a part of a microbe that is essential for its invasion and survival [Mian et al., 1991].

When the immune system fails to recognize a body's normal protein as 'self', it produces another type of antibody, called 'autoantibody', directed against that protein. This response of the immune system against individual's own tissues is called 'autoimmunity' and the related diseases are named ADs. ADs are broadly

classified into two categories; *systemic*, which are not tissue specific (e.g. rheumatoid and vasculitis diseases) and *local* syndromes which affect a specific organ or tissue (e.g. Diabetes and thyroiditis) [Cotsapas and Hafler, 2013].

Diagnosis of immune disorders differs for each class. In *systemic* disorders specific autoantibodies can be detected by serological assays (the scientific study of serum). *Localized* disorders are best diagnosed by *immunofluorescence* (see Section 2.1) of biopsy specimens. In both cases, the levels of autoantibodies are measured to determine the progress of the disease.

Antinuclear Antibodies (ANAs), which are found in many disorders including autoimmunity, cancer and infection, are kind of antibodies that bind to contents of the cell nucleus. By screening the blood serum, presence of ANA can be confirmed which in turn leads to diagnosis of some autoimmune disorders. According to American College of Rheumatology, the golden standard test for detecting and qualifying ANAs is called Indirect Immunofluorescence (IIF) which uses the HEp-2 tissue.

2.1 Indirect Immunofluorescence Imaging

Immunofluorescence is an imaging technique which uses fluorescence microscope on microbiological samples that are stained with fluorescent chemical compound. This technique can be broadly divided into two categories, namely direct and indirect.

The direct immunofluorescence uses a single antibody linked chemically to a fluorphore (the fluorescent chemical compound). This antibody detects the

antigen and binds to it which enables the fluorescent microscope to capture the specific wavelength of emitted light excited from fluorphore. This technique has two main limitations. First, due to use of a single antibody, it is less sensitive and may result in false negatives. Second, many single antibodies are needed in the experiment, which makes it very expensive.

The IIF uses two antibodies instead of single one, where the first antibody is unlabeled and binds to the target antigen. The second antibody, labeled with fluorphore, detects the first antibody and binds to it. One of the good properties of IIF is that multiple secondary antibodies can bind to the primary one and amplify the emitted light for each antigen, which results in high contrast of the captured images [Storch, 2000].

The HEp-2 cell is a protein that contains hundreds of antigens used as an ideal substrate for the IIF test. Antibodies are first stained in HEp-2 tissue and then bound to a fluorescent chemical compound. Depending on the antibody present in the blood serum and the localization of the antigen in the cell, the patterns of fluorescence will be seen on the HEp-2 cells [González-Buitrago and González, 2006]. These patterns are then classified to diagnose ADs. The patterns' characteristics and their relations to specific ADs are studied in Section 3.3.

Image quality variation makes interpretation of fluorescence patterns very challenging. To make the pattern interpretation more consistent, automated methods for classifying the cells are essential.

The imaging of the IIF test consists of five stages [Hiemann et al., 2006], starting with image acquisition with autofocus to reduce Photobleaching effects [Soda et al., 2006]. The second stage involves automated cell segmentation using meth-

ods such as the similarity based watershed and adaptive edge-based segmentation [Huang et al., 2008a, Huang et al., 2008b]. This is followed by the mitotic cell segmentation stage which has been investigated using morphological and textural features and LBP [Foggia et al., 2010]. The fourth stage further classifies intensity level images into three classes, namely, *negative*, *intermediate* and *positive* intensities [Soda and Iannello, 2006]. Finally, the last stage classifies the cell staining patterns into several classes corresponding to different ADs.

2.2 Literature Review

A number of HEp-2 cell classification techniques have been reported in recent years. The technique by [Perner et al., 2002] is one of the earliest methods that handles the HEp-2 cell classification problem, where Otsu's global thresholding [Otsu, 1975] is used for cell segmentation and texture features are exploited for classification. [Huang et al., 2012] utilized the texture and statistical features and classified the cells using Self-Organizing Maps. [Soda and Iannello, 2009] aggregated the binary classifiers on spectral textural features and introduced a reliability measure of the classification. Techniques on intensity level and staining pattern classifications have also been reported by [Hiemann et al., 2007, Soda et al., 2009, Sack et al., 2003].

Most works described above use their own datasets, which make a fair comparison of different methods a nearly impossible task. The need for effective benchmarking led to the first publicly available dataset "MIVIA HEp-2 images dataset" [Foggia et al., 2013] referred to as ICPR2012 dataset as it was released

for the HEp-2 Cells Classification Contest at the 2012 International Conference on Pattern Recognition. At the 2013 International Conference on Image Processing, an expanded dataset which is referred to as the ICIP2013 dataset[1] was introduced (see Section 3.3).

Several attempts have been made to facilitate the automatic HEp-2 cell classification by evaluating on these datasets. Different classifiers are designed including k-NN, Random Forest, naïve Bayes, etc. However, Support Vector Machine (SVM) is the most used classifier in the literature. It is shown in the literature that the choice of classifier does not affect the final classification result as much as the type of features selected [Han et al., 2014, Foggia et al., 2014].

The features used in the literature can be categorized as follows:

- *intensity-based*: Grey Level Co-occurrence Matrix (GLCM), statistical and morphological features including area, convex-hull, eccentricity, roundness, etc.
- *Feature engineering*: LBP, Discrete Cosine Transform (DCT), Difference of Gradient (DoG), Histogram of Gradient (HoG), SIFT, Gabor wavelet, etc.
- *Feature Learning*: Bag of Words (BoW), unsupervised Dictionary Learning and Deep neural networks feature descriptors.

2.2.1 intensity-based Features

Table. 2.1 summaries some methods which use intensity-based features. There are three main drawbacks to these approaches. First, extracting large number of

[1] `http://i3a2014.unisa.it/?page_id=126`

Table 2.1: Intensity-based methods for HEp-2 cell classification.

name	preprocessing	Feature	classifier
Ghosh[⋆]	Grayscale	Area, Eccentricity, GLCM	SVM, linear kernel
Gilbert[†]	Green channel	Pixel intensity, number and size of blobs	SVM, RBF kernel
Hassaine[†]	RGB, HLS and lab	Morphological features	Logistic regression
Kazanov[‡]	-	Morphological features and pixel intensities	Nave Bayes
Kovacs[†]	-	Pixel intensity, area	SVM, k-NN and Nave Bayes
Maree[•]	Normalized RGB	Pixel intensity	SVM
Rezvani[†]	Grayscale	Eccentricity, Compactness, Roundness, etc.	RBF Kernels
Shen[†]	-	Histogram of RGB channels	SVM, RBF kernel
Snell[∓]	Grayscale	Shape, Different statistics	SVM, RBF kernel
Strandmark[°]	Projection to Principal components	Geometrical, Pixel intensity, GLCM	Random Forest
Wang[†]	Intensity normalization	Pixel intensity	SVM, linear kernel
Yang[◇]	Projection to Principal components	Pixel intensity	SVM, linear kernel

[⋆] [Ghosh and Chaudhary, 2012] [†] [Foggia et al., 2013] [‡] [Ponomarev et al., 2014] [•] [Marée et al., 2013]
[∓] [Snell et al., 2012] [°] [Strandmark et al., 2012] [◇] [Yang et al., 2014b]

various features does not necessarily result in representative and discriminative ones. Second, the possibility of obtaining redundant features is very high and it increases the curse-of-dimensionality problem [Friedman, 1997]. This in turn calls

Table 2.2: Methods for HEp-2 cell classification using engineered features.

name	preprocessing	Feature	classifier
Kastaniotis[⋆]	Grayscale	SIFT, Rotation invariant LBP	SVM, RBF kernel
Kuan[†]	-	LBP, DCT, Wavelet	SVM, linear kernel
Nosaka[‡]	Green Channel	CoALBP	SVM, linear kernel
Stoklasa[•]	Green Channel	LBP, Color Structure, SIFT	k-NN
Wafa[∓]	-	Pyramid of DoG	k-NN
Xiangfei[°]	-	Frequency Histogram of Textons	k-NN
Faraki[◇]	-	Covariance Descriptor from Gabor filter response	SVM, linear kernel

[⋆] [Theodorakopoulos et al., 2012] [†] [Li et al., 2012] [‡] [Nosaka and Fukui, 2014] [•] [Stoklasa et al., 2014]
[∓] [Bel Haj Ali et al., 2012] [°] [Foggia et al., 2013] [◇] [Faraki et al., 2014]

for a post processing stage (e.g. PCA) to reduce the feature dimension and also make it more discriminative. Lastly, there are two types of HEp-2 cell images in terms of intensity levels (see Section 3.3). When dealing with *intermediate* level images, where the pixel values are much lower than *positive* intensity images, the intensity-based methods are prone to misclassification and need a preprocessing stage to obtain representative features.

2.2.2 Feature Engineering

Engineered features are widely used in pattern recognition problems because they are engineered for application independent processes. Table. 2.2 shows several methods that have exploited such features.

[Di Cataldo et al., 2014] used DCT features in several binary images, which are obtained by thresholding the input image with different values. The LBP [Ojala et al., 1996] and its modifications (Co-Occurrence among LBP and Rotation Invariant LBP) are used in many works in the literature of the HEp-2 cell image classification. [Nosaka and Fukui, 2014], the winner of the ICPR2012 contest, introduced an extension of LBP for feature selection, into this problem. They developed Co-Occurrence among LBPs to consider the spatial relation among the LBPs. They claimed that their method is robust to rotation of the input cell image.

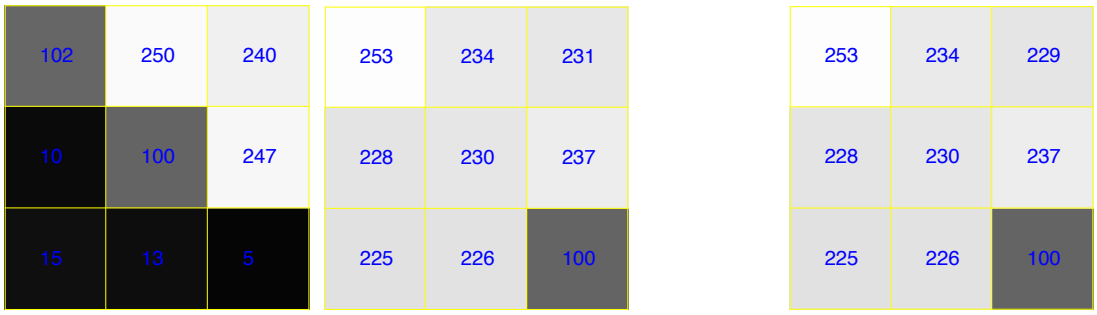
The LBP operator for pixel position (x_i, y_i) with intensity value of I_i is formulated as:

$$LBP(x_i, y_i) = \sum_{j=0}^{7} T\left(I_i^j - I_i\right) \times 2^j \tag{2.1}$$

$$T(k) = \begin{cases} 1, & if \;\; k \geqslant 0 \\ 0, & if \;\; k < 0 \end{cases}$$

where $I_i^j, j \in \{1, 2, \ldots, 7\}$ are intensity values of eight neighboring pixels around the pixel (x_i, y_i).

However, the main problem of LBP-based approaches is the possibility of coding two very different local patterns into the same LBP value, and two similar local patterns into very different LBP values. These problems are because of assigning binary values according to the difference of the center pixel value with the ones of the surrounding pixels. For instance, it is evident in Fig. 2.1a, although two patterns are totally different, the LBP values of their center pixels are the



(a) LBP = 225 (b) LBP = 97

Figure 2.1: Shortcoming of LBP in assigning (a) similar binary values for two different patterns and (b) different binary values for almost same patterns.

same (LBP = 225) for both patterns. On the other hand, the visual patterns of right image in Fig. 2.1a and Fig. 2.1b are almost the same but their LBP values for the center pixels are significantly different (225 versus 97). Therefore, applying LBP operator on the images to represent the local structures could result in non-discriminative features which degrades the final classification performance.

The performance of these methods dominates the intensity-based approaches because the engineered features are specifically designed and tuned for the problem at hand. However, there is no intervening procedure between the feature extraction stage and the classifier to make the features more representative and discriminative.

2.2.3 Feature Learning and Sparse Coding

There are different methods for feature learning which are categorized into supervised and unsupervised. Neural Networks and their modifications, Convolutional Neural Network (CNN) and deep learning are among supervised methods in which features are learned with labeled input data [Guo et al., 2016]. On the other hand,

Independent Component Analysis (ICA), Auto-encoders and dictionary learning are examples of the unsupervised approaches [Khorsandi et al., 2015a]. The input data for these methods are the raw data to induce the algorithm to learn features from it. However, the input of such methods could be the engineered features instead of raw values to help the algorithm learn new features on top of the engineered ones and map the feature space to a higher/lower dimension.

Recently, there has been an increasing interest in sparse coding and dictionary learning in computer vision and image processing research for classification tasks [Mairal et al., 2012, Jiang et al., 2013b], which we will discuss more details in Section 2.2.4. [Kong et al., 2014] used the HoG and responses of filter banks as input features and learned a dictionary. The corresponding sparse codes are then used to minimize the class reconstruction error. Additionally, [Manivannan et al., 2014a] extracted variety of features such as multi-resolution Local Pattern (mLP), SIFT, Random Projection (RP) and Intensity Histogram (IH). Then, the Locality-constrained Linear Coding (LLC) is exploited to calculate the sparse codes.

Table. 2.3 shows different methods that are exploited the feature learning procedures. The CNN and Deep learning methods required very large amount of input data to perform reasonably and learn the features. This is problematic when dealing with the HEp-2 cell datasets as they have limited number of images. Additionally, the training procedure of CNNs is time consuming and complex. Specially, by increasing the hidden layers and nodes in the network, the number of parameters which should be learnt increases exponentially. To solve the problem of not having enough input data, the augmentation and synthesizing schemes are suggested [Gao et al., 2014, Thibault and Angulo, 2012], but the final per-

Table 2.3: The methods for HEp-2 cell classification problem by using feature learning methods.

name	**preprocessing**	**Feature**	**classifier**
Wang†	Intensity normalization	Pixel intensity + BoW	SVM, linear kernel
Siyamalan⋆	Intensity normalization	LBP, SIFT, Intensity Histogram + Sparse Coding	SVM, linear kernel
Larsen‡	Green Channel	Shape-index histogram + BoW	SVM, linear kernel
Malon†	Contrast stretching	Pixel Intensity + CNN features	CNN
Thibault∓	-	Pattern spectrum	Neural Network
Wiliem°	-	SIFT, DCT + Sparse Coding	k-NN
Gao◇	Contrast stretching, image resized, image augmentation	CNN features	CNN

† [Foggia et al., 2013] ⋆ [Manivannan et al., 2014a] ‡ [Larsen et al., 2014] ∓ [Thibault and Angulo, 2012]
° [Wiliem et al., 2013] ◇ [Gao et al., 2014]

formance is not significantly improved because the synthesized images are only the regeneration of the available data that can be ignored by the network. The CNNs and deep learning algorithms require a large number of real data rather than synthesized information.

2.2.4 Compressed Sensing, Sparse Coding and Dictionary Learning

One of the inseparable tasks of image/signal processing is acquiring and sampling of signals in order to save, transmit and reconstruct them. The Shannon-Nyquist sampling theorem stated that a signal can perfectly reconstructed if the sampling rate is more than twice of the highest frequency of the signal. Compressed Sensing (CS) theory is then proposed which stated that the signals can be reconstructed by fewer samples than the sampling theorem in terms of having a prior knowledge about the signal's sparsity. a signal ($\mathbf{s} \in \mathbb{R}^n$) is considered sparse, if it has small amount of nonzero entries.

The CS method first takes the samples and then compresses the signal. The sampling stage is consist of acquiring m weighted linear combination of samples where $m \ll n$. The m-measurements creates the $m \times n$ measurement matrix ($\mathbf{\Psi}$),

$$\mathbf{y} = \mathbf{\Psi}\mathbf{s} \tag{2.2}$$

and if the noise of the system is considered, the (2.2) can be written as $\mathbf{y} = \mathbf{\Psi}\mathbf{s}+\eta$ where η is the noise. Due to the lower number of equations than the signal dimension, the CS faces an under-determined system of linear equations which has either no solution or infinitely many solutions generally. However, the sparsity constraints of the CS employs nonlinear optimization-based methods to search for the sparsest signal.

By assumption of the CS, the input signal $\mathbf{s}$ has the sparsity property. Therefore, it can be represented as a linear combination of dictionary bases $\mathbb{D} =$

$[D_1, D_2, \ldots, D_K]^\intercal$.

$$\mathbf{s} = \mathbb{D}\mathbf{z} \tag{2.3}$$
$$||\mathbf{z}||_0 \ll n$$

where n is the dimension of the input signal $\mathbf{s}$ and $\mathbf{z}$ is the coefficients of the dictionary bases, which is sparse.

By combining the (2.2) and (2.3) the formulation is as follows

$$\mathbf{y} = \mathbf{\Psi}\mathbf{s} = \mathbf{\Psi}\mathbb{D}\mathbf{z} = \Theta\mathbf{z} \tag{2.4}$$

According to the CS theory, a signal ($\mathbf{s}$) can be reconstructed by its coefficients ($\mathbf{z}$) when Θ satisfies the restricted isometry property (RIP) [Candès et al., 2006]. A matrix Θ is said to satisfy the RIP of order k with constants $\delta_k \in (0, 1)$ if $(1 - \delta_k)||\mathbf{v}||_2^2 \leq ||\Theta\mathbf{v}||_2^2 \leq (1 + \delta_k||\mathbf{v}||_2^2)$ for any $\mathbf{v}$ such that $||\mathbf{v}||_0 \leq k$.

Normally, in the literature of CS, the ℓ_0-norm, which is simply the number of non-zero elements of a signal, is used as sparsity measurement;

$$\underset{\hat{\mathbf{s}} \in \mathbb{R}^n}{\operatorname{argmin}} ||\hat{\mathbf{s}}||_0 \tag{2.5}$$
$$\text{s.t. } \mathbf{\Psi}\hat{\mathbf{s}} = \mathbf{y}$$

However, the above equation is NP-hard and alternative ℓ_1-norm is suggested to relax the problem.

Dictionary Learning - As can be derived from (2.4), an input signal can be reconstructed by a linear combination of a few dictionary columns (words) as the

weights are forced to be sparse.

One of the most popular unsupervised methods for dictionary learning is Vector Quantization (VQ) by using k-means [Philbin et al., 2008]. Let $\mathbb{F}$ be a set of features in a $\mathcal{D}$-dimensional space, i.e. $\mathbb{F} = [F_1, F_2, \ldots, F_N] \in \mathbb{R}^{(\mathcal{D}\times N)}$, and $\mathbb{D} = [D_1, D_2, \ldots, D_K] \in \mathbb{R}^{\mathcal{D}\times K}$ are K words (cluster centers) of the dictionary. In VQ, the objective is to learn $\mathbb{D}$ by solving the following optimization problem:

$$\min_{\mathbb{D}} \sum_{i=1}^{N} \min_{k=1\ldots K} \left\| F_n - \mathbb{D}_k \right\|^2 \tag{2.6}$$

where $\|.\|$ denotes the ℓ_2-norm. In this formulation, all the feature points in the feature space will become a member of only one of the K cluster centers. By introducing an indicator function Z, which contains the weights of the cluster centers $Z = [z_1, z_2, \ldots, z_N] \in \mathbb{R}^{K\times N}$, the equation (2.6) can be reformulated as:

$$\min_{Z,\mathbb{D}} \sum_{i=1}^{N} \left\| F_n - \mathbb{D} z_n \right\|^2 \tag{2.7}$$

$$\text{s.t. } Card(z_n) = 1, \quad |z_n| = 1, \quad z_n \geq 0, \forall n$$

The cardinality constraint on z_n ($Card(z_n)$) means that only one element of z_n can be nonzero. Moreover, this value should be nonnegative and the ℓ_1-norm (summation of all elements) of Z should be equal to one. Because of the hard constraints on cardinality and ℓ_1-norm of z_n, this dictionary learning is computationally complex which affects the reconstruction error. To relax these hard constraints a sparse coding method is proposed [Yang et al., 2009]

$$\min_{Z,\mathbb{D}} \sum_{i=1}^{N} \left\| F_n - \mathbb{D} z_n \right\|^2 + \lambda |z_n| \tag{2.8}$$
$$\text{s.t. } ||D_k|| \leq 1, \quad \forall k = 1, 2, \ldots, K$$

Here the ℓ_1-norm of weights is relocated to the objective function using Lagrange Multiplier and just an ℓ_2-norm of the dictionary is used as a constraint. This constraint is to prevent all the elements of dictionary from becoming zero, which is a trivial minimum of the objective function.

There are two simple iterative algorithms to calculate the dictionary; K-SVD and Method of Optimal Directions (MOD). Because the problem in (2.8) is non-smooth convex, it should be solved iteratively till a sparsest solution is obtained [Zonoobi et al., 2011]. Firstly the values of the dictionary is initialized by using k-means method which provide the cluster centers of the input features as the dictionary words. Then the main procedure consists of the following two stages:

- Sparse Coding: In this step, the dictionary ($\mathbb{D}$) is fixed and the (2.8) is reformulated as:

$$\min_{Z} \sum_{i=1}^{N} \left\| F_n - \mathbb{D} z_n \right\|^2 + \lambda |z_n| \tag{2.9}$$

  This is a linear regression problem with ℓ_1-norm regularization on the coefficients which is known as LASSO (least absolute shrinkage and selection operator) method [Tibshirani, 1996] in the literature. Other greedy methods in-

cluding Orthogonal Matching Pursuit (OMP) [Mallat and Zhang, 1993, Pati et al., 1993] is often employed due to its efficiency [Tropp, 2004].

- Dictionary Update: In this step, the coefficients are assumed fixed and the dictionary updated by following formulation:

$$\min_{\mathbb{D}} \sum_{i=1}^{N} \left\| F_n - \mathbb{D} z_n \right\|^2 \tag{2.10}$$
$$\text{s.t. } ||D_k|| \leq 1, \quad \forall k = 1, 2, \ldots, K$$

  Two algorithms of K-SVD and MOD differ in this stage where in K-SVD, the dictionary is updated atom-by-atom where MOD updates the whole dictionary. Additionally, the MOD is computationally expensive rather than K-SVD due to the matrix inversion operation in its algorithm [Patel and Chellappa, 2011].

In this thesis, we proposed other dictionary learning methods by using Non-Parametric Bayesian Method (see Section 4.2.1), Adaptive Distributed Dictionary Learning (see Section 5.2) and Joint Multi-Cue Dictionary Learning (see Section 7.3).

Chapter 3

Sparse Coding and Feature Combination

3.1 Introduction

The idea of Sparse BoW (SBoW) for Hep-2 cell classification is investigated. This chapter presents the latest development and improvements in several aspects. First, a *hierarchical pooling* approach is proposed to perform the max-pooling operator on the different image patch sizes according to the level of the regions in the SPM. Second, different parameters are investigated and the optimal ones are identified which lead to much higher cell classification accuracy. For example, our study shows that the iteration of the dictionary learning is closely correlated with the cell classification accuracy: a certain number of iterations give the best accuracy which cannot be either too large or too small. Third, this work studies different features including SIFT, SURF and the complementary combination of

both features. Fourth, the work performs comprehensive evaluations on two public benchmarking datasets as well as detailed analysis and discussion of different selections of parameters and strategies. For example, different pooling strategies are investigated and the best-performing max-pooling strategy is identified and analyzed.

In [Ensafi et al., 2014a, Ensafi et al., 2014b] a dictionary learning method is applied and the sparse codes of image patches are aggregated with Spatial Pyramid Matching (SPM) [Lazebnik et al., 2006]. In particular, an image is first partitioned in 1, 4 and 16 regions in three pyramid layers and the max-pooling operator is then applied on the sparse codes of each region to form the final feature vector by concatenation (see Fig. 3.3). One limitation of this approach is the mismatch between the resolution of the regions and the size of the image patches. In other words, certain local information will be lost when the large size image patches are aggregated in a small image region. In this chapter, we use the pyramid of the image patches with various patch sizes to alleviate this problem.

The rest of this chapter is organized as follows. Section 3.2 describes the proposed HEp-2 cell classification technique. The Publicly available datasets which are used for the experiments are described in Section 3.3. Experimental results are then presented in Section 3.4. Several concluding remarks are finally drawn in Section 3.4.3.

3.2 Method

Figure 3.1 shows the framework of the proposed system [Ensafi et al., 2014b] which is named Sparse Bag of Words (SBoW). In the preprocessing stage, if the bounding box of the cells are not given, the cell extraction method (see Section 3.2.2) is applied to get the masks and the bounding box of the cells in each specimen image. Then, for each cell, multiple overlapping equal-distanced image patches are determined. For each image patch, grid SIFT and SURF features are then extracted and concatenated to produce the patch-level features. During the training stage, the extracted patch-level features are sampled to learn a dictionary of visual words under the BoW framework, where the sparse coding scheme is adopted to learn each visual dictionary word. Multi-scale and max-pooling strategies are implemented to transform the visual features into feature vectors, which are further fed to train a multi-class SVM classifier. Then for each test image, the SIFT and SURF features are similarly extracted and transformed to feature vectors leveraging on the learned dictionary. The type of the HEp-2 cells can be identified by using the pre-trained multi-class SVM classifier.

3.2.1 Preprocessing

To perform cell classification, the cells are first extracted from the specimen image using the provided masks of the images. However, there are two problems with these masks. First, the cell masks for images are not accurate and contain some non-cell areas depending on the segmentation algorithm used. Second, those cells that 'touch' each other in the provided masks cannot be extracted based on the

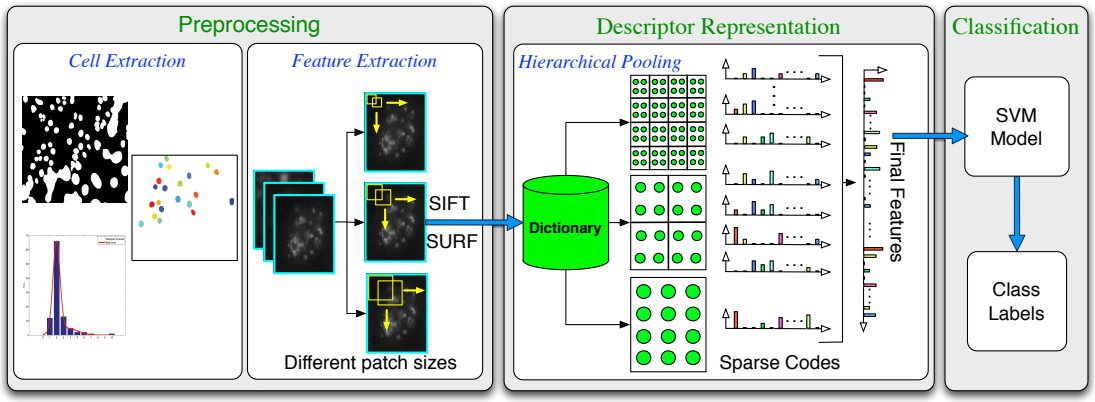


Figure 3.1: The SBoW framework for HEp-2 cell classification CAD system.

provided masks. The two problems are illustrated in two sample images as shown in Fig. 3.2.

The first problem originates from the mask preparation process. In particular, a special dye is used to stain the cytoplasm of the cells and a global thresholding technique is then applied to produce the cell masks [Foggia et al., 2014]. However, this method fails in some cases and creates noisy masks, where some non-cell elements in the tissue are wrongly extracted as masks. These non-cell mask elements are often very large (see Fig. 3.2b), and can be distinguished and discarded according to their sizes. For the touched and overlapped cells problem, the majority of cells are selected which are well segmented for the cell classification.

3.2.2 Cell Extraction

In the cell extraction stage, the goal of finding cells with similar shapes and sizes is achieved using morphological operations which result in connected pixels extracted from the image masks that are analyzed to select those that correspond

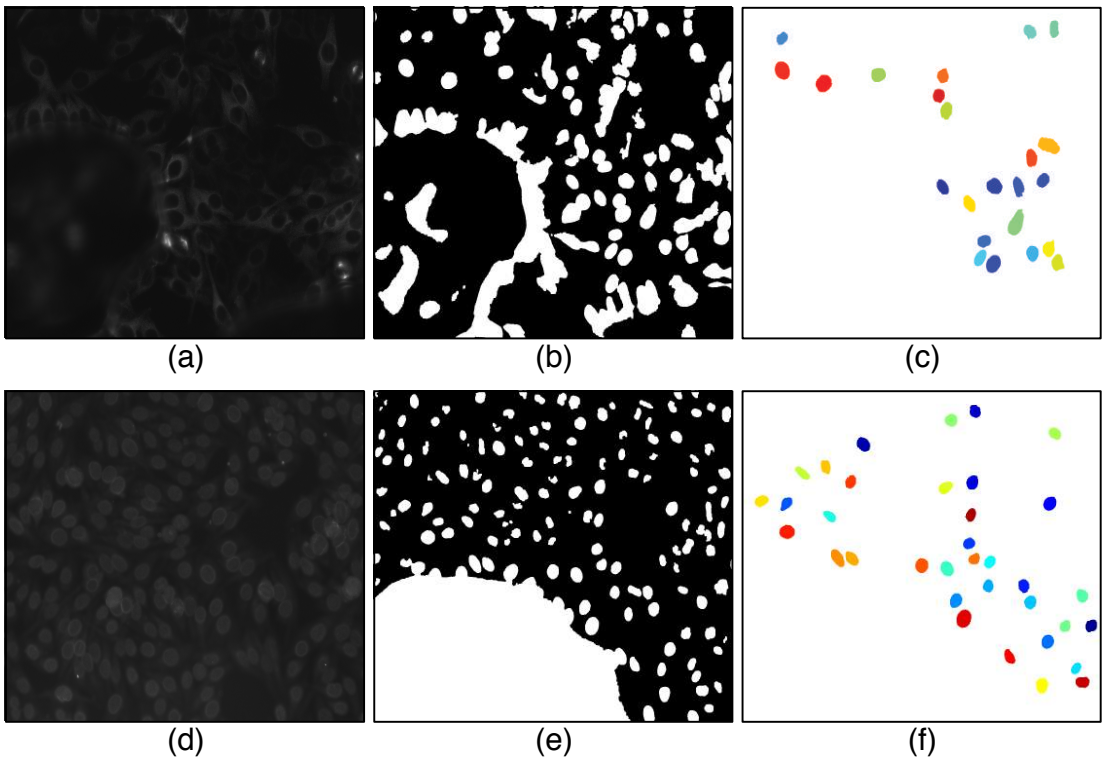


Figure 3.2: Inaccurate masks of the cells in the ICIP2013 dataset in terms of noisy segmentation and overlapping cells. (a) and (d) are the specimen images. (b) and (e) are provided masks. (c) and (f) are the extracted cells by the proposed method.

to real cells for cell classification as shown in the 'Cell Extraction' stage of Fig. 3.1. There is a need to distinguish those large connected pixels which could have been wrongly segmented or represent overlapped cells.

The *area* and *solidity* morphological features are extracted from each connected pixel for cell classification. The histogram of the *area* features are quantized into bins and the maximum bin (b_{mx}) is taken to represent the *area* for most connected pixels that are likely to be proper cells with no overlaps with other cells. We then proceed to select those cells with *area* that are close to b_{mx}, based on the standard deviation of the histogram (std) resulting in range ($R_{mx}; R_{mn}$

respectively) as follows:

$$R_{mn} = \max(0, \quad b_{mx} - std) \tag{3.1}$$
$$R_{mx} = \min(b_{mx}, \; b_{mx} + std)$$

The *std* is calculated by fitting a Gaussian probability distribution function on the histogram. Only those connected pixels with areas more than a threshold (A_t) are considered as correctly identified cells. Next, cells with similar shape are identified using the *solidity*, S, property which is related to the roundness of the cells and is defined as follows:

$$S = \frac{A}{CA} \tag{3.2}$$

where A is the area of the connected pixel and CA (Convex Area) is the number of pixels in the convex hull of the area. The *solidity* S, would be close to one if the cells are of circular shape.

As shown in Algorithm 1, in each iteration for one specimen image, the number of bins (nb) and the values of S_{mn} (minimum solidity) are decreased gradually to select at least 5 cells in each image. By decreasing these values, we gradually relax the constraints for selecting the cells, because in some mask images, the connected pixels have irregular shapes rather than circular shapes. The initial values are selected with a cross validation strategy: $nb = 12$; $S_{mn} = 0.98$; $A_t = 45$. This helps choose those cells that of average size (i.e., area) and circular in shape.

Algorithm 1: Cell Extraction Algorithm

```
   Input: Specimen Images and Masks
   Output: Extracted Cells
1  begin
2      nb = 12, S_mn = 0.98, A_t = 45;
3      foreach mask image do
4          Get the connected pixels;
5          Get the morphological properties (Area, Solidity);
6          Fit a Gaussian probability distribution on the histogram of the
           areas with nb bins;
7          Calculate the standard deviation (std) of fitted Gaussian
           distribution;
8          R_mn ← max(0,  b_mx − std);
9          R_mx ← min(b_mx, b_mx + std);
10         # selected cells ← 0;
11         foreach area(cell) ∈ [R_mn, R_mx] do
12             if solidity > S_mn then
13                 if area > A_t then
14                     if area ∩ boundary = {∅} then
15                         Get the cell;
16         if # selected cells < 5 then
17             S_mn ← 0.95 × S_mn;
18             nb ← nb − 1;
19             Go to 6;
```

3.2.3 Feature Extraction

Using intensity values of the images directly as features has some problems. First, in each dataset, there are two intensity levels namely, *positive* and *intermediate* levels, which differ significantly. Particularly, the intensity values of *positive* images are greater than that of *intermediate* images, where *positive* cells can be easily seen by naked eyes but not for the *intermediate* cells. Second, the ICIP2013 dataset (sea Section. 3.3.2) has gray-scale values but the ICPR2012 dataset has

color images. Finally, the noise in the images makes the intensity level analysis inaccurate.

To this end, the SIFT [Lowe, 2004] and SURF [Bay et al., 2008] features are utilized to capture the appearance characteristics of the different types of HEp-2 cells. In particular, SIFT features are computed by down sampling of the image in different smoothed image levels and SURF features are computed using the Hessian matrix. As a result, SURF provides better features in the presence of illumination changes (in *positive* and *intermediate* intensity levels) whereas SIFT performs better in the presence of image rotation and blur [Juan and Gwun, 2009]. The two types of features therefore complement each other and the combination of them produces features with better representation and discrimination capability.

Grid SIFT and SURF features are used as illustrated in Fig. 3.1. The standard SIFT feature are not used because it first runs the corner detection [Harris and Stephens, 1988,Shi and Tomasi, 1994] to capture the interest points and then extracts the features of these points [Ensafi et al., 2014a]. But this approach would not perform well for the cell classification problem because the HEp-2 cell patterns within the immunofluorescence images are usually of a very small size. In particular, the number of interest points for the Homogeneous cell with homogeneous visual pattern is much lower than that of Centromere class that contains many shinny points as illustrated in Figs. 3.4 and 3.5. We therefore utilize grids over the whole cell region to capture the visual features.

To produce these features, the entire cell image is divided to overlapping patches. In each patch, the SIFT and SURF features are captured and combined together. In this regard, 128 SIFT and 64 SURF features are extracted which

creates the 192 features for each patch in total.

3.2.4 Descriptor Representation

To describe the input images, we adopted the idea of BoW which was originally applied for the representation of text documents but recently used widely in image classification and retrieval [Zhang et al., 2010]. The BoW learns a variety of visual words that are literally the basements of the input images. Therefore, the images can be reconstructed by assigning weights to the basements.

$$\mathbf{F} = \mathbf{D}\mathbf{Z} \tag{3.3}$$

where $\mathbf{F}$ is a set of features in a $\mathcal{D}$-dimensional space extracted from images, $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N] \in \mathbb{R}^{(\mathcal{D}\times N)}$, and $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K] \in \mathbb{R}^{(\mathcal{D}\times K)}$ is K words of the dictionary. Additionally, $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N] \in \mathbb{R}^{(K\times N)}$ is the coefficients of the features which indicates the specific words that are used to reconstruct the input image.

Equation 3.3 does not have enough information to calculate a unique solution and it is not a well posed problem [Tikhonov, 1963]. Moreover, some noise parameter is added to the formulation in reality because the information gathering contains noise and the reconstruction of the images using the learned dictionary is not perfect.

On the other hand, a sparse representation of codes are required to have as

sparse as possible words to reconstruct the input image.

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} ||\mathbf{z}||_0 \tag{3.4}$$
$$\text{s.t.} \quad \mathbf{f} = \mathbf{D}\mathbf{z} + \mathbf{e}$$

where $\mathbf{e}$ is assumed to be zero mean Gaussian noise with σ^2 variance. Additionally, the solution should be regularized, where the prior knowledge is needed about the solution. Therefore, the equation can be perceived by using a probabilistic approach. To this end, the Laplacian distribution can be chosen for the prior information on codes.

$$p(\mathbf{z}) = \prod_{i=1}^{K} \frac{\lambda}{2} \exp\left(-\lambda \left|\mathbf{z}_i\right|\right) \tag{3.5}$$

here the assumption is that the $\mathbf{z}$ elements are i.i.d and λ is the regularization parameter. By calculating the maximum a posteriori (MAP) estimate of $\mathbf{z}$ by using the following Bayes' theorem, the sparse representations of input images can be calculated: [Amiri and Haykin, 2014]

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} p(\mathbf{z}|\mathbf{F}, \mathbf{D}) = \arg\max_{\mathbf{z}} \frac{p(\mathbf{F}|\mathbf{z}, \mathbf{D})p(\mathbf{z})}{p(|\mathbf{D})} \tag{3.6}$$

The negative logarithm of posterior can also be minimized instead of MAP,

to come up to the sparse coding formulation.

$$\min_{\mathbf{z},\mathbf{D}} \sum_{n=1}^{N} \left\|\mathbf{F}_n - \mathbf{D}\mathbf{z}_n\right\|^2 + \lambda|\mathbf{z}_n| \tag{3.7}$$
$$\text{s.t. } ||\mathbf{d}_k|| \leq 1, \quad \forall k = 1, 2, \ldots, K$$

here the dictionary is also added to the parameters of the sparse coding formulation to capture a good dictionary which can minimize the sparse codes as well.

Because the problem in Eq. 3.7 is non-smooth convex, it should be solved iteratively till a sparsest solution is obtained [Zonoobi et al., 2011]. Firstly the values of the dictionary is initialized by using k-means method which provide the cluster centers of the input features as the dictionary words. By fixing these values, the weights are learned by using conjugate gradient method. The weights are then fixed and the dictionary words are optimized [Ensafi et al., 2014b].

A dictionary with dimensionality K is said to be *critically complete* if K is close to the dimensionality of feature vectors ($\mathcal{D}$). In practice, K should be much larger than $\mathcal{D}$ to provide an *over complete* dictionary to insure a proper representation of the input features. This fact is biologically inspired from human visual cortex that is estimated to be over complete by a factor of 500. For example, a 14×14 input patch is coded by 100 000 neurons. For the HEp-2 cell classification, an optimal K of 1024 is selected based on extensive tests on a large amount of cell images (Note that the dimension of SIFT and SURF features is 192) [Ensafi et al., 2014b].

To this end, the features (sparse codes) of each patch of the input images are extracted. The next issue is how to combine them to represent the features

of whole image. Simply, concatenating the features of each patch together is not feasible because of the different sizes of the images. A pooling strategy is therefore needed to create feature vectors of the same size for each image.

Another issue is that by feature extraction of each patch separately, the spatial information of the images may be ignored because the features are extracted locally in each patch. Thus the scaled sparse codes is used, where the coded space of an image is divided to equal regions in different levels. Then the histogram of codes for each region is calculated with a fixed number of bins as shown in Fig. 3.3. In this regard, the same size feature vector of the regions are calculated although the input features and size of the regions are different. Finally, according to the chosen pooling strategy, the final feature vector is calculated, which is a concatenation of feature vectors of all regions.

However, the difference between the sizes of the image patches and regions in the SPM layers, will introduce the loss of the local information of the patches. As all the image patches have the same size and would not change according to the shrinking size of the regions in the SPM method, a hierarchical strategy is proposed to use the information of the different size image patches for different regions. To this end, three sizes of image patches are extracted to be used in three levels of the SPM method. In particular, the patch size of 18×18 is used in the first layer where all the image patches are pooled. In the second layer, the image is divided to 4 regions where the information of the 12×12 image patches is used. Finally, the 8×8 image patches are used for the third layer which has 16 regions. These image patches of three different sizes are employed by extensive examination through cross-validation.

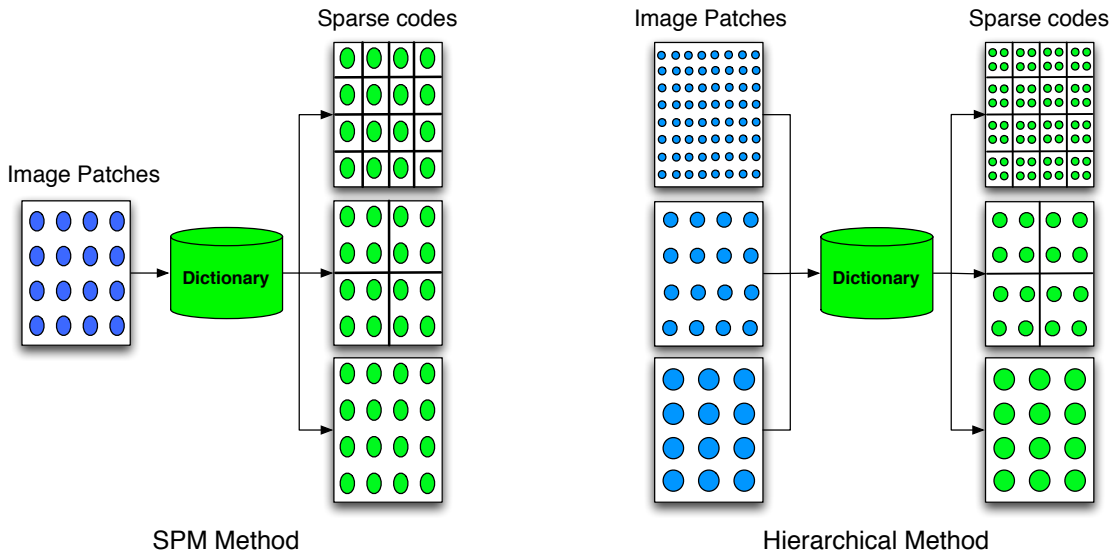


Figure 3.3: The pooling strategy in SPM (Left) and Hierarchical (Right) for capturing the final sparse codes.

There are different pooling strategies to obtain the final features of images according to their sparse codes. These strategies, which are examined in Section. 3.4.3, include *one-hot encoding*, *average pooling* and *max-pooling*. In one-hot encoding, just one representative feature is selected for each region in the scaled levels. But in average and max pooling, the average and maximum values for each bin of histogram is selected for each region and all the final histograms are concatenated.

3.2.5 Classification Process

A multi-class linear SVM is trained to classify the feature vector of input images. To this end, a strategy of one-versus-all is adopted. In particular, L linear binary classifiers are learned as specified in (3.8) and concatenated to form a multi-class classifier. In this regard, the class label of an input image (the input image is

converted to sparse feature vectors) is $y_i^c = \{+1, -1\}$ where $y_i^c = +1$ means that image i belongs to class c, $c \in \mathcal{Y} = 1, \ldots, L$. Therefore, we have $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ input pairs for training sets [Yang et al., 2009, Ensafi et al., 2014a].

$$\min_{\mathbf{w}_c} \left\{ J(\mathbf{w}_c) = \|\mathbf{w}_c\|^2 + C \sum_{i=1}^{n} \ell(\mathbf{w}_c; y_i^c; \mathbf{x}_i) \right\} \tag{3.8}$$

The optimization problem in (3.8) uses the hinge loss as the cost function, which can be solved using the conjugate gradient method. As the hinge loss function is not differentiable in all the points, the quadratic hinge loss [Yang et al., 2009] is used

$$\ell(\mathbf{w}_c; y_i^c; \mathbf{x}_i) = [\max(0, \mathbf{w}_c^\intercal \mathbf{x}.y_i^c - 1)]^2 \tag{3.9}$$

By learning the multi-class SVM model, the class of the sparse feature vector is finally assigned as follows:

$$y = \max_{c \in \mathcal{Y}} < \mathbf{w}_c, \mathbf{x} > \tag{3.10}$$

Here, the maximum distance of each feature point with respect to L classification lines is calculated in the feature space and the corresponding class is assigned to the testing point [Ensafi et al., 2014b].

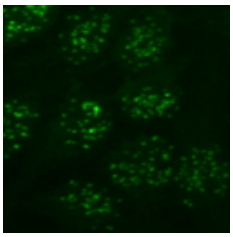
3.3 Datasets

The two publicly available datasets are ICPR2012 [Foggia et al., 2013] and ICIP2013 datasets [Foggia et al., 2014] which were used in two contests held with ICPR2012

and ICIP2013 conferences. The ICPR2012 dataset has a training set and a test set both available to the public, whereas ICIP2013 dataset has only a training set public available. Both datasets have six classes in *Cell Level*. But at *specimen level*, there are six classes for the ICPR2012 dataset and seven classes for the ICIP2013 dataset, respectively. Cell masks are provided in order to classify the cells without considering other neighboring cells for the specimen image.

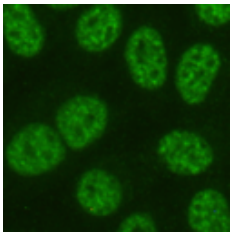
3.3.1 ICPR2012

This dataset is introduced as a contest in conjunction with International Conference on Pattern Recognition (ICPR) 2012 and is publicly available[1]. ICPR2012 dataset consists of 28 HEp-2 cell images where each image has a resolution of 1388×1038 pixels. The images are captured by using a fluorescence microscope (40-fold magnification) that is coupled with a 50W mercury vapor lamp and a digital camera. Each of the 28 RGB images contains one of the six staining patterns, which have following characteristics and figures:

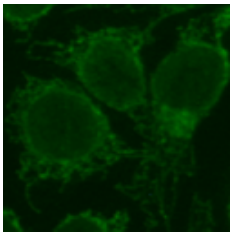


Centromere (Ce): Contains several discrete small spots that are scattered throughout the nuclei area. These speckled can be observed in the nuclear chromatin.

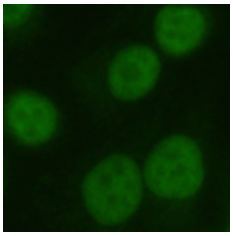
[1] http://nerone.diem.unisa.it/hep2contest/index.shtml



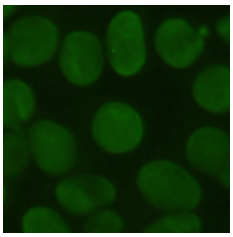
Coarse Speckled (Cs): Contains coarse granular nuclear staining pattern. Some holes are observed on the nuclei area and the pattern is so close to Fine Speckled cell images.



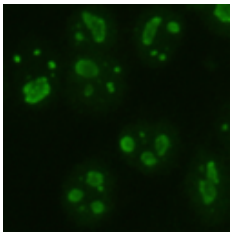
Cytoplasmatic (Cy): contains a very fine dense resembling homogeneous staining which covering part or the cytoplasm.



Fine Speckled (Fs): Contains fine granular nuclear staining pattern. In most of the cell images of this type, the patterns are observed blurry.



Homogeneous (H): The staining pattern is spread out over the interphase nuclei. These type of cell images have smooth diffused characteristics all over the nuclei area.



Nucleolar (N): Small compact particles can be observed in the cells' nucleoli. Without these granules, this pattern is close to Homogeneous class. At most six granules can be observed.

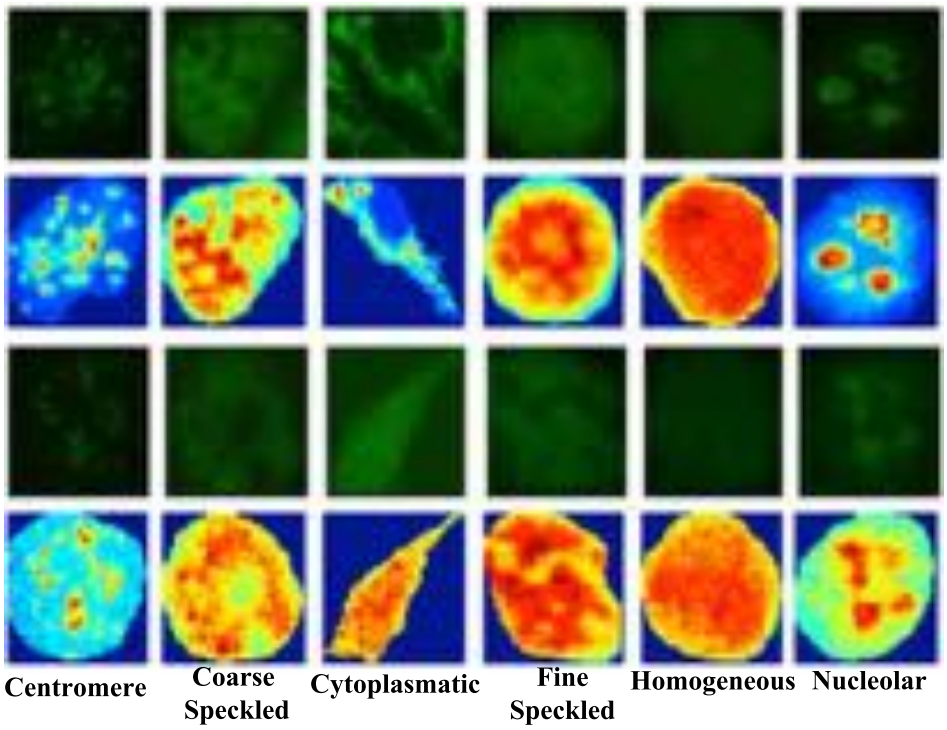


Figure 3.4: The *cell level* images of ICPR2012 for *positive* (top two rows) and *intermediate* (bottom two rows) intensity levels and their heat-maps to show the underlying pattern.

For more information regarding the related AD to each staining patterns and high resolution images, please refer to the Appendix.

The cell mask and cell label are provided for each image. In addition, there are two levels of intensity images, namely, *intermediate* images and *positive* images as illustrated in Fig. 3.4. The heat map of the cells are also shown for better understanding of underlying patterns. In particular, the intensity values of the *positive* cells are more than *intermediate* cells, which can be easily investigated by naked eye. The intensity values of foreground in *intermediate* cells are relatively close to the background pixels, which makes the classification problem more challenging for this type of cells.

Table 3.2: The statistics of cell and specimen images in the ICPR2012 training and testing datasets. The values in parentheses are the number of cells in specimen images.

ICPR2012	Training set		Test set		Total		**Overall**
Dataset	Intermediate	Positive	Intermediate	Positive	Intermediate	Positive	
Centromere	2 (119)	1 (89)	1 (65)	2 (84)	3 (184)	3 (173)	6 (357)
Coarse speckled	1 (41)	1 (68)	1 (33)	2 (68)	2 (74)	3 (136)	5 (210)
Cytoplasmatic	1 (24)	1 (34)	1 (13)	1 (38)	2 (37)	2 (72)	4 (109)
Fine speckled	1 (48)	1 (46)	1 (63)	1 (51)	2 (111)	2 (97)	4 (208)
Homogeneous	1 (47)	2 (103)	1 (61)	1 (119)	2 (108)	3 (222)	5 (330)
Nucleolar	1 (46)	1 (56)	1 (66)	1 (73)	2 (112)	2 (129)	4 (241)
Total	7 (325)	7 (396)	6 (301)	8 (433)	13 (626)	15 (829)	28 (1455)
	14 (721)		**14 (734)**		**28 (1455)**		

Table. 3.2 shows the number of cells in each specimen image for different patterns and intensities. In total, there are 1455 cells in the 28 images, including 14 specimen images (containing 721 cells) for training and 14 (containing 734 cells) for testing.

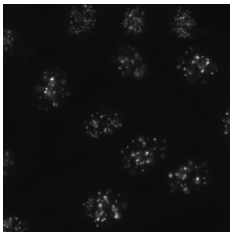
As can be seen in Table. 3.2 the distribution of cells for different patterns are not the same. For instance, the number of Cytoplasmatic cells are almost one-third of the number of Homogeneous cells. These kind of information can be used as a prior knowledge in the experiment.

3.3.2 ICIP2013

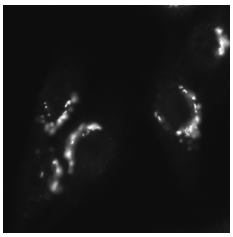
This dataset is provided by Sullivan Nicolaides Pathology Laboratory. It contains 419 samples of patients, which were prepared on the 18-well slide of HEP-2000 IIF assay from Immuno Concepts N.A. Ltd. The images were captured using a monochrome high dynamic range microscopy camera. Approximately 100-200 cell images were extracted from each patient serum. This dataset contains 252

specimen images with the size of 2776×2080 pixels in seven classes. For ease of computation, each image is divided to four parts and the mask of them are also provided. In total there were 68,429 cell images extracted including 13,596 for training (publicly available[2]) and 54,833 for testing.

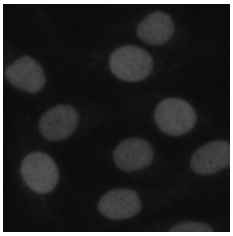
Unlike the ICPR2012 dataset, this dataset contains gray-scale images and the number of data is much more than previous dataset. The specimen level images with different classes are described as:



Centromere (Ce): Discrete small spots that are observed in the nuclear chromatin in a scattering fashion.

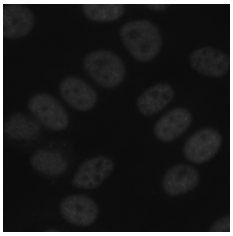


Golgi (G): Composed of irregular large granules which are stained adjacent to the boundaries of nucleus and around chromosomal material.

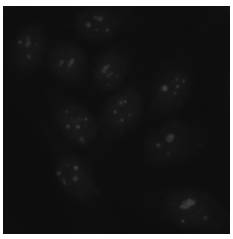


Homogeneous (H): The staining pattern is spread out over the interphase nuclei. These type of cell images have smooth diffused characteristics all over the nuclei area.

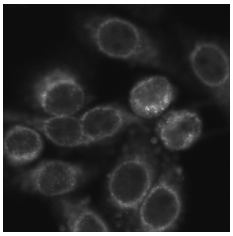
[2] http://i3a2014.unisa.it/?page_id=126



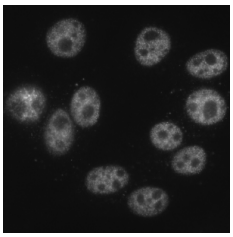
Mitotic Spindle (MP): staining only of the triangular or "banana-shaped" pole area of the mitotic spindle in the metaphase cells. This pattern is rare.



Nucleolar (N): Clustered particles can be observed in the cells' nucleoli. Without these granules, this pattern is close to Homogeneous class. At most six granules can be observed.



Nuclear Membrane (NM): Contains a tube-like of smooth homogeneous fluorescence in the interphase cells.



Speckled (S): Contains two subcategories of fine- and coarse-speckled. Various sized speckled can be observed densely distributed throughout nucleoplasm.

For more information regarding the related AD to each staining patterns and high resolution images, please refer to the Appendix.

Each annotated cell image in this dataset contains information of cell pattern, intensity level (*positive* or *intermediate*), mask and specimen image number. The intensity level can be used as an informative prior knowledge in the experiments.

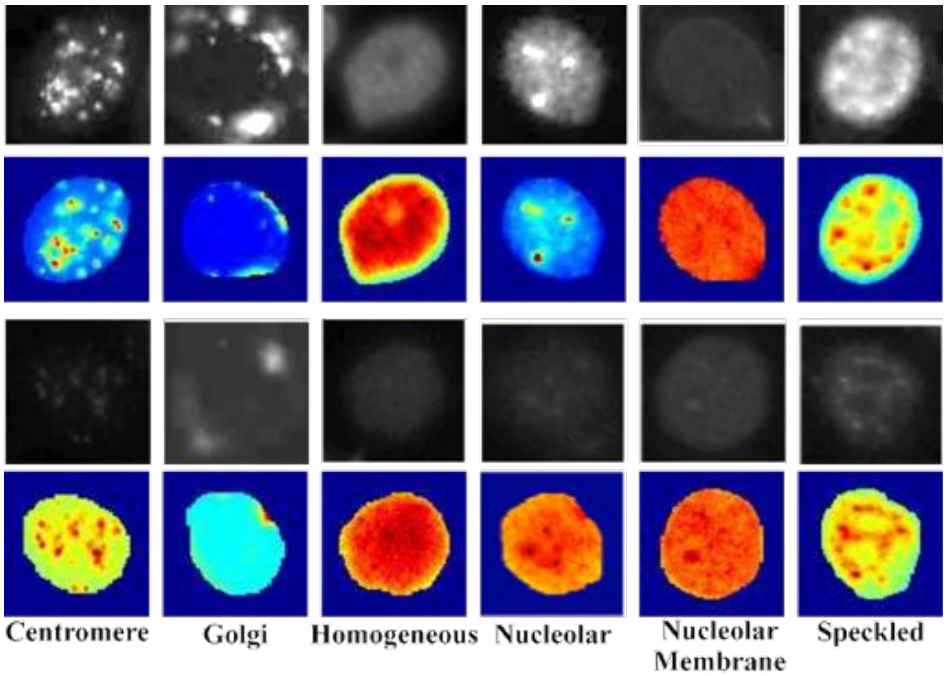


Figure 3.5: The *cell level* images of ICIP2013 for *positive* (top two rows) and *intermediate* (bottom two rows) intensity levels and their heat-maps to show the underlying pattern.

The cell level classification in ICIP2013 dataset contains six classes which are shown in Fig. 3.5. The top two rows are the *positive* cell images and the bottom two rows are indicated the *intermediate* cell images. For both types, the heat maps are also shown to understand the underlying patterns.

Table. 3.4 shows the number of cells in each image for different patterns and intensities for this dataset. The numbers in parenthesis are the number of cells in corresponding specimen image.

This number for 'Mitosis Spindle' is zero, because this class is not considered in the Cell level classification. As evident in Table. 3.4, the number of *intermediate* cells are more than *positive* ones and the Golgi cell class has lower number of images. These priors can also be considered in the experiments.

Table 3.4: The statistics of cell and specimen images for the ICIP2013 training dataset. The values in parentheses are the number of cells in specimen images.

ICIP2013 Dataset	Centromere	Golgi	Homogeneous	Mitosis Spindle	Nucleolar	Nucleolar Membrane	Speckled	**Total**
Positive	26 (1378)	4 (349)	26 (1087)	5 (0)	17 (934)	7 (943)	25 (1457)	110 (6148)
Intermediate	25 (1363)	6 (375)	27 (1407)	10 (0)	33 (1664)	14 (1265)	27 (1374)	142 (7448)
Total	51 (2741)	10 (724)	53 (2494)	15 (0)	50 (2598)	21 (2208)	52 (2831)	**252 (13596)**

3.4 Experiments and Results

The proposed method is tested on two publicly available datasets. For each dataset, two experiments on *Cell Level* and *Specimen Level* are designed to classify the individual cell and the specimen image, respectively. For evaluations, the Mean Class Accuracy (MCA) is adopted as used in the ICIP2013 contest: $MCA = \frac{1}{K}\sum_{k=1}^{K} CCR_k$ where CCR_k is the correct classification rate for class k and K denotes the number of classes.

3.4.1 Evaluation on ICPR2012 Dataset

Cell Level Accuracy

The *Cell Level* classification is performed under the typical setup as describe in Section 3.2, namely, SIFT and SURF feature concatenation for dictionary learning, 40 iteration dictionary learning with 2048 words produced, sparse coding with max-pooling for SVM classification. Overall, a *Cell Level* accuracy of 78.2% is obtained for the ICPR2012 dataset. This accuracy is 3.2% higher than the best accuracy reported in [Theodorakopoulos et al., 2014] and almost 9% higher than the winning accuracy in the ICPR2012 contest (Nosaka [Nosaka and Fukui, 2014]). Table. 3.5 shows the accuracy of the proposed technique and ten state-of-the-art

Table 3.5: The cell-level accuracy for the ICPR2012 dataset.

Name	Cell Level	Positive	Intermediate	Average Accuracy	Specimen Level
SBoW	**78**	**82**	**65**	**74**	93
Ensafi⋆	73	81	62	71	86
Kastaniotis*	75	70	31	51	86
Shen†	74	27	7	17	86
DiCataldo⋄	72	60	35	48	93
Kazanov‡	71	62	41	52	**100**
Faraki°	70	-	-	-	79
Nosaka•	69	74	35	55	86
Wiliem∓	67	69	48	59	71
Xiangfei⊗	67	78	48	63	93
Stoklasa⊛	64	74	35	55	79

⋆ [Ensafi et al., 2014a] * [Theodorakopoulos et al., 2014] † [Shen et al., 2014]
⋄ [Di Cataldo et al., 2014] ‡ [Ponomarev et al., 2014] ° [Faraki et al., 2014]
• [Nosaka and Fukui, 2014] ∓ [Wiliem et al., 2014] ⊗ [Kong et al., 2014]
⊛ [Stoklasa et al., 2014]

techniques as reported in the ICPR2012 contest [Foggia et al., 2013].

The dataset also contains the intensity information of the cells, namely *positive* and *intermediate* classes. Our proposed technique obtains superior cell classification accuracy of 82% and 65% for the *positive* and *intermediate* intensity level, respectively, as shown in Table. 3.5. Table 3.6a further shows the confusion matrix of *Cell Level* accuracy, which indicates high misclassification rate between Fine-, Coarse-Speckled and Homogeneous classes due to the similar underlying patterns of these classes.

Additionally, the Fig. 3.6 and 3.7 show the comparison of cell level accuracies on the test set with all the participants in the ICPR 2012 contest. As it is evident we outperform other methods.

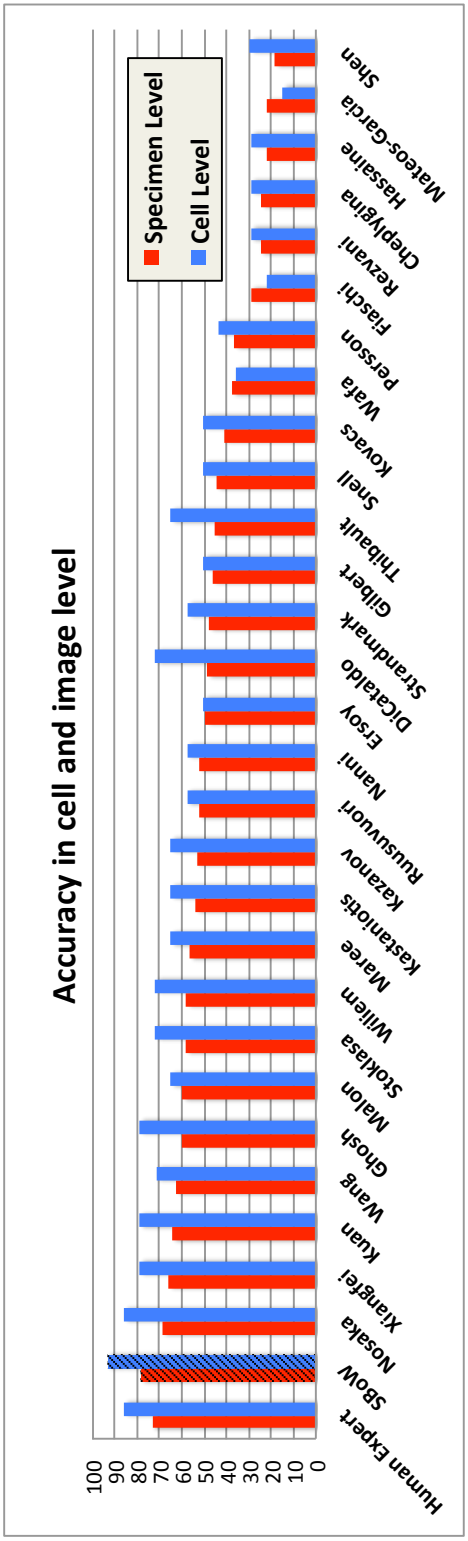


Figure 3.6: Cell and Specimen Level accuracies on test set vs. all the participants' method in ICPR2012 contest.

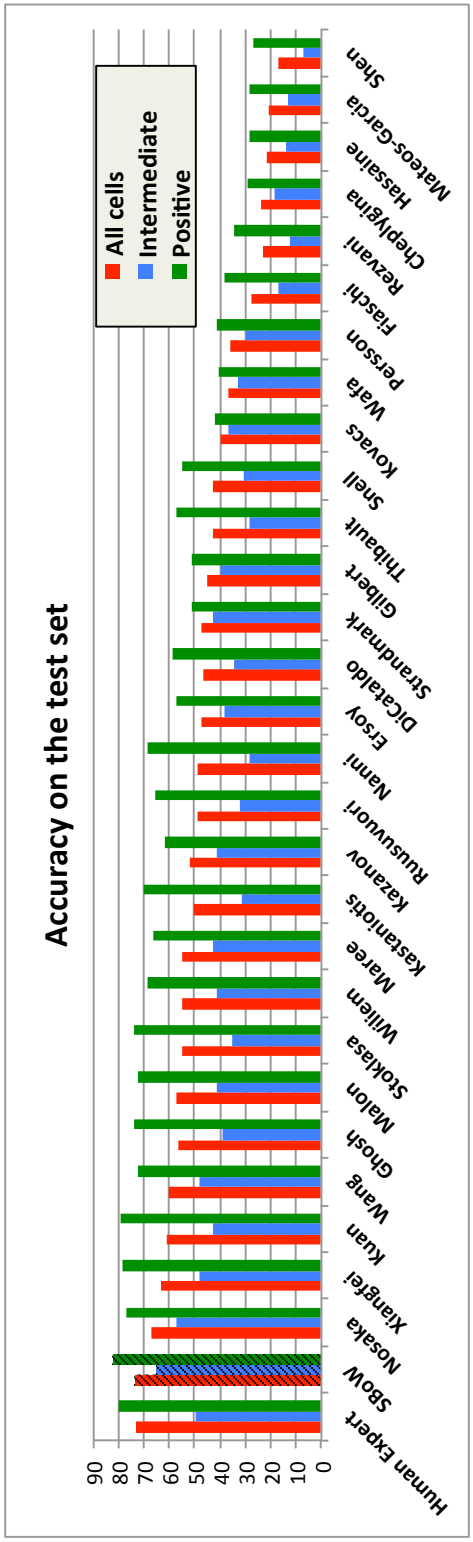


Figure 3.7: Two intensity Level accuracies on test set vs. all the participants' method in ICPR2012 contest.

Table 3.6: The confusion matrices for *Cell Level* (a) and *Specimen Level* (b) by using SIFT+SURF for ICPR2012 dataset.

Confusion Matrix	Ce	CS	Cy	FS	H	N
Ce	**89.93**	0.67	0.00	0.00	0.00	9.40
CS	1.98	**73.27**	2.97	18.81	0.99	1.98
Cy	0.00	1.96	**98.04**	0.00	0.00	0.00
FS	5.26	15.79	2.63	**52.63**	23.68	0.00
H	2.22	1.67	0.00	18.33	**75.00**	2.78
N	10.07	5.76	0.72	1.44	1.44	**80.58**

(a) Cell Level

Confusion Matrix	Ce	CS	Cy	FS	H	N
Ce	**100.0**	0.0	0.0	0.0	0.0	0.0
CS	0.0	**66.77**	0.0	33.3	0.0	0.0
Cy	0.00	0.0	**100.0**	0.0	0.0	0.0
FS	0.0	50.0	0.0	**50.0**	0.0	0.0
H	0.0	0.0	0.0	0.0	**100.0**	0.0
N	0.0	0.0	0.0	0.0	0.0	**100.0**

(b) Specimen Level

Specimen Level Accuracy

For the *Specimen Level* classification, all cells in one specimen image are classified and the maximum votes of the cell classes are taken as the specimen class. For the specimen images, the proposed technique achieves an accuracy of 93% as shown in Table. 3.5. The best specimen classification accuracy is achieved by Kazanov [Foggia et al., 2013] but we have two misclassifications between Fine- and Coarse-Speckled classes as can be seen in the confusion matrix in Table. 3.6b due to the similar patterns of these two classes.

3.4.2 Evaluation on ICIP2013 Dataset

Cell Level Accuracy

This dataset has only a training set publicly available. To provide fair comparison with other methods, we follow the evaluation method as reported in [Han et al., 2014], where 600 cell images are randomly selected for each cell class as the training set (except Golgi class where 300 cell images are selected) and the rest for testing. Then the *Cell Level* classification is performed 20 times on *positive* and

intermediate cells and the average accuracy is reported. Table. 3.7 shows the classification accuracies in cell and specimen images. As Table. 3.7 shows, there are up to 3% and 5% improvements for the *positive* and *intermediate* cell classification, respectively. This much better accuracy is largely because of the sparse representation of the input features and their combinations in different scales which reduces the global information lost and provides better discriminative features.

Additionally, Table. 3.8a shows the confusion matrix which highlights the misclassification between Homogeneous, Speckled and Centromere, Nucleolar-Membrane classes. The misclassification is largely due to the very similar underlying pattern of these classes especially in *intermediate* cells as can be seen in Fig. 3.5.

Specimen Level Accuracy

The cells in each specimen image should be extracted and classified in order to estimate the *Specimen Level* classification. We therefore extract cells by combining several image feature instead of using the provided cell masks as described in Section. 3.2.2. For the 252 specimens, our cell extraction method extract 5816 cells which is used as the *Cell Level* dataset as shown in Table. 3.9. For evaluation, half of specimen images from each class is randomly selected for training and the rest for testing. The evaluation is performed 10 times and the average accuracy is reported as shown in Table. 3.7.

In this experiment, the results are compared using the extracted cells (see Section. 3.2.2) versus the cells provided by the cell masks. As Table. 3.7 shows, the proposed cell extraction method helps improve the classification accuracy

significantly. Additionally, the time complexity of the proposed cell extraction method is an order of magnitude lower than that considering all the cells in the specimen images.

The confusion matrix of *Specimen Level* classification is shown in Table. 3.8b, which shows high misclassification between Nucleolar Membrane and Homogeneous due to their very similar patterns. There is also high misclassification between Golgi and Mitosis Spindle classes, which is largely due to the low number of training samples of these classes compared with other classes as shown in Table. 3.9.

Table 3.7: Accuracy on the ICIP2013 dataset.

ICIP2013	Cell Level			Specimen Level	
	Han⋆	Ensafi†	**SBoW**	All Cells	**SBoW**
Positive cell images	95.5%	95.8%	**98.12%**	89.21%	**93.26%**
Intermediate cell images	80.9%	87.9%	**92.78%**	88.43%	**92.12%**
Average	88.2%	91.9%	**95.45%**	88.82%	**92.69%**

⋆ [Han et al., 2014] † [Ensafi et al., 2014b]

Table 3.8: The confusion matrices for *Cell Level* (a) and *Specimen Level* (b) by using SIFT+SURF for ICIP2013 dataset.

Confusion Matrix	Ce	G	H	N	NuMem	S
Ce	**96.34**	0.00	0.00	0.00	2.17	1.49
G	0.00	**98.47**	1.12	0.00	0.00	0.41
H	0.00	0.97	**96.14**	0.91	0.00	1.98
N	0.00	1.14	2.12	**96.19**	0.00	0.55
NuMem	2.24	1.12	0.00	0.00	**95.32**	1.32
S	1.08	0.21	7.26	0.92	0.29	**90.24**

(a) Cell Level

Confusion Matrix	Ce	G	H	MitSp	N	NuMem	S
Ce	**94.61**	0.00	1.12	0.00	0.00	0.00	4.27
G	0.00	**93.21**	0.03	6.76	0.00	0.00	0.00
H	0.00	0.44	**93.78**	0.00	0.00	0.00	5.78
MitSp	0.00	5.34	5.46	**86.19**	0.00	3.01	0.00
N	0.00	0.00	0.78	0.04	**99.18**	0.00	0.00
NuMem	0.00	0.00	14.74	0.00	0.00	**85.26**	0.00
S	1.70	0.00	1.73	0.00	0.00	0.00	**96.57**

(b) Specimen Level

Table 3.9: The statistics of extracted cells for the ICIP2013 training database. The values in parentheses are the number of cells in specimen images.

Extracted Cells	Centromere	Golgi	Homogeneous	Mitosis Spindle	Nucleolar	Nucleolar Membrane	Speckled	**Total**
Positive	26 (603)	4 (90)	26 (623)	5 (106)	17 (379)	7 (158)	25 (639)	110 (2598)
Intermediate	25 (523)	6 (124)	27 (664)	10 (211)	33 (766)	14 (294)	27 (649)	142 (3218)
Total	51 (1126)	10 (214)	53 (1287)	15 (317)	50 (1145)	21 (452)	52 (1288)	**252 (5816)**

3.4.3 Discussion

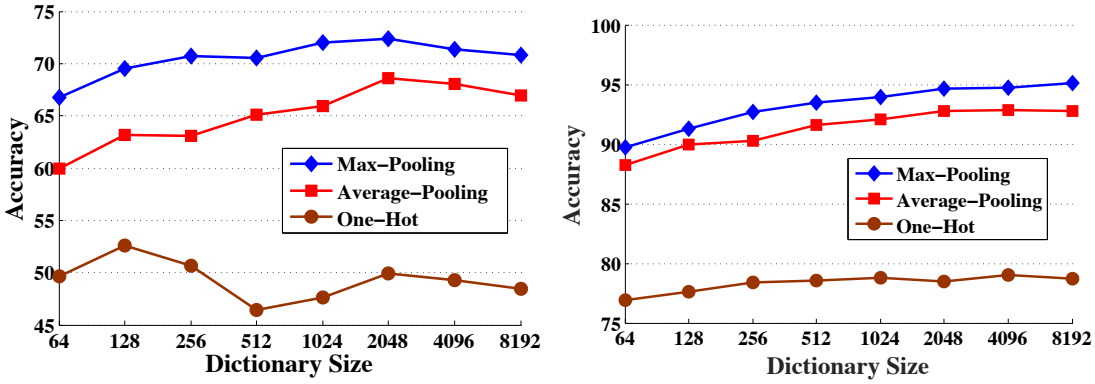
Dictionary Size

The performance of the proposed technique is closely related to the dictionary size. We study this issue by doubling the dictionary size in each iteration starting from 64. Fig. 3.8 shows the cell classification accuracy with different dictionary sizes. As Fig. 3.8 shows, the classification accuracy first increases with dictionary size and saturates when the dictionary size reaches certain number. One possible explanation for this result is that dictionaries with a larger size produce more discriminative sparse codes which lead to better classification accuracy.

Generally the accuracy of ICIP2013 dataset is clearly higher than that of ICPR2012 dataset. The better accuracy is largely due to the large number of available training cells in ICIP2013 dataset (see Section. 3.3.2), which helps learn more discriminative features and better classifiers.

Pooling Strategy

Three types of pooling strategy are studied including one-hot encoding, average- and max-pooling. The first strategy is performed on the sparse codes of different regions in scaled images which measures only one maximum code in each region



(a) ICPR2012 (b) ICIP2013

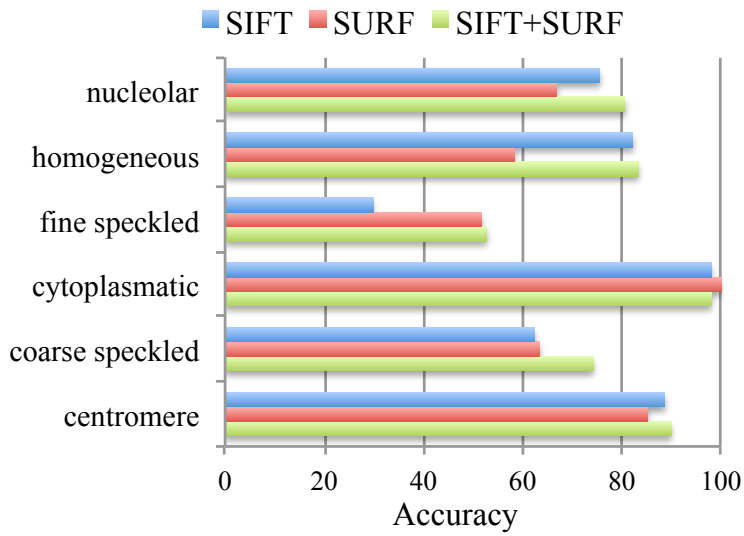
Figure 3.8: The study of different dictionary size and pooling strategy.

and concatenates to build the final feature vector. This method underestimates the large amount of information which can be extracted from the input images and results in poor performance as shown in Fig. 3.8. As a comparison, both average-pooling and max-pooling outperform the one-hot encoding clearly, and the max-pooling obtains the best accuracy.

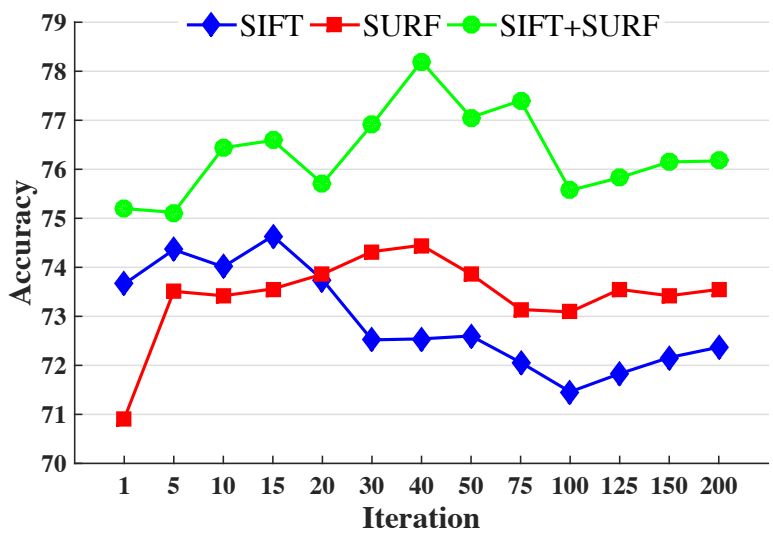
The different performance can be explained that in one-hot encoding strategy, important information could be lost by just focusing on the most representative feature rather than a set of features. Additionally, by performing the average-strategy, the sharpness of the histogram of features are blared which results in the loss of important information of the input images.

SIFT, SURF and Learning Iteration

We study the complementation between the SIFT and SURF for the HEp-2 classification problem. Three different dictionaries and SVM classifiers are trained by using SIFT, SURF and SIFT+SURF features, respectively. Fig. 3.9a shows



(a)



(b)

Figure 3.9: Classification accuracy on ICPR2012 dataset by using different feature sets (a) and correlation between the number of iterations in the dictionary learning process (b).

experimental results on the ICPR2012 dataset by using the three different sets of features.

As Fig. 3.9a shows, SIFT and SURF perform similarly for Cytoplasmatic, Coarse-speckled, and Centromere cells. Besides, SIFT outperforms SURF for Nucleolar and Homogeneous cells and SURF outperforms SIFT for Fine-speckled cells. On the other hand, the combination of SIFT and SURF outperforms the SIFT greatly for Fine-speckled and Coarse-speckled and perform similarly to the SIFT for the rest four cell types. Overall, the combination of SIFT and SURF obtains an average accuracy of 78.2% over all six cell types, which is clearly higher than 74.5% and 70.9% that are obtained by using the SIFT and SURF alone, respectively.

To study the effects of the iteration number on the performance of our algorithm, we conducted a series of experiments and have three observations. First, the iteration number does have certain effects on the performance of our algorithm. Second, the effect is still within a constrained range where the classification accuracy changes within the range of 70% - 78% when the iteration number changes as shown in Fig. 3.9b in the revised manuscript. Third, the experiments show that the optimal classification accuracy is obtained when the iteration number is around 30 - 50. This can be a useful reference for the ensuing research and development on this topic. More interestingly, the proposed technique is capable of obtaining state-of-the-art accuracy (around 75%) with the first iteration of the dictionary learning. This is very important for the reduction of the computation costs which is often heavy due to the extraction of the SIFT and SURF features as well as the minimization problem involved.

3.5 Summary

The BoW model on sparse coding scheme is proposed for the HEp-2 cell classification problem. The dictionary size, pooling strategy, effects of integrating the SIFT and SURF features and the optimal number of iterations for dictionary learning are investigated. Our experiments show that a larger dictionary size usually leads to a better classification performance. Additionally, the max-pooling strategy works better than both average-pooling and one-hot encoding method. Experiments on two public benchmarking datasets show superior classification accuracy at both cell level and specimen level.

Manually assigning the dictionary size increases the computational complexity of dictionary learning procedure and tuning this parameter is done by try and error scheme. In the next chapter a non-parametric Bayesian method is proposed to estimate the dictionary size.

Chapter 4

Non-Parametric Bayesian Method for Dictionary Learning

4.1 Introduction

In this chapter we propose a Sparse Non-Parametric Bayesian (SNPB) model and implement it for the HEp-2 cell classification problem, targeting applications for computer-aided AD diagnosis. In the SNPB model, the dictionary learning exploits the non-parametric sparse factor analysis (NSFA) [Knowles et al., 2011, Zonoobi et al., 2014a, Zonoobi et al., 2014b] that is capable of determining the dimension of the dictionary words automatically. In particular, the "non-parametric" here means that the dimension of the dictionary can be extended to infinity at the beginning of the learning process and lead to the correct and most efficient value at the end. The learned sparse representation of the codes is used as the final feature for the HEp-2 cell classification.

Variety of methods are proposed to use the dictionary learning method where most of them assigned manual values for the dictionary learning procedures and nearest neighbors [Di Cataldo et al., 2014, Shen et al., 2014, Ensafi et al., 2014a, Wiliem et al., 2014]. However, there is a trade off between performance of the system and the dimension of the dictionary [Ensafi et al., 2014b], where the large values for dictionary dimension results in better accuracies but affect the computational cost. To the best of our knowledge, there is no study of finding the best low-dimensional dictionary for the HEp-2 cell classification problem, which is one of the novelties of this work.

In the rest of the chapter, we describe the proposed CAD system in Section 4.2 and explain the experiments and results in Section 4.3 on two benchmarking datasets.Finally, we have the discussion and summary sections in 4.3.3 and 4.4 respectively.

4.2 Method

The proposed SNPB method is depicted in Fig. 4.1. First the SIFT and SURF features are extracted from the masked images in a grid manner. Then a dictionary is learned by using the non-parametric Bayesian method, which can estimate the dimension of the dictionary automatically. By transferring the input features to sparse codes by means of the learned dictionary, we scaled them to three layers and then the max pooling approach makes the output feature vectors.

In other words, three layers of codes are used. The first layer is the all sparse codes, second one is the divided image to four regions and the last layer is divided

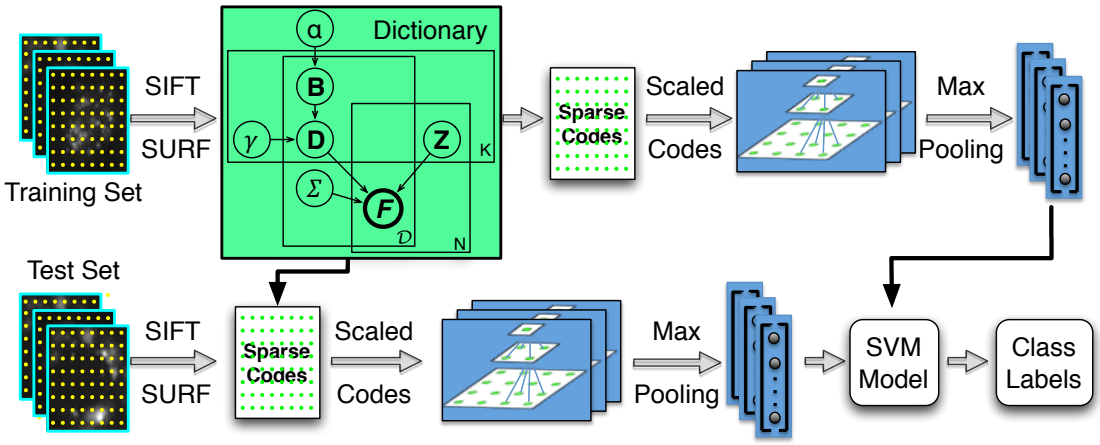


Figure 4.1: The SNPB framework of the HEp-2 cell classification.

to 16 regions. Totally there are 21 regions where in each region the maximum bin of the histograms of codes is calculated. Finally, the concatenated sparse code of each region is used as the final feature vector for each image. By assuming the dimension of each feature vector is $\mathcal{D}$, the final feature vector dimension is $21\times\mathcal{D}$. In training stage, these features are then fed to the multi class (One-Versus-All) linear SVM to classify the input training images with their ground truth labels. The same procedure is applied on the test images by using the pre-learned dictionary and SVM classifier.

4.2.1 Dictionary Learning

The dictionary learning method which is used is based on non-parametric Bayesian method [Gershman and Blei, 2012] that makes use of the sparse prior knowledge on coefficients of dictionary words based on Indian Buffet Process (IBP) [Griffiths and Ghahramani, 2011]. The graphical model of this method is shown in Fig. 4.1 in the dictionary part [Knowles et al., 2011].

Let $\mathbf{F}$ be a set of features in a $\mathcal{D}$-dimensional space, $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N] \in \mathbb{R}^{(\mathcal{D}\times N)}$, and $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K] \in \mathbb{R}^{(\mathcal{D}\times K)}$ are K words of our dictionary to be estimate. We can write:

$$\mathbf{f}_n = \mathbf{D}\mathbf{z}_n + \mathbf{e}_n, \qquad \mathbf{D} = \gamma \odot \mathbf{B} \tag{4.1}$$

where γ_k is the precision (inverse variance) of the kth word in dictionary and $\mathbf{B}$ is a binary matrix. The indicator function $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N]^T$ contains the weights of the words in the dictionary and we need it to be as sparse as possible and $\mathbf{e}$ is the noise vectors for each dictionary words, usually assumed to be Gaussian with diagonal covariance matrix Σ_e for each dimension. Here we assume both indicator function $\mathbf{Z}$ and dictionary $\mathbf{D}$ are hidden variables of our non-parametric model and we want to infer the posterior distribution given the input feature vectors. Now we can model our dictionary by "spike and slab" distribution as

$$P(\mathbf{D}_{dk}|\mathbf{B}_{dk}, \gamma_k) = \mathbf{B}_{dk}\mathcal{N}(\mathbf{D}_{dk}; 0, \gamma_k^{-1}) + (1 - \mathbf{B}_{dk})\delta_0(\mathbf{D}_{dk}) \tag{4.2}$$

where δ_0 is the delta function. In this model we want to estimate the number of dictionary words K. Therefore, the $\mathbf{B}$ matrix should have infinite columns in initial step. To do so, we make use of IBP, which provides a sparse matrix of intuitively infinite dimension. In this regard, we can assume that we have finite K model and then take the limit to $K \to \infty$. To provide the $\mathbf{B}$ matrix by IBP, we assume that the rows are generated separately and a probability of source k

contributing to any dimension is π_k. Then we can write

$$P(\mathbf{B}|\pi) = \prod_{k=1}^{K}\prod_{d=1}^{\mathcal{D}} P(\mathbf{B}_{dk}|\pi_k) = \prod_{k=1}^{K} \pi_k^{m_k}(1-\pi_k)^{\mathcal{D}-m_k} \tag{4.3}$$

where m_k is the number of nonzero elements of column k in $\mathbf{B}$. Because the product is the binomial distribution, we can use the conjugate Beta(r,s) distribution for π_k. We can assume α as the strength parameter in $r = \frac{\alpha}{K}$ and $s = 1$. Then we can define the model as

$$\mathbf{B}_{dk}|\pi_k \sim Bernoulli(\pi_k) \tag{4.4}$$

$$\pi_k|\alpha \sim Beta\left(\frac{\alpha}{K}, 1\right) \tag{4.5}$$

By integrating out π we have

$$P(\mathbf{B}) = \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(\mathcal{D} - m_k + 1)}{\Gamma(\mathcal{D} + 1 + \frac{\alpha}{K})} \tag{4.6}$$

where $\Gamma(.)$ is the Gamma function. By defining the method proposed by [Griffiths and Ghahramani, 2011] we can have the infinite limit of equation (4.6) as

$$P(\mathbf{B}) = \frac{\alpha^{K_+}}{\prod_{h>0}} K_h! \exp(-\alpha H_{\mathcal{D}}) \prod_{k=1}^{K_+} \frac{(\mathcal{D} - m_k)!(m_k - 1)!}{N!} \tag{4.7}$$

where K_+ is the number of non-zero column of $\mathbf{B}$, $H_{\mathcal{D}} = \sum_{j=1}^{\mathcal{D}} \frac{1}{j}$ is the $\mathcal{D}^{th}$ harmonic number and K_h is the number of rows whose entries correspond t the binary number h.

To provide a sparse matrix with the distribution in (4.7), the Indian Buffet

Process starts from the first row and samples Poisson(α) columns. To generate the i_{th} row, IBP samples from the columns which have been sampled in previous rows with the probability of $\frac{m_k}{i}$ and samples Poisson($\frac{\alpha}{i}$) from the new columns. Here m_k is the number of nonzero elements of column k in $\mathbf{B}$. The large values of α produce the matrix with relatively large number of columns.

For inference, the Markov Chain Monte Carlo (MCMC) method is used, which defines a Markov chain on the hidden variables ($\mathbf{B}$, $\mathbf{Z}$) and maximizes the posteriors. In other words, in each iteration the $\mathbf{B}$, $\mathbf{Z}$ matrices are sampled using Gibbs sampling strategy (which is a simple form of MCMC) and the posterior probability is maximized.

Additionally, we can sample the hyper-parameter of IBP (α) as well using conjugate $Gamma(a_1, a_2)$ prior by the likelihood term of equation (4.7),

$$P(\alpha|\mathbf{B}) \propto P(\mathbf{B}|\alpha)P(\alpha) = Gamma(K_+ + a_1, H_{\mathcal{D}} + a_2). \tag{4.8}$$

where a_1 and a_2 are constant values [Knowles et al., 2011].

4.2.2 Sparse Coding

Next stage after calculating the optimal dictionary is to code the input images sparsely [Zonoobi and Kassim, 2013]. We use the efficient sparse coding scheme [Lee et al., 2006] by fixing the dictionary words, which results in (4.9),

$$\min_{\mathbf{z}} \sum_{n=1}^{N} \left\|\mathbf{f}_n - \mathbf{D}\mathbf{z}_n\right\|^2 + \lambda|\mathbf{z}_n| \tag{4.9}$$

However, if we know the sign of each code in $\mathbf{z}_n$ then we can replace it with either $\mathbf{z}_n$ or $-\mathbf{z}_n$, then the resulting formulation will change to a simple Quadratic optimization Problem (QP). By guessing the initial values, we can refine it by solving this QP using least squares.

4.3 Experiments And Results

Datasets

Two publicly available datasets namely MIVIA HEp-2 (ICPR2012) and ICIP2013 are used in this experiment. The former dataset has training and test sets in six classes but the latter one has huge number of cells in its training set and the test set remained as an evaluation set for the organizers, which is not published so far. Both datasets contain IIF images with several cells in them. The maskS of the cells are provided in order to classify the cells without considering other neighboring cells named *cell level* classification (see Chapter 3.3). Additionally, it is assumed that the cells in each image belong to one class, which defines the *image level* classification problem.

4.3.1 Optimizing the Dimension of Dictionary

The proposed non-parametric Bayesian method is used to estimate the optimum dictionary dimension in both datasets. As can be seen in Fig. 4.2, the dimension of dictionaries (K), increase in the starting iterations and finally converge to their optimums. By this method, the dimension of the *positive* and *intermediate* intensity level dictionaries in ICPR2012 are calculated 28 and 18, respectively. For

Table 4.1: The MCA on test set of ICPR2012 dataset [Foggia et al., 2013]. The Values in the parentheses are dictionary dimensions.

ICPR2012	SNPB-(D)	Ensafi-(D)*	Shen*	DiCataldo†	Kazanov$^{\diamond}$	Siram‡	Nosaka$^{\circ}$	Wiliem$^{\bullet}$	Xiangfei$^{\mp}$	Stoklasa$^{\circledast}$
Cell Level	**75 (38)**	72 (1024)	74	72	71	62	69	67	67	64
Image Level	93 (38)	86 (1024)	86	93	**100**	86	79	71	93	79
Positive	**82 (28)**	81 (1024)	27	60	62	63	74	69	78	74
Intermediate	59 (18)	**62 (1024)**	7	35	41	60	35	48	48	35
Average Accuracy	69	**72**	17	48	52	62	55	59	63	55

* [Ensafi et al., 2014a] * [Shen et al., 2014] † [Di Cataldo et al., 2014] $^{\diamond}$ [Foggia et al., 2013]
‡ [Sriram et al., 2014] $^{\circ}$ [Nosaka and Fukui, 2014] $^{\bullet}$ [Wiliem et al., 2014] $^{\mp}$ [Han et al., 2014]
$^{\otimes}$ [Foggia et al., 2013] $^{\circledast}$ [Foggia et al., 2013]

ICIP2013 datasets, these values are estimated 139 and 123 respectively. Whereas, these values are manually selected to 1024 as the state-of-the-art results in [Ensafi et al., 2014a, Ensafi et al., 2014b]. The increasing slope of the charts in Fig. 4.2 proves that the size of the dictionary matrix is intuitively infinite and by optimizing the model, it decreases to a minimum value in its steady state. Additionally, by having low dimensional dictionaries, the dimension of final sparse codes and complexity of calculating them are decreased as well.

4.3.2 Evaluation

To evaluate the method, the MCA (see Section 3.4) is used. Additionally, in all the evaluation procedures, the dictionaries are learned on the training set only.

For the ICPR2012 dataset, the test set is available for evaluation. Table. 4.1 shows the accuracies of ICPR2012 dataset versus the other methods. As can be seen in this table, although we have learned a low dimension dictionary, a better accuracy is achieved in cell level and *positive* intensity level in comparison with

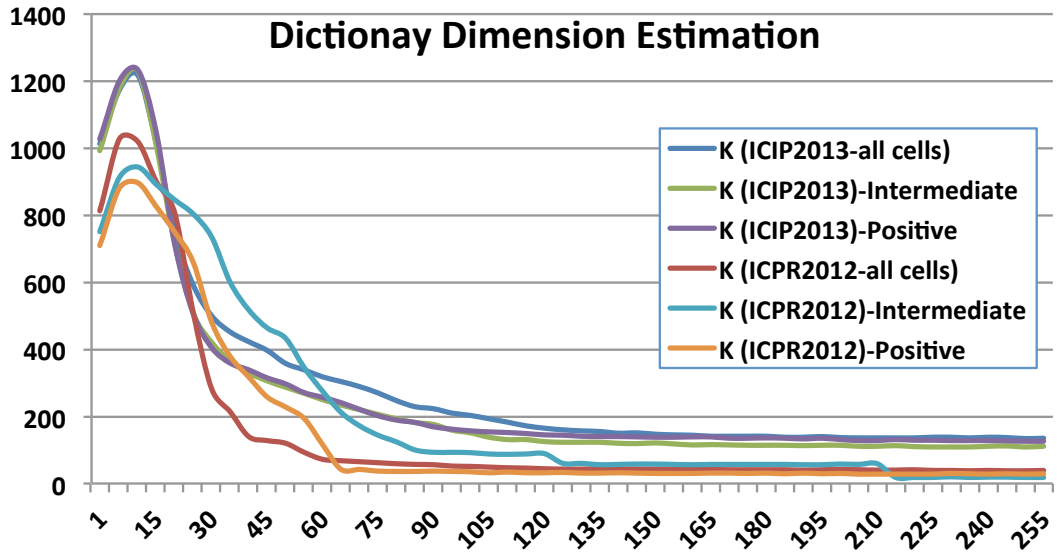


Figure 4.2: The dictionary dimensions for ICIP2013 and ICPR2012 datasets versus the iteration of the algorithm.

other methods.

For the ICIP2013 dataset, in order to compare our results with the method of [Han et al., 2014], the accuracies are achieved using 600 randomly selected images for training and the rest for testing. This evaluation is performed on *positive* and intensity level images as well and the results are stated in Table. 4.2. Additionally, the method of [Ensafi et al., 2014b], which we call it Sparse Coding (SC) method is evaluated using the same randomly selected images. In this method the dictionary dimension is manually defined to 1024 as the authors suggest. As we can see in the TABLE 4.2, the SNPB results are better than the SC method by having low dimensional dictionaries as well.

Table 4.2: The MCA of ICIP2013 dataset.

ICIP2013 Dataset	**Accuracy**		
	SNPB-(Dimension)	SC-(dimension)	[Han et al., 2014]
Positive images	**96.8% (139)**	95.1% (1024)	95.5%
Intermediate images	**88.8% (123)**	87.9% (1024)	80.9%
Average	**92.8%**	91.5%	88.2%

4.3.3 Discussion

The SNBP model obtained state-of-the-art result in *Cell Level* (75.2%) and *positive* intensity cells (82.6%) respectively in comparison with other methods for ICPR2012 dataset. Additionally, the dimension of the learned dictionary for *positive* and *intermediate* intensity levels are 28 and 18 respectively, which are more than 36 times smaller than the other dictionary based models as in [Ensafi et al., 2014a], which is manually selected to 1024. This dominant reduction of dictionary size is a great beneficial for calculating the sparse codes and classifying the test images.

For the ICIP2013 dataset, the state-of-the-art accuracies are obtained with comparing to [Han et al., 2014] and [Ensafi et al., 2014b] by considering the 600 randomly selected images for training and the rest for testing. Additionally, the advantage of the proposed method is its lower dimension of the learned dictionary, 139 and 123 for *positive* and *intermediate* intensity level dictionaries respectively, which are almost 8 times smaller than the previous models.

4.4 Summary

A Sparse Non-Parametric Bayesian (SNPB) model is proposed for automatic classification of the HEp-2 cell images. The prevalent approach uses sparse coding and Bag of Words models which depends highly on the dictionary size that is usually selected in a manual manner. The Indian Buffet Process provides prior knowledge of sparse codes and takes advantage of intuitively infinite matrix dimension which is exploited to produce an optimal dictionary size automatically. Experiments show that the dimension of the proposed model is 28 and 8 times smaller than the similar BoW methods in ICPR2012 and ICIP2013 datasets respectively. Additionally, the lower dimension of learned dictionary leads to lower computational time in the test procedure.

One bottleneck of this method is the large amount of overlapped image patches which are participated in the dictionary learning procedure. Additionally, the size and shape of the patches correlate the amount of information carried by the patches. Therefore, an efficient method is needed to locate and shape the patches to satisfy the problems. In the next chapter, a distributed dictionary learning method is proposed to addresses the HEp-2 cell classification problem in a computationally efficient and less memory intensive way.

Chapter 5

Adaptive Distributed Dictionary Learning

5.1 Introduction

In this chapter, we propose Adaptive Distributed Dictionary Learning (ADDL) method which addresses the HEp-2 cell classification problem in a computationally efficient and less memory intensive way compared to the other methods. To the best of our knowledge, this is the first time that a distributed dictionary learning method has been successfully implemented for image classification. In our proposed method, the dictionary matrix and the coding vector are partitioned into N blocks where each block is associated with a sub-dictionary and a sub-vector. Considering these blocks, we form a connected network of N nodes where each node is in charge of updating its own sub-dictionary. Each node is connected to a number of neighboring nodes sharing their information to update the sub-dictionaries. Es-

sentially, we reformulate the dictionary learning problem as a distributed learning task over the network and use the diffusion adaptation strategy [Sayed, 2014, Tu and Sayed, 2012] to solve this distributed problem. Moreover, we propose to combine the information of neighboring nodes in an adaptive way which results in superior performance. This adaptive approach enables the nodes to learn about the usefulness of the information received from their neighbors which helps the nodes to ignore misleading information.

The rest of this chapter is organized as follows. We introduce dictionary learning in a distributed manner in Section 5.2. The experiments and results obtained by testing the method on two datasets are provided in Section 5.3. The chapter is concluded in Section 5.4.

5.2 Overview of the ADDL Method

Fig. 5.1 shows our proposed ADDL method which extracts SURF (speeded-up robust features) and SIFT features of the images and uses them as inputs to the distributed dictionary learning. The learned dictionary is then used for HEp-2 cell classification where the sparse coding of image patches are combined with spatial pyramid matching (SPM) [Lazebnik et al., 2006]. As can be seen in Fig. 5.1, each input image is divided into 1, 4 and 16 regions within three pyramid layers and max-pooling is applied to the sparse codes of each region to obtain the final feature vector. Then SVM is learned to classify the cell images. Further details of the dictionary learning which is performed in a distributed manner is presented in the following sub-sections.

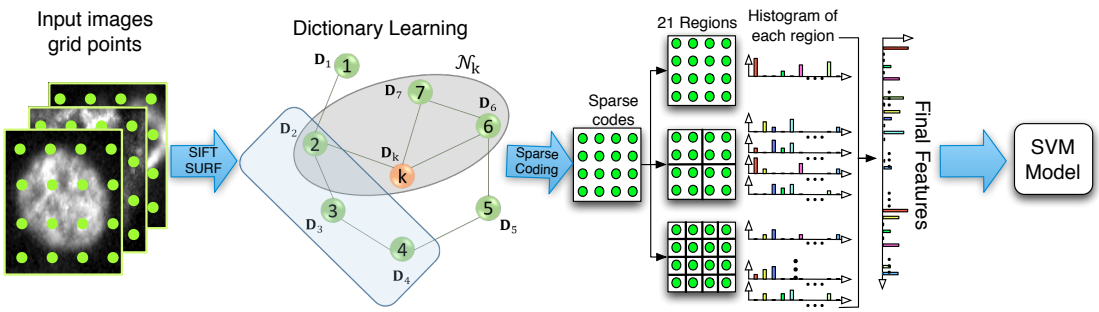


Figure 5.1: The proposed ADDL framework.

5.2.1 Dictionary Learning

By extracting the features of each image patch the input feature vector $\mathbf{F}_t$ for the dictionary learning algorithm is calculated. The dictionary learning problem can then be formulated as:

$$\min_{\boldsymbol{z}_t, \mathbf{D}} \left(\|\mathbf{F}_t - \mathbf{D}\boldsymbol{z}_t\|_2^2 + \lambda \|\boldsymbol{z}_t\|_1 + \frac{\beta}{2} \|\boldsymbol{z}_t\|_2^2 \right) \tag{5.1}$$

where $\mathbf{F}_t$ is the $M \times 1$ input feature vector at time t, $\mathbf{D}$ is an $M \times K$ dictionary matrix, $\boldsymbol{z}_t$ is the $K \times 1$ sparse code vector, and λ and β are the adjustable penalty (regularization) terms. It should be noted that the role of the ℓ_1-norm term $\|\boldsymbol{z}_t\|_1$ is to promote sparsity of the code vector while the Euclidean norm $\|\boldsymbol{z}_t\|_2$ ensures that the estimated values are small. To solve the optimization problem in (5.1), we introduce the distributed learning method.

5.2.2 Distributed Dictionary Learning

To learn the dictionary in a distributed manner we adopt the recently proposed approach presented in [Chen et al., 2015b]. In this method, the dictionary matrix

$\mathbf{D}$ and the coding vector $\boldsymbol{z}$ are partitioned into block forms:

$$\mathbf{D} = [\mathbf{D}_1 \dots \mathbf{D}_N], \quad \boldsymbol{z} = \text{col}\{\boldsymbol{z}_1, \dots, \boldsymbol{z}_N\} \tag{5.2}$$

where $\mathbf{D}_k$ is a sub-dictionary matrix of size $M \times N_k$ and $\boldsymbol{z}_k$ is a sub-vector of size $N_k \times 1$. Moreover, the summation over the sizes of sub-dictionaries is equal to the total size of the dictionary:

$$N_1 + \cdots + N_N = K \tag{5.3}$$

Now we form a connected network of N agents where each agent k in the network is responsible to update its own sub-dictionary $\mathbf{D}_k$ that is distributed over the network. As shown in Fig. 5.2, each agent in the network has a number of neighboring agents that it can interact with. Moreover, the input features $\mathbf{F}_t$ can only be presented to a subset of agents represented by $\mathcal{N}_{\mathcal{I}}$. Our experiments show that (see Section 5.3) providing the input data only to a subset of agents is computationally efficient while retaining comparable performance with other methods. This is due to the distributed nature of the network where the agents are allowed to interact and cooperate with their neighbors, resulting in dispersion of information over the network.

Considering (5.2) in the dictionary learning problem we can reformulate (5.1) as:

$$\min_{\boldsymbol{z},\mathbf{D}} \left(\|\mathbf{F}_t - \sum_{k=1}^{N} \mathbf{D}_k \boldsymbol{z}_k\|_2^2 + \sum_{k=1}^{N} (\lambda \|\boldsymbol{z}_k\|_1 + \frac{\beta}{2} \|\boldsymbol{z}_k\|_2^2)\right) \tag{5.4}$$

The linear combination of the sub-dictionaries $\mathbf{D}_k$ represents the input features

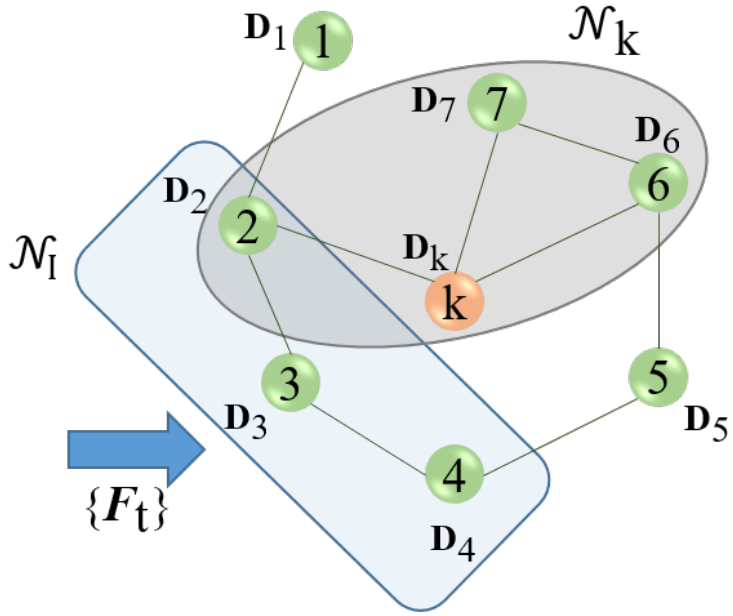


Figure 5.2: In this connected network, each agent k is responsible for learning a sub-dictionary $\mathbf{D}_k$ and is able to share information with its neighbors represented by $\mathcal{N}_k$. Also, input data $\mathbf{F}_t$ is presented to a subset of agents represented by $\mathcal{N}_{\mathcal{I}}$.

$\mathbf{F}_t$. It should be noted that the first term of (5.4) ensures that the reconstruction error is small while the role of the second term is to make the code vector sparse and small.

To solve the optimization problem of (5.4) in a distributed manner, the cost function should have a *"sum-of-costs"* form. Specifically, in order to apply distributed methods to tackle the problem at hand, the global cost function of the optimization problem, $J^{glob}(\boldsymbol{\omega})$, should be the aggregation of individual cost functions of the agents $J_k(\boldsymbol{\omega})$:

$$J^{glob}(\boldsymbol{\omega}) = \sum_{k=1}^{N} J_k(\boldsymbol{\omega}) \tag{5.5}$$

It has been shown in [Chen et al., 2015b] that the problem in (5.4) does not follow the form in (5.5) as it is *"cost-of-sums"* and not *"sum-of-costs"*. Therefore, it is not feasible to use distributed techniques for solving the problem in (5.4) directly. However, it has been shown that the dual problem of (5.4) has a distributed form similar to (5.5) and the optimal primal variables $\{\mathbf{D}_k\}$ and $\boldsymbol{z}$ can be recovered from the solution of the dual problem (see (5.14) and (5.15)). According to [Chen et al., 2015b], the dual problem can be formulated as:

$$\min_{\boldsymbol{\nu}} -g(\boldsymbol{\nu}, \mathbf{F}_t) = \|\boldsymbol{\nu}\|_2^2 - \boldsymbol{\nu}^T \mathbf{F}_t + \sum_{k=1}^{N} \mathcal{S}_{\frac{\lambda}{\beta}}(\frac{\mathbf{D}_k^T \boldsymbol{\nu}}{\beta}) \tag{5.6}$$

where $\boldsymbol{\nu}$ is the auxiliary vector variable of size $M \times 1$ in the dual problem, λ and β are the regularization coefficients in (5.1), and $\mathcal{S}_{\frac{\lambda}{\beta}}(x)$ is a function defined as:

$$\mathcal{S}_{\frac{\lambda}{\beta}}(x) \triangleq -\frac{\beta}{2} \cdot \|\mathcal{T}_{\frac{\lambda}{\beta}}(x)\|_2^2 - \lambda \cdot \|\mathcal{T}_{\frac{\lambda}{\beta}}(x)\|_1^2 + \beta \cdot x^T \mathcal{T}_{\frac{\lambda}{\beta}}(x) \tag{5.7}$$

Here $\mathcal{T}_\gamma(x)$ is the entry-wise soft-thresholding operator on vector x that can be formulated for the n_{th} element as:

$$[\mathcal{T}_\gamma(x)]_n \triangleq (|[x]_n| - \gamma)_+ \mathrm{sgn}([x]_n) \tag{5.8}$$

where $(x)_+ = \max(x, 0)$ and $\mathrm{sgn}(x)$ represents the signum function.

We can consider the dual function in (5.6) as the global cost function. Therefore, the individual cost function of each node k can be defined as (please see [Chen

et al., 2015b, Chen et al., 2014] for details):

$$J_k(\boldsymbol{\nu};\mathbf{F}_t) \triangleq \begin{cases} -\frac{\boldsymbol{\nu}^T\mathbf{F}_t}{|\mathcal{N}_\mathcal{I}|} + \frac{1}{N}\|\boldsymbol{\nu}\|_2^2 + \mathcal{S}_{\frac{\lambda}{\beta}}(\frac{\mathbf{D}_k^T\boldsymbol{\nu}}{\beta}), & k \in \mathcal{N}_I \\ \frac{1}{N}\|\boldsymbol{\nu}\|_2^2 + \mathcal{S}_{\frac{\lambda}{\beta}}(\frac{\mathbf{D}_k^T\boldsymbol{\nu}}{\beta}), & k \notin \mathcal{N}_I \end{cases} \tag{5.9}$$

where $|\mathcal{N}_\mathcal{I}|$ is the cardinality of $\mathcal{N}_\mathcal{I}$.

It should be noted that the summation over the individual cost functions $J_k(\boldsymbol{\nu};\mathbf{F}_t)$ is equal to the cost function in (5.6) and the dual problem for estimating the optimal solution $\boldsymbol{\nu}^o$ can be rewritten as:

$$\boldsymbol{\nu}^o = \min_{\boldsymbol{\nu}} \sum_{k=1}^{N} J_k(\boldsymbol{\nu};\mathbf{F}_t) \tag{5.10}$$

Therefore, according to (5.5) the dual problem can be solved using distributed learning strategies and the optimal primal variables $\{\mathbf{D}_k\}$ and $\boldsymbol{z}$ can be recovered afterwards (see (5.14) and (5.15)).

Several distributed learning methods have been proposed in the literature such as incremental strategies [Bertsekas, 1997, Nedic and Bertsekas, 2001], consensus strategies [Xiao and Boyd, 2004, Nedic and Ozdaglar, 2009], and diffusion adaptation strategies [Sayed, 2014, Chen and Sayed, 2012, Monajemi et al., 2014]. It has been shown that diffusion strategies have superior performance and stability compared to the other methods while being robust, scalable, and capable of real time adaptation and learning [Tu and Sayed, 2012]. The details of the diffusion strategy adapted to solve the distributed optimization problem in (5.10) is explained in the next sub-section.

5.2.3 Diffusion Adaptation Method

In the diffusion adaptation strategy, there is a network of N nodes where each node k is connected to its neighboring nodes represented by $\mathcal{N}_k$ shown in Fig. 5.2. Each node can share information with and receive information from its neighbors. Each node also has an individual cost function to minimize and the global cost function of the network is the aggregation of all these individual costs similar to (5.5). The diffusion adaptation method consists of two steps: the adaptation step and the combination step. In the adaptation step, each node k updates its own estimate for the optimization problem via a gradient descent step. This estimate is considered as an intermediate estimate, $\boldsymbol{\psi}_{k,i}$, which is further updated in the combination step. During the combination step, the neighboring nodes share their intermediate estimates. Afterwards, each node k updates its own final estimate, $\boldsymbol{\nu}_{k,i}$, by combining the intermediate estimates received from the neighbors in the ith time instant (further explanation can be found in [Sayed, 2014]). Therefore, the diffusion adaptation strategy can be formulated as:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{\nu}_{k,i-1} - \mu \nabla_{\boldsymbol{\nu}} J_k(\boldsymbol{\nu}_{k,i-1}; \mathbf{F}_t) \qquad \text{(Adaptation step)} \tag{5.11}$$

$$\boldsymbol{\nu}_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k}(i) \boldsymbol{\psi}_{\ell,i} \qquad \text{(Combination step)} \tag{5.12}$$

where $\boldsymbol{\nu}_{k,i}$ is the estimate of node k of the optimal solution $\boldsymbol{\nu}_t^o$ at iteration i, $\boldsymbol{\psi}_{k,i}$ is the intermediate estimate, and $\mu > 0$ is the updating step-size selected to be sufficiently small. The weights $a_{\ell k}(i)$ in (5.12) are called *combination weights* and as seen in Fig. 5.3, each $a_{\ell k}(i)$ is the weight that node k assigns to the information received from node ℓ at time instant i. The combination weights $a_{\ell k}(i)$ must

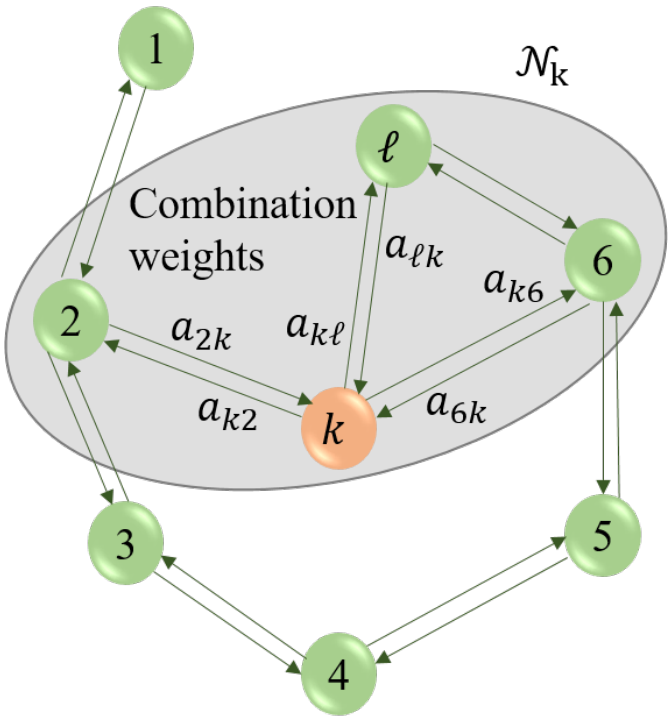


Figure 5.3: An example of a connected network where the neighboring nodes share information with each other. The combination weight $a_{\ell k}(i)$ is the weight that node k assigns to the information received from node ℓ at time instant i.

satisfy:

$$\sum_{\ell \in \mathcal{N}_k} a_{\ell k}(i) = 1, \quad a_{\ell k}(i) > 0 \text{ if } \ell \in \mathcal{N}_k, \quad a_{\ell k}(i) = 0 \text{ if } \ell \notin \mathcal{N}_k \tag{5.13}$$

It should be noted that there are several ways to design the combination weights which can have a significant impact on the performance of the algorithm [Monajemi et al., 2015, Chen et al., 2015a]. In Section 5.2.4 we discuss the role of these weights and introduce an adaptive method to learn the weights over time.

After the optimal dual variable $\boldsymbol{\nu}_t^o$ is estimated by (5.11) and (5.12), the optimal primal variables of the dictionary learning problem, including the sparse

codes $\boldsymbol{z}_t^o$ and the sub-dictionaries $\mathbf{D}_{k,t}$, can be obtained by [Chen et al., 2015b]:

$$\boldsymbol{z}_{k,t}^o = \arg\max_{\boldsymbol{z}_k}[(\mathbf{D}_k^T \boldsymbol{\nu}_t^o)^T \boldsymbol{z}_k - (\lambda \|\boldsymbol{z}_k\|_1 + \frac{\beta}{2}\|\boldsymbol{z}_k\|_2^2)] \tag{5.14}$$

$$\mathbf{D}_{k,t} = \Pi_{\mathcal{D}_k}(\mathbf{D}_{k,t-1} + \mu \cdot \boldsymbol{\nu}_t^o \, \boldsymbol{z}_{k,t}^o) \tag{5.15}$$

where $\Pi_{\mathcal{D}_k}[\cdot]$ is the projection operator onto the constraint set $\mathcal{D}_k$.

In the next section, we propose an adaptive approach to design the combination weights in (5.12). These weights play an important role in combining the information received from the other nodes of the network which can affect the performance of the algorithm.

5.2.4 Selection of the Combination Weights

Selection of the combination weights in (5.12) can affect the performance of the network in solving the optimization problem. Here, we propose to use an adaptive approach for estimating the weights to tackle the dictionary learning task. In the previously proposed distributed dictionary learning methods, the combination weights are determined in a static manner where the nodes allocate the same weights to their neighbors without considering the reliability of the received information [Chen et al., 2015b, Chen et al., 2014, Towfic et al., 2014]. For instance, consider the case where uniform weights are selected and the combination step (5.12) is simply an averaging over all the estimates:

$$a_{\ell k} = \frac{1}{|\mathcal{N}_k|} \quad \text{if } \ell \in \mathcal{N}_k \quad \text{(uniform combination weights)} \tag{5.16}$$

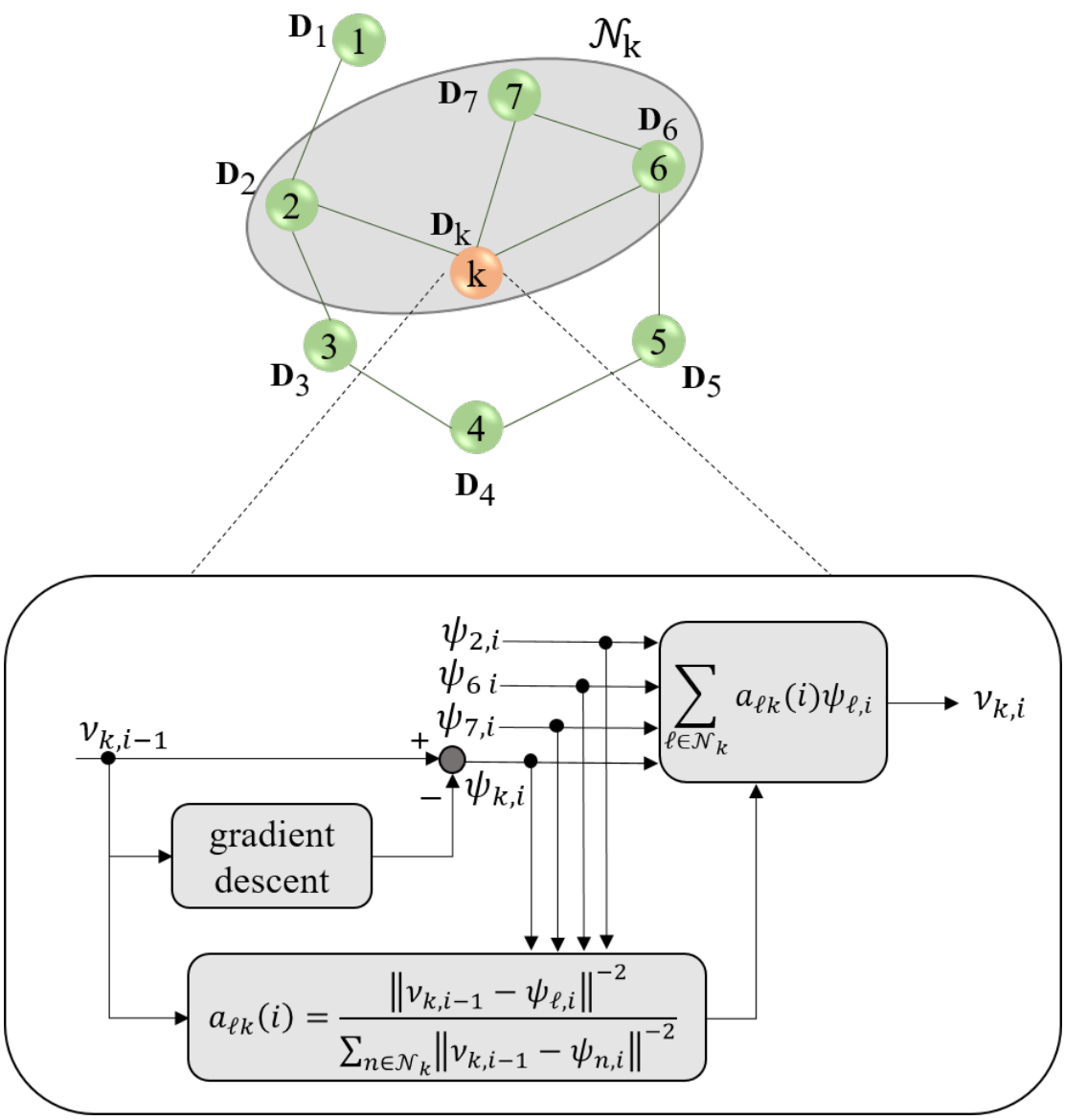


Figure 5.4: The proposed adaptive diffusion method to solve the dual of the dictionary learning problem where each node combines the estimates of its neighbors by the adaptive weights.

By designing the combination weights in a uniform manner the nodes assign the same weight to all of their neighbors without considering the reliability of the information they receive from them.

It has been shown that it is important to design the weights such that the nodes can learn about the reliability of the information received form their neighbors over time [Monajemi et al., 2015, Chen et al., 2015a]. Therefore, the combination weights must be estimated in a manner that helps the nodes to ignore misleading information and cooperate only with neighbors that share the same objective. We do so following the approach proposed in [Zhao and Sayed, 2012] which minimizes the instantaneous Mean Square Deviation (MSD) of the network defined as:

$$\mathrm{MSD}(i) \triangleq \frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\|\widetilde{\boldsymbol{\nu}}_k(i)\|^2, \tag{5.17}$$

where $\widetilde{\boldsymbol{\nu}}_k(i) \triangleq \boldsymbol{\nu}_t^o - \boldsymbol{\nu}_k(i)$ is the error vector at node k at iteration i. Then, the combination coefficients $a_{\ell k}(i)$ can be obtained by solving the optimization problem:

$$\min_{\{a_{\ell k}(i)\}} \mathrm{MSD}(i) = \frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\|\widetilde{\boldsymbol{\nu}}_k(i)\|^2 \tag{5.18}$$

It is shown in [Zhao and Sayed, 2012] that the optimal solution can be approximated by:

$$a_{\ell k}(i) \approx \begin{cases} \frac{\|\boldsymbol{\nu}_k(i-1)-\boldsymbol{\psi}_\ell(i)\|^{-2}}{\sum_{n\in\mathcal{N}_k}\|\boldsymbol{\nu}_k(i-1)-\boldsymbol{\psi}_n(i)\|^{-2}}, & \ell \in \mathcal{N}_k \\ 0, & \text{otherwise} \end{cases} \tag{5.19}$$

One important observation from (5.19) is that the combination weight $a_{\ell k}(i)$ is inversely proportional to the distance between the estimate of node k and the intermediate estimate $\boldsymbol{\psi}_\ell(i)$ of node ℓ. In other words, the combination weights are estimated such that the nodes allocate higher weights to neighbors with similar objectives while learning to ignore misleading information. As a result, using this

Algorithm 2: The proposed Adaptive Distributed Dictionary Learning (ADDL) method for HEp-2 cell classification.

Input: Sub-dictionaries $\mathbf{D}_k$ are initialized randomly and projected onto the constraint set. The dual solution is initialized as $\nu_{k,0} = 0$ for all $k = 1, \ldots, N$.

Set the values for β, λ, and μ.

begin

  foreach *input feature sample* $\mathbf{F}_t$ **do**

    calculate the optimal dual variable $\boldsymbol{\nu}_t^o$ until convergence by $\boldsymbol{\psi}_{k,i} = \boldsymbol{\nu}_{k,i-1} - \mu \nabla_{\boldsymbol{\nu}} J_k(\boldsymbol{\nu}_{k,i-1}; \mathbf{F}_t)$

    $$a_{\ell k}(i) \approx \begin{cases} \frac{\|\boldsymbol{\nu}_k(i-1) - \boldsymbol{\psi}_\ell(i)\|^{-2}}{\sum_{n \in \mathcal{N}_k} \|\boldsymbol{\nu}_k(i-1) - \boldsymbol{\psi}_n(i)\|^{-2}}, & \ell \in \mathcal{N}_k \\ 0, & \text{otherwise} \end{cases}$$

    $\boldsymbol{\nu}_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k}(i) \boldsymbol{\psi}_{\ell,i}$

    foreach *agent* k **do**

      Calculate the sparse codes $\boldsymbol{z}_{k,t}^o$ by $\boldsymbol{z}_{k,t}^o = \arg\max_{\boldsymbol{z}_k} [(\mathbf{D}_k^T \boldsymbol{\nu}_t^o)^T \boldsymbol{z}_k - (\lambda \|\boldsymbol{z}_k\|_1 + \frac{\beta}{2} \|\boldsymbol{z}_k\|_2^2)]$ Obtain the sub-dictionaries $\mathbf{D}_{k,t}$ by:

      $\mathbf{D}_{k,t} = \Pi_{\mathcal{D}_k}(\mathbf{D}_{k,t-1} + \mu \cdot \boldsymbol{\nu}_t^o \, \boldsymbol{z}_{k,t}^o)$

  Obtain the dictionary $\mathbf{D}$ and sparse codes $\boldsymbol{z}$ by:

  $\mathbf{D} = [\mathbf{D}_1 \ldots \mathbf{D}_N]$ $\boldsymbol{z} = \text{col}\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\}$

combination method enables the nodes to continuously learn about the objective of their neighbors so that they can distinguish between the useful and misleading information. Estimating the combination weights in this manner helps the agents to benefit from the cooperation with their neighbors. Moreover, for exploiting the similarity among the nodes with similar objectives, this method results in a more discriminative dictionary which leads to better classification results (Section 5.3). A schematic of the diffusion adaptation method to solve the dictionary learning problem is shown in Fig. 5.4. The summary of the proposed ADDL method is given in Algorithm 1.

5.3 Experiments and Results

5.3.1 Evaluation Methods

Two publicly available datasets including ICPR2012 [Foggia et al., 2013] and ICIP2013 [Foggia et al., 2014] are used for evaluation. Both datasets contain many *cells* in each *specimen* image as discussed in Section 3.3.

ICPR2012. Two evaluation strategies have been performed in the literature for this dataset including "test set" evaluation and "leave-one-specimen-out" (LOSO). The "test set" evaluation uses the provided training and test set, while the LOSO method uses all the cells in one specimen image for test and the rest of the cells for training.

ICIP2013. Due to the lack of a test set, two evaluation methods are used in the literature. The first is the HSM method reported in [Han et al., 2014], where 600 cells (300 for Golgi class) from each class are used for training and the rest for test. The other method is LOSO as performed for the ICPR2012 dataset.

5.3.2 Classification Results

ICPR2012. Table 5.1 shows the classification results for the proposed ADDL and the comparison with other dictionary and non-dictionary based methods. The ADDL results are reported in two forms of adaptive and uniform weights according to (5.19) and (5.16) respectively, where adaptive weights (72%) outperform the uniformly weighted (69%) by 3% on average based on the "Test set" evaluation. The best accuracy in positive images is reported by SNPB method (82%) but for the intermediate level, the ADDL with adaptive weights outperforms other

Table 5.1: The classification accuracies for the ICPR2012 dataset by using Test Set and LOSO evaluation methods.

ICPR2012 (%)			**ADDL**		**Other DLs**			**Others**	
			Adaptive	Uniform	Ensafi⋆	SNPB°	Kastaniotis†	Nosaka◇	DiCataldo‡
Test set	*Cell Level*	Positive	80	78	81	**82**	70	79	60
		Intermediate	**63**	59	62	59	31	58	35
		Average	**72**	69	**72**	70	51	69	48
	Specimen Level		86	79	86	**93**	86	79	**93**
LOSO	*Cell Level*	Positive	94	91	91	92	72	80	**95**
		Intermediate	**83**	78	72	70	55	60	80
		Average	**89**	85	82	81	64	70	88
	Specimen Level		**93**	79	79	86	79	86	**93**

⋆ [Ensafi et al., 2014a] ° [Ensafi et al., 2015] † [Theodorakopoulos et al., 2014]
◇ [Nosaka and Fukui, 2014] ‡ [Di Cataldo et al., 2014]

methods with 63%. The specimen level accuracy is also comparable with other methods at 86%.

With the "LOSO" evaluation method, 89% and 85% accuracies are obtained for adaptive and uniformly weighted ADDL respectively, where the adaptive method outperforms other performances. At the specimen level the accuracy is 93% which is also obtained by [Di Cataldo et al., 2014].

It should be noted that due to the low number of input images in the ICPR2012 dataset (28 images in total) the obtained accuracies are comparable with other methods. However, by increasing the number of input images, as in the case of ICIP2013, the advantage of ADDL method can be seen clearly which is described in the next sub-section.

ICIP2013. Table 5.2 shows the experimental results for the ICIP2013 dataset. The proposed ADDL method with adaptive weights outperforms other methods significantly. By using the HSM evaluation method the ADDL with adaptive weights obtained 93.7% accuracy which is 2% higher than ADDL with uniform weights and other dictionary learning (DL) methods. Additionally, it outperforms non-DL methods by 5%.

With the LOSO evaluation method, ADDL with adaptive weights obtained 81.6% accuracy on average which is 4% higher than that achieved by ADDL with uniform weights. This performance is better than the other DL based methods and 4% higher than other classification methods. Additionally, the performance in the specimen level is better than other methods offering 90.4% accuracy. The superior classification accuracy can be largely explained by the use of combination wights in the diffusion adaptation method that enables the nodes to share information and solve the optimization problem in a cooperative manner. Note that accuracies on the ICPR2012 and ICIP2013 datasets are different because the quality and amount of images within the two datasets are very different.

5.3.3 Computational cost

Dictionary learning is a computationally expensive and time consuming task. Table 5.3 shows different dictionary learning procedures with their computation times. These measurements were done on a machine with Intel Core i7 CPU 2.6GHz and 16 GB RAM with 64-bit operating system. As shown in Table 5.3, the proposed ADDL method, by giving the information to a single node, takes less time (56.64 sec) than the other methods. This is 2 and 5 times better than

Table 5.2: The classification accuracies for the ICIP2013 dataset by using HSM and LOSO evaluation methods.

ICIP2013 (%)			ADDL		Other DL Methods				Others	
			Adaptive	Uniform	Ensafi⋆	SNPB°	Gragnaniello†	manivannan◇	HSM ‡	Larsen ▽
HSM	*Cell Level*	Positive	**97.9**	95.4	95.8	96.8	-	-	95.5	-
		Intermediate	**89.4**	87.6	87.9	88.8	-	-	80.9	-
		Average	**93.7**	91.5	91.9	92.8	-	-	88.2	-
LOSO	*Cell Level*	Positive	**88.5**	84.2	83.4	83.8	-	-	-	-
		Intermediate	**74.7**	71.4	71.2	72	-	-	-	-
		Average	**81.6**	77.8	77.3	77.9	81.1	80.3	-	78.7
	Specimen Level		**90.4**	86.7	88	89.2	86.7	89.9	-	-

⋆ [Ensafi et al., 2014b] ° [Ensafi et al., 2015] † [Gragnaniello et al., 2014]
◇ [Manivannan et al., 2014a] ‡ [Han et al., 2014] ▽ [Larsen et al., 2014]

the methods of [Ensafi et al., 2014b] and SNPB [Ensafi et al., 2015], respectively. The proposed method takes 15 seconds more when the information is given to all nodes to process.

For the ICIP2013 dataset, the proposed method takes 286.21 seconds to calculate the dictionary when the information is given to a single node to process. This is 47 seconds lower than passing information to all nodes to process but significantly better than the results of other dictionary methods. For example, ADDL is 9 and 20 times better than the [Ensafi et al., 2014b] and SNPB [Ensafi et al., 2015], respectively. Therefore, it can be seen clearly that the proposed method can enhance the performance of dictionary learning task significantly in both computational cost and classification accuracy.

Table 5.3: Computation times of different dictionary learning methods.

Dictionary Learning	**ADDL**		Ensafi⋆	SNPB°
Computation time (sec)	All Nodes	Single Node		
ICPR2012	71.63	56.64	126.34	354.38
ICIP2013	333.73	286.21	2751.91	5742.64

⋆ [Ensafi et al., 2014b] ° [Ensafi et al., 2015]

5.4 Summary

In this chapter, we proposed an adaptive distributed dictionary learning method that benefits from lower computational cost with lower number of tuning parameters which is an important advantage in solving classification problems. The ADDL method is applied to HEp-2 cell images and obtained state-of-the-art results for both public datasets. The proposed method enhances the accuracy of the cell classification problem compared to other methods while reducing the computational time significantly as the dictionary is learned in a distributed manner. Moreover, learning the combination weights adaptively is an important contribution for the proposed method which makes it capable of adjusting itself for different datasets according to the input images. This approach is also a foundation for big data analysis where the information is available on the nodes of a computer cluster or cloud.

Chapter 6

Superpixel Approach versus Patch Base Classification

6.1 Introduction

Although some success has been achieved for HEp-2 cell classification as reported in the recent benchmarking tests at ICPR2012 [Foggia et al., 2013] and ICIP2013 [Foggia et al., 2014], the best accuracy obtained is generally lower than what physicians can provide. The major constraints of these image patch based techniques include the high computational cost due to a huge amount of overlapped image patches [Wang et al., 2016] to be processed and the tedious parameter tuning (for patch size, scanning step size, etc.) for optimal cell classification performance.

In this chapter we propose a novel superpixel based HEp-2 cell classification technique by using sparse coding scheme. We call our proposed method the Sparse

Coding of Superpixels (SCS).

In the superpixel approach which is widely used in segmentation problems [Lucchi et al., 2010], the input image is divided to relatively small and non-overlapped regions. Each superpixel contains many connected pixels which have similar features. In segmentation problems, the features of each superpixels are used to decide whether the superpixel should aggregate with the neighboring superpixels to make a bigger region and the process continues until the final segmentation result is obtained. However, for classifying images, not only the features of pixels within the superpixels but also superpixels boundaries which are usually aligned with the high gradient regions should be analyzed. Sparse Coding technique is widely used in different machine learning problems [Khorsandi et al., 2015b, Taalimi et al., 2015b, Wang et al., 2013, Wang et al., 2014, Zonoobi and Kassim, 2014], where image patches are used to calculate the features to train the dictionary and the classifiers.

We designed a novel superpixel technology for the HEp-2 cell classification problem. To the best of our knowledge, this is also the first work that uses superpixel to classify the HEp-2 cells (see Section 6.2.1). Our major contributions can be summarized as follows:

- Superpixels are used (instead of regular sampling of overlapped image patches) to guide the selection of the right image patches that contain more 'informative' features.

- "Extended superpixels" are designed by dilating the boundary of each superpixel which capture more discriminative gradient information across the boundaries of the HEp-2 cell.

- A cell extraction method is designed which extracts better-quality cells than the originally provided ones.

The rest of this chapter is organized as follows. The proposed SCS method is described in Section 6.2, including the cell extraction, superpixel and dictionary learning scheme. In Section 6.3 the experiments on two publicly available datasets are investigated and validation on the parameters of the proposed superpixel method is discussed. Finally, we provide our conclusions in Section 6.4.

6.2 Sparse Coding of Superpixels (SCS)

Our proposed SCS technique comprises the following four stages as illustrated in Fig. 6.1: *cell extraction*, *superpixel extraction*, *dictionary learning* and *cell classification*. The cell extraction stage is specially needed when the bounding boxes of the cells are not provided (as in specimen classification of ICIP2013). Given a cell image, the superpixels are first extracted to determine informative image patches. SIFT and SURF features are then extracted from each superpixel to learn an over-complete dictionary. Finally, a linear SVM classifier is trained for HEp-2 cell classification.

Cell Extraction

The cell extraction method (see Section 3.2.2) is used where morphological features (area, solidity, etc) are exploited to select the cells which are of the same shape (roundness) and size. This method first uses the histogram of the area of the connected components to estimate the range of typical size of the cells and then

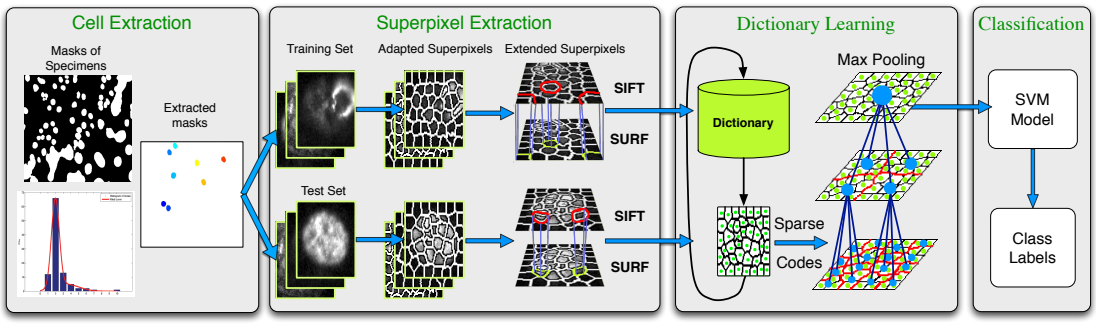


Figure 6.1: Framework of the proposed HEp-2 cell classification technique.

selects cells in the specimen image according to their roundness property. This method helps reduce the number of required cells for specimen classification and improves performance by removing noisy segments.

6.2.1 Superpixel Extraction

We extract the superpixels based on the Simple Linear Iterative Clustering (SLIC) [Achanta et al., 2012] due to its distinct properties of low computational cost and close adherence to the object boundaries in comparison with similar methods [Levinshtein et al., 2009, Felzenszwalb and Huttenlocher, 2004]. The original SLIC method, places grid points (P points) to be the initial superpixel centers. If an initial superpixel center lies along the cell boundary, another cell image pixel with the minimum gradient value lying within the 3×3 neighborhood of the original center is selected as the initial superpixel center. k-means clustering is then performed to calculate the distance between the cluster centers and the neighboring pixels according to their intensity (color or gray scale) values and positions. The distance calculation is accomplished in the $2W \times 2W$ window size (W denotes the superpixel size which can be determined by $\sqrt{s}$, where s is the number of

desired pixels in each superpixel) to find similar neighboring pixels instead of the whole image area in order to increase the speed of the algorithm.

This method tries to minimize the distance of the color and pixel positions with the cluster centers $[l, a, b, x, y]^T$, where the first three elements are the CIELAB color space parameters and the last two are the position of pixels. These different distances are normalized in order to be aggregated together in one formulation as in equation 6.1.

$$\begin{aligned} d_c &= \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \\ d_s &= \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \\ D &= \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \end{aligned} \tag{6.1}$$

where N_c and N_s are color and spatial proximities by their maximum distances within a cluster [Achanta et al., 2012].

Our modifications to the SLIC superpixel algorithm are as follows.

i. besides the color and spatial proximities, we added the gradient information to also enable the algorithm to evaluate the texture information resulting in the following extended set of parameters: $[l, a, b, x, y, g]^T$; where g is the magnitude of gradient in each pixel. The gradient distance (d_g) is normalized with N_g which is the maximum gradient distance between the pixels and the cluster center in one cluster as formulated in equation 6.2

$$D = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2 + \left(\frac{d_g}{N_g}\right)^2} \tag{6.2}$$

where d_g is

$$d_g = |g_j - g_i| \tag{6.3}$$

ii. while the SLIC uses the number of the desired superpixels P to control the superpixel size which does not work well as different images have different sizes and accordingly very different superpixel sizes s (the number of pixel within a superpixel), we use the superpixel size s as the input parameter to ensure that superpixels from images of different sizes will have a similar superpixel size. Note that the superpixel size s can be translated to the number of superpixels P by $P = [N/s]$, where N denotes the number of pixels within the image.

iii. the SLIC method adheres to the boundaries (i.e., high gradient regions) which is desirable for object segmentation but for HEp-2 cell classification, it could cause serious problem as many discriminative features lie across the high gradient regions. We solve this problem by introducing an "extended superpixel" which is derived by dilating the original superpixel to include the "cross-boundary" information. We extend each superpixel to include the high gradient information to be discussed in Section 6.3.3.

The proposed superpixel extraction method is applied on the cells, before convolving with masks. Then those superpixels which are outside of the masks, are omitted. The extended superpixels can then be used for dictionary learning and cell classification.

6.2.2 Dictionary Learning

SIFT and SURF are used as visual features for dictionary and classification model learning. In particular, SIFT and SURF features are first extracted from each extracted superpixel patch and then processed through the max-pooling of the extracted feature histograms. The processed features are then concatenated to form the feature description. The features of the superpixels are finally sampled from each input image and creates the $\mathcal{D}$-dimensional feature matrix $\mathbf{F}= [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N]^\intercal \in \mathbb{R}^{(N\times\mathcal{D})}$ as illustrated in Fig. 6.1. The dictionary learning and classification method is explained in Sections 3.2.4 and 3.2.5, respectively.

6.3 Experiments and Results

6.3.1 Evaluation Metric and Protocol

The proposed technique is evaluated by using the MCA (see Section 3.4) as suggested by the contest organizers. For the ICPR2012 dataset, the Leave-One-Specimen-Out (LOSO) strategy is adopted for the fair comparison with the state-of-the-art results which are reported by the dataset organizers [Foggia et al., 2014]. In LOSO strategy all the training and test images are used. Separately, the accuracies on 'test set' is also evaluated by training the classification model using the training image set.

In the ICIP2013 dataset, because we only have access to the training set, two protocols which are used in the literature are exploited including Leave-One-Specimen-Out strategy as used in [Manivannan et al., 2014a, Larsen et al., 2014]

and the evaluation protocol used in [Han et al., 2014], which we named HSM representing the title of the paper (High-order Statistics of Microtexton for HEp-2 Staining Pattern Classification) for cell classification problem. All the training cells which are provided by the ICIP2013 dataset are extracted from 83 specimen images. In the first method, Leave-One-Specimen-Out strategy, in each run, the cells of one specimen image is used for testing and the rest cells of 82 specimen images are used for training. In the second evaluation method, HSM [Han et al., 2014], 600 cells are randomly selected from each of the classes (except 300 cell images for Golgi class) for training and the rest for testing. This task is performed 20 times and the average accuracies of all iterations are reported. Particularly, in each iteration, $5 \times 600 + 300 = 3300$ cells are selected for training and $13596 - 3300 = 10296$ cells for testing (13596 is the total number of training cells in the ICIP2013 dataset).

To obtain the image level classification result, a modification of the masks should be applied to get the correct cell information from the images as described in sect. 6.2. By performing the proposed cell extraction method on 252 specimens (1008 images), 5012 cells are extracted. The Leave-One-Specimen-Out strategy is also used where all the cells from one specimen image are used for testing and the rest cells of 251 specimen images are used for training. To obtain the class label of each specimen image, maximum voting is performed on the cell labels of that specimen image.

In our experiments we extract superpixels of size 100 pixels and extend them for 9 and 12 pixels from each side for ICPR2012 and ICIP2013 datasets respectively, which results in better accuracies as shown in Fig. 6.2. It should be noted,

because the ICIP2013 dataset contains gray scale images, the intensity values of the pixels are used instead of CIELAB color space values to calculate color distance (d_c) in (6.1).

6.3.2 Classification Results

ICPR2012 - Table. 6.1 shows experimental results on the ICPR2012 dataset. As Table. 6.1 shows, our method obtains the best accuracy among other methods as reported in [Foggia et al., 2014]. For the *Cell Level* classification, an accuracy of 79% is obtained when the model is trained by using all training images including positive and intermediate images and then evaluated on the test set as shown in second column of Table. 6.1. For the Leave-One-Specimen-Out evaluation strategy, an interesting result is obtained for the intermediate level images (forth column) where 92% accuracy is obtained which is 12% higher than other methods. On average, we got an accuracy of 93% which is 5% higher than the-state-of-art based as shown in fifth column of Table. 6.1.

The *intermediate* cell classification plays an important role in HEp-2 cell classification problem. As the intensity values of *intermediate* cells are much lower than *positive* cells, they may affect the final evaluation if they are considered together when training the classifier. In fact, those methods which have separate models for these two categories, usually perform better than those which use both categories simultaneously. Although there is an increased number of training data in the Leave-One-Specimen-Out strategy, improved accuracies were not observed because the intensity levels in the test set are ignored.

We also measured the effect of our modified superpixel extraction method

Table 6.1: Accuracy on all training and test images of the ICPR2012 dataset based on Leave-One-Specimen-Out and accuracy when the model is trained by the training set and applied to the test set [Foggia et al., 2013].

ICPR2012 Dataset Accuracy (%)	**Cell Level**				**Image Level**
	Test set	*Leave-One-Specimen-Out*			*Leave-One-Specimen-Out*
		Positive	Intermediate	Average	
SCS	**79**	**95**	**92**	**93**	**93**
SCS (no Gradient)	76	90	86	88	85
SBoW	78	82	65	74	93
Kastaniotis	75	72	55	64	78
Shen	74	71	65	68	78
DiCataldo	72	**95**	80	88	**93**
Kazanov	71	63	49	51	75
Faraki	70	81	59	70	78
Nosaka	69	80	60	70	85
Wiliem	67	63	49	56	71
Xiangfei	67	89	74	82	89
Stoklasa	64	80	44	62	71

which incorporates the gradient information (see Section 6.2.1). TABLE 6.1 shows the results when our method without gradient information named 'SCS (no Gradient)' is applied on the dataset. Despite the high accuracies for some scenarios, the average accuracy is lower than that of using the gradient information in our method.

From the confusion matrices of *Image Level*, *Intermediate* and *Positive* images in ICPR2012 obtained using Leave-One-Specimen-Out method shown in Table. 6.2, it is evident that the two Fine- and Coarse-speckled classes are hard to classify because of the similar patterns that they have as can be seen in the heat-map show in Fig. 3.4.

ICIP2013 - Table. 6.3 and 6.4 show experimental results on the ICIP2013 dataset. For *Cell Level* classification task, the HSM [Han et al., 2014] and Leave-

Table 6.2: The confusion matrices for Positive (a) and Intermediate (b) images and image level (c) by using Leave-One-Specimen-Out method.

	Ce	CS	Cy	FS	H	N
Ce	100.0	0.0	0.0	0.0	0.0	0.0
CS	0.0	66.7	0.0	33.3	0.0	0.0
Cy	0.0	0.0	100.0	0.0	0.0	0.0
FS	0.0	0.0	0.0	100.0	0.0	0.0
H	0.0	0.0	0.0	0.0	100.0	0.0
N	0.0	0.0	0.0	0.0	0.0	100.0

(a) Positive Cells

	Ce	CS	Cy	FS	H	N
Ce	100.0	0.0	0.0	0.0	0.0	0.0
CS	0.0	100.0	0.0	0.0	0.0	0.0
Cy	0.0	0.0	100.0	0.0	0.0	0.0
FS	0.0	50.0	0.0	50.0	0.0	0.0
H	0.0	0.0	0.0	0.0	100.0	0.0
N	0.0	0.0	0.0	0.0	0.0	100.0

(b) Intermediate Cells

	Ce	CS	Cy	FS	H	N
Ce	100.0	0.0	0.0	0.0	0.0	0.0
CS	0.0	80.0	0.0	20.0	0.0	0.0
Cy	0.0	0.0	100.0	0.0	0.0	0.0
FS	0.0	25.0	0.0	75.0	0.0	0.0
H	0.0	0.0	0.0	0.0	100.0	0.0
N	0.0	0.0	0.0	0.0	0.0	100.0

(c) Image Level

One-Specimen-Out method (see Section 6.3.1) are used. In the first method, as shown in Table. 6.3, where all the reported methods have used HSM [Han et al., 2014] evaluation method, we obtained 97.73% and 90.54% accuracies for positive and intermediate level images which are higher than the state-of-the-art results. In the Leave-One-Specimen-Out method, higher accuracies were achieved in comparison with other state-of-the-arts as shown in Table. 6.4 in *Cell Level* columns. We have also studied the effect of using gradient information to our superpixel extraction method. As can be seen in the Table. 6.3, the accuracy is lower without gradient information (SCS(no Gradient)) than when this information is used. Although the achieved average accuracy of 91.1% is higher than by HSM [Han et al., 2014] method, it is 3% lower than the SCS proposed method.

For the *Image Level* classification task, when we compare our proposed SCS

Table 6.3: Accuracy on ICIP2013 dataset for Cell Level by using HSM [Han et al., 2014] evaluation method.

ICIP2013 (%)	**Cell Level**				
	HSM	[Ensafi et al., 2014b]	[Ensafi et al., 2015]	SCS (no Gradient)	**SCS**
Positive	95.5	95.8	96.8	95.4	**97.73**
Intermediate	80.9	87.9	88.8	86.8	**90.54**
Average	88.2	91.9	92.8	91.1	**94.14**

Table 6.4: Accuracy on ICIP2013 dataset for Cell and Image Level by using Leave-One-Specimen-Out evaluation method.

ICIP2013 (%)	**Cell Level**			**Image Level**			
Leave-One-Specimen-Out	Larsen*	manivannan†	**SCS**	Gragnaniello$^\diamond$	manivannan‡	All Cells	**SCS**
Positive	-	-	**86.79**	-	-	87.36	**92.38**
Intermediate	-	-	**76.87**	-	-	74.16	**88.74**
Average	78.70	80.25	**81.83**	86.87	89.93	80.76	**90.56**

* [Larsen et al., 2014] † [Manivannan et al., 2014a] $^\diamond$ [Gragnaniello et al., 2014] ‡ [Manivannan et al., 2014b]

results with two other state-of-the-art methods including [Gragnaniello et al., 2014] and [Manivannan et al., 2014b] by using Leave-One-Specimen-Out evaluation strategy, we achieved higher accuracies for the Image Level part as seen in Table. 6.4. We also evaluated our proposed cell extraction method by comparing the results of proposed SCS 'with' and 'without' cell extraction stage; the latter is called 'All Cells' strategy. In the 'All Cells' strategy, all connected pixels of the specimen images are extracted by using the provided masks without filtering based on *area* and *solidity* as explained in Section 6.2. This gives around 110,000 cells in total. Then the proposed SCS method (without cell extraction part) is performed. As can be seen from the last two columns of *Image Level* part in Table. 6.4, the cell extraction method achieves significantly better results. Specifically, we obtained an average accuracy of 90.56%, which is 10% higher than that of using all cells of the specimen images. Additionally, the time complexity of the

Table 6.5: The confusion matrices for Cell Level by using HSM [Han et al., 2014] evaluation method (a), Leave-One-Specimen-Out method (b) and Image Level by using Leave-One-Specimen-Out method (c).

	Ce	G	H	N	NuMem	S
Ce	95.11	0.00	0.42	0.00	2.56	1.92
G	0.00	96.67	2.47	0.00	0.00	0.87
H	0.00	2.44	94.44	1.64	0.00	1.49
N	0.00	2.27	2.39	95.08	0.00	0.26
NuMem	2.63	1.36	0.00	0.00	94.24	1.78
S	1.58	0.20	7.55	0.89	0.51	89.29

(a) Cell Level HSM method

	Ce	G	H	N	NuMem	S
Ce	87.88	0.00	1.25	0.00	6.33	4.54
G	4.78	60.33	15.67	6.91	5.22	7.09
H	1.67	3.15	85.63	2.92	2.30	4.33
N	0.67	0.00	3.89	90.71	1.33	3.40
NuMem	6.33	1.04	0.98	0.90	89.01	1.74
S	6.55	1.33	7.12	6.33	1.25	77.42

(b) Cell Level Leave-One-Specimen-Out

	Ce	G	H	MitSp	N	NuMem	S
Ce	94.65	0.00	0.00	0.00	0.00	0.00	5.35
G	0.00	93.24	0.00	6.76	0.00	0.00	0.00
H	0.00	0.00	94.80	0.00	0.00	0.00	5.20
MitSp	0.00	5.47	5.13	86.23	0.00	3.17	0.00
N	0.00	0.00	0.00	0.00	100.00	0.00	0.00
NuMem	0.00	0.00	15.86	0.00	0.00	84.14	0.00
S	2.00	0.00	2.00	0.00	0.00	0.00	96.00

(c) Cell Level Leave-One-Specimen-Out

proposed method on extracted cells is an order of magnitude lower than using all the cells due to the lower amount of training data.

The confusion matrices for *Cell* and *Image Level*s by using HSM [Han et al., 2014] and Leave-One-Specimen-Out methods are shown in Table. 6.5. For the *Cell Level* classification, the Homogenous and Speckled classes are misclassified more than the others due to their similar patterns as evident from Table 6.5a and 6.5b. For the *Image Level* classification, where one new class (MitSp) is added to the dataset, the confusion matrix in Table. 6.5c shows that the misclassification rate between MitSp, NuMem and Speckled are high. To achieve better results, more informative features are needed which is an interesting topic for further research.

The superior cell classification accuracy can be largely explained by the use of informative patches of cell images that are obtained by proposed superpixel

method. Note that accuracies on the ICPR2012 and ICIP2013 datasets are very different because the quality and amount of images within the two datasets are very different.

6.3.3 Superpixel Parameters

In this sub-section, we present our study on the robustness of the proposed method to changes in superpixel size and extension using the test set of the ICPR2012 dataset and the evaluation set of ICIP2013 (see Section 6.3.1). As evident from Fig. 6.2, the accuracy increases from very small superpixel sizes to a value which performs the best (100 pixels in one superpixel). In contrast, the accuracy decreases when larger superpixels are extracted. In other words, when the superpixel size is very small, it contains less information for classification. Therefore, the accuracy is not acceptable for very small size superpixels. On the other hand, the large superpixels contains many informative features but categorizing them by the dictionary learning process may increase the reconstruction error. Additionally, when the large superpixels are used, the number of superpixels will decrease.

The correlation of training and test accuracies are shown in Fig. 6.2a and 6.2b, when the superpixel size is increased. For example in ICIP2013 dataset, where we have enough representing data, the training accuracy is also drops by increasing the superpixel size. This study shows that the proposed technique prefers a larger number of small superpixels instead of a smaller number of larger superpixels.

In addition, applying the extension to the superpixel improves the cell classification accuracy clearly which can be observed when the superpixel boundaries extend from 0 pixel to 3 pixels (see Fig. 6.2c and 6.2d). It shows that using the

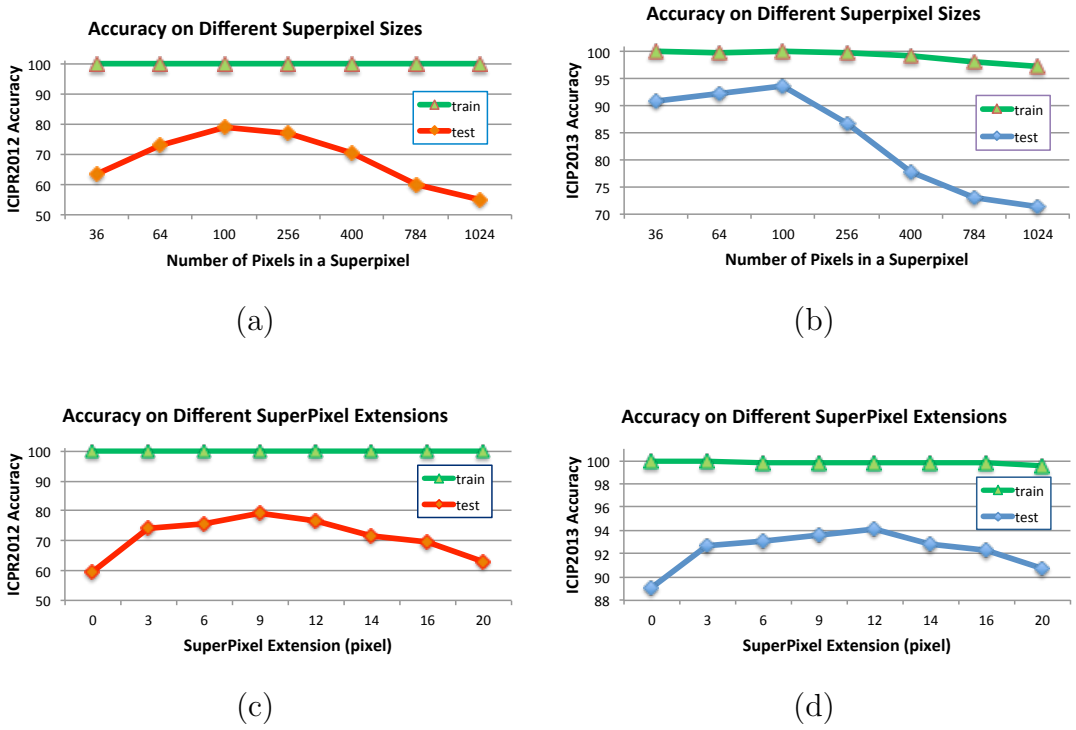


Figure 6.2: Cell classification accuracy for ICPR2012 (left graphs) on the 'test set' and ICIP2013 (right graphs) by using HSM [Han et al., 2014] evaluation protocol, where superpixel sizes (a, b) and different superpixel extensions (c, d) are applied.

original superpixels (with no extension) results in low accuracy as the informative features of the superpixels are the edges of the images which now overlap with the boundaries of superpixels and are thus omitted in the feature extraction process. Therefore, by extending the superpixel sizes, we bring these important information into the superpixels and provide better image patches for classification purposes. At the same time, the accuracy stabilizes when the extension lies around 6-14 pixels.

Table 6.6: Efficiency improvement for both model training and cell classification.

Timing	Dictionary Learning Time (sec)			Testing Time (sec)		
	Superpixels	Overlapping Patches	Ratio	Superpixels	Overlapping Patches	Ratio
ICIP2013	10725	137595	12.83	1505	3048	2.03
ICPR2012	924	6317	6.84	145	324	2.23

6.3.4 Timing

The proposed superpixel based technique is much faster than the traditional overlapped patched based methods for both training and testing tasks. As Table 6.6 shows, the dictionary learning is around 12 and 6 times faster for ICIP2013 and ICPR2012 dataset, respectively. In addition, the testing time for feature sparse coding and SVM classification is around 2 time fast. These measurements are accomplished in a machine with Intel Core i7 CPU 2.6GHz and 16 GB of RAM with 64-bit operating system.

6.4 Summary

This chapter presents a superpixel based HEp-2 cell classification technique. Unlike traditional image patch based approaches, the proposed technique makes use of superpixels to select image patches in a more intelligent way. In addition, several adaptations of the superpixel such as the minimizing the gradient distance and extension idea have been carefully introduced for the optimal cell classification. Extensive experiments on two public datasets show superior HEp-2 cell classification performance.

This method is based on the unsupervised dictionary learning which does not necessarily result in discriminative sparse code. Additionally, the different

features in each superpixel are concatenated which can increase the redundancy. To facilitate these problem a supervised dictionary learning which satisfies the feature fusion, is proposed and studied in the following chapter.

Chapter 7

Joint Multi-Cue Dictionary Learning

7.1 Introduction

The majority of existing dictionary learning methods, supervised or unsupervised, can handle only single source of data [Jiang et al., 2013a,Ramirez et al., 2010,Yang et al., 2014a, Ensafi et al., 2014b]. Fusion of information from different sensor modalities can be more robust to single sensor failure. For example, in [Shekhar et al., 2014] the classification results of using face, fingerprint, and hand signatures are fused using a majority vote to achieve better performance in identity verification. The information fusion happens in feature level or classifier level [Ruta and Gabrys, 2000]. In feature fusion, different types of features are combined to make one representation while in classifier fusion, for each modality one classifier is trained independent of others and their decisions would be fused. While clas-

sifier fusion has been well-studied in research papers, fusion at feature level is a relatively less-studied problem, specifically for fusing different modalities [Rattani et al., 2007] mostly because feature modalities are not in the same size. In Bag-of-Words, feature fusion is imposed by concatenating all of the features in one vector. The dimension of this vector is high and suffers from curse-of-dimensionality while it does not contain the valuable information of correlation between feature types. In this chapter, we use "modality", "source", "cue" and "feature" interchangeably.

Joint sparsity with fix dictionary Joint sparsity prior can do fusion between multiple features which makes them suitable to reconstruct samples originated from different sources [Shekhar et al., 2014,Bahrampour et al., 2014,Minaee et al., 2015]. The dictionary in [Shekhar et al., 2014], is predefined and is made by putting together all training samples. In other words, their method is designed to answer the following question: Given a multimodal signal and a set of modality-specific dictionary, how we can find the multimodal sparse codes while joint sparsity is applied. That is, the only optimization variable is the multimodal sparse codes and they neglect the significant role of designing dictionary for the task. We believe that, to get a proper representation of the signals, the design of the dictionary has a significant role. Hence, we design our method to learn a dictionary for each modality that is "good"at reconstructing the pure signal and "bad" at modeling the noise, while at the same time the sparse decomposition coefficients of all modalities have the similar sparsity pattern.

In [Taalimi et al., 2015a] multi-modal dictionary learning and classifier training is proposed. The proposed joint optimization problem is formulated as task-driven dictionary learning [Mairal et al., 2012] and solved using LC-KSVD [Jiang

et al., 2013b]. However, the basic idea of task driven dictionary learning is to learn optimal parameters (dictionary) to make the coefficients fit the task. So it needs to solve a bi-level optimization problem to make the coefficients consistent with the training and testing phase. While LC-KSVD does not utilize the bi-level strategy, the algorithm in LC-KSVD can not guarantee the optimal solution, in which modality-based sparse representations are the solution of another optimization problem.

While calculating the sparse codes of each image patch provides the local information stored in the patches, the spatial information is also essential for classification and this is obtained by aggregating the local information. A naive approach is to concatenate the features of all patches in each image to obtain a long vector of sparse codes. However, the final feature vector size for each image would be different due to the various number of patches for each image according to the image size. To this end, we introduced a novel pooling strategy to combine the patches' sparse codes that benefit from two important properties of small size feature vector and wisely selected image regions where their patches should be aggregated. This is performed by dividing the image into three layers as in Spatial Pyramid Matching (SPM) [Lazebnik et al., 2006] (see Section 7.3.3) including whole image, a tube around the cell boundary and the inner side of the tube. The last two layers are then divided to 4 regions and the max-pooling operator is performed to combine the information of the image patches.

Our main contributions to solve the above limitations are the following proposals:

- A new multi-modal dictionary learning method that produces discrimina-

tive dictionaries with few atoms from many training samples, where one dictionary is trained in all-vs-all fashion for each modality.

- Our target is to show that in the presence of multi-modal data where each sample is seen from highly related feature modalities with various sizes (here, SIFT and SURF), we can get better classification accuracy by encoding the a priori known correlation between feature modalities in space of sparse codes. The correlation (or relation) between different features/modalities is translated in space of sparse codes as the similarity between zero/nonzero pattern of the channels. This is done using the notion of grouping in space of sparse codes and applied with the joint sparse regularization to enforce the multi-modal sparse representations of each class to share the same sparsity patterns at the column level.

- A novel pooling strategy that combines the information of the image patches in certain regions. The regions are adaptively produced according to the boundary of the cell in the image.

The optimization problem over multi-modal dictionaries and multi-modal sparse representations is solved jointly. This method can combine information from different feature types and force them to have common sparsity patterns for each class, which is presented in Fig. 7.1.

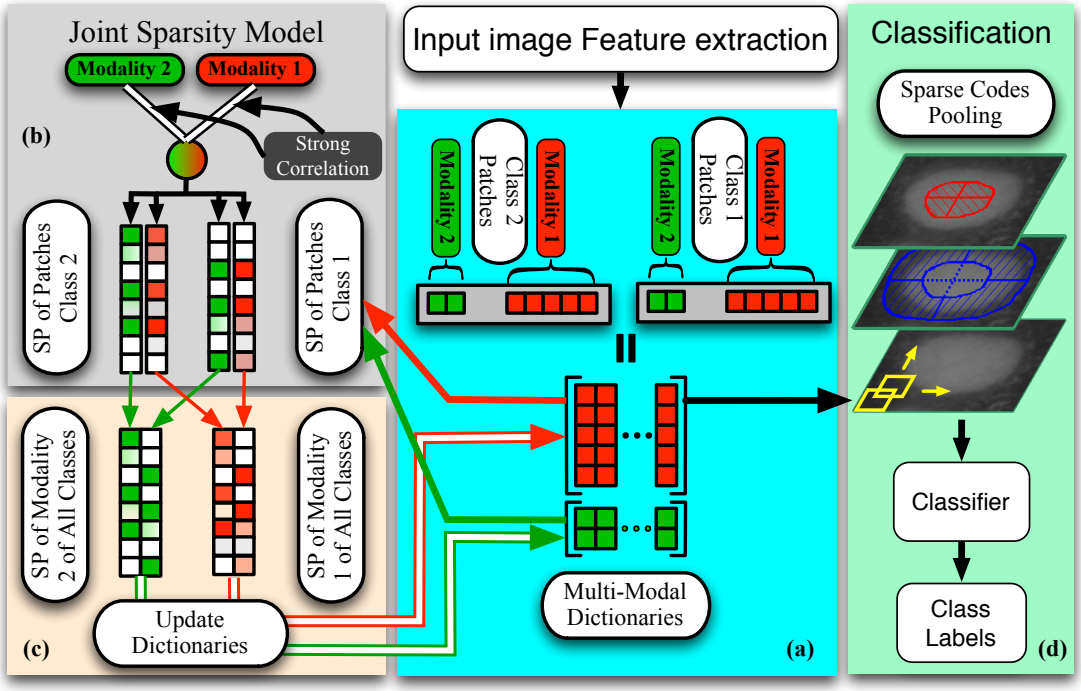


Figure 7.1: Joint sparse modeling for two classes with two modalities and color coded dictionaries (a). The joint sparsity regularizer (b) imposes high correlation between the sparse representations. Modality-based sparse codes of all classes (c) updates dictionaries followed by sparse codes pooling and training SVM classifier(d).

7.2 Sparse representation classification

7.2.1 Fixed dictionary

The SRC method is proposed in [Wright et al., 2009] for the face recognition problem. Let C represent the number of classes, training data as $\{\boldsymbol{Y}_{i,c}\}_{i=1}^{N}, c \in \{1, \cdots, C\}$ and $N = \sum_{c=1}^{C} N_c$ training samples where each class c has N_c samples as $\boldsymbol{Y}_{i,c}|_{i=1}^{N_c}$. The class specific dictionary $\boldsymbol{D}_c$ is made by concatenating all training samples as $\boldsymbol{D}_c = [\boldsymbol{Y}_{1,c}, \ldots, \boldsymbol{Y}_{N_c,c}] \in \mathbf{R}^{n \times N_c}$ where n is the dimension of the feature modality. The final dictionary is made by putting together all class-specific

dictionaries as $\boldsymbol{D} = [\boldsymbol{D}_1, \ldots, \boldsymbol{D}_C] \in \mathcal{R}^{n \times N}$. The task is to identify the label of test sample $\boldsymbol{y}_t \in \mathcal{R}^n$. In SRC it is assumed that the test sample $\boldsymbol{y}_t$ lies in the space formed by $\boldsymbol{D}_c$ and can be approximated using few number of training samples of the c-th class:

$$\boldsymbol{y}_t = \boldsymbol{D}\boldsymbol{\alpha}_t + \boldsymbol{e} \tag{7.1}$$

where $\boldsymbol{\alpha}_t$ is the $N \times 1$ sparse representation of test signal using $\boldsymbol{D}$. Assume $\delta_c \in \mathcal{R}^N$ as an operator that is applied on $\boldsymbol{\alpha}$ and it only keeps coefficients that are corresponding to atoms of the c-th class and makes the rest coefficients zero. To obtain the class label of $\boldsymbol{y}_t$ the above equation should be solved such that $\boldsymbol{\alpha}_t$ is sparse:

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_{\ell_1} \quad \text{s.t.} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha}\|_{\ell_2} \leq \epsilon \tag{7.2}$$

The test data is reconstructed using atoms that belong to the c-th class: $\hat{\boldsymbol{y}}_c = \boldsymbol{D}\delta_c(\boldsymbol{\alpha}_t)$ and it belongs to the class with minimum reconstruction error: $\min_c \|\boldsymbol{y} - \hat{\boldsymbol{y}}_c\|_{\ell_2}$.

The dictionary in SRC scheme is made by concatenation of all training samples hence the atoms are not designed carefully for the desired task. This issue limits the usage of the SRC method to cases with a huge amount of training samples. The number of training samples should be high enough so that we can be sure that the probability distribution of the data is sampled enough.

7.2.2 Joint sparse representation classification

So far the classification was based on a single source of feature. In this section, we go through the classification using multiple feature types. The idea is to

exploit all the available sources of information and their correlation for the desired classification task.

$$\underset{\boldsymbol{\Gamma}=[\boldsymbol{\alpha}_t^1,\boldsymbol{\alpha}_t^2,\ldots,\boldsymbol{\alpha}_t^M]}{\operatorname{argmin}} f(\boldsymbol{\Gamma}) + \lambda\Omega(\boldsymbol{\Gamma}) \tag{7.3}$$

where $f(\boldsymbol{\Gamma}) \triangleq \frac{1}{2}\sum_{m=1}^{M} \|\boldsymbol{y}_t^m - \boldsymbol{D}^m\boldsymbol{\alpha}_t^m\|_{\ell_2}^2$ is a convex loss function to calculate the reconstruction error and $\Omega(\boldsymbol{\Gamma})$ is the non-smooth regularization function with λ being the regularization parameter [Bach et al., 2012]:

$$\|\boldsymbol{\Gamma}\|_{1,2} = \sum_{r=1}^{p}\left\{\sum_{c=1}^{M}|\boldsymbol{\Gamma}_{r,c}|^2\right\}^{1/2} \tag{7.4}$$

where $\boldsymbol{\Gamma}_{r,c}$ is the r-th row and c-th column element of $\boldsymbol{\Gamma}$. The $\ell_{1,2}$ is the summation over the ℓ_2 norm of each row of $\boldsymbol{\Gamma}$ and it results in a matrix $\boldsymbol{\Gamma}$ that has a few nonzero rows [Nguyen et al., 2011]. The test signal $\boldsymbol{Y}_t = \{\boldsymbol{y}_t^m, m \in \mathcal{M}\}$ is assigned to the class c based on minimization of reconstruction error of all M modalities:

$$\min_c \sum_{m=1}^{M} \|\boldsymbol{y}_t^m - \hat{\boldsymbol{y}}_c^m\|_{\ell_2} \tag{7.5}$$

where $\hat{\boldsymbol{y}}_c^m = \boldsymbol{D}^m\delta_c(\boldsymbol{\alpha}_t^m)$ and the sparse representations $\boldsymbol{\alpha}_t^m$ are the m-th column of the solution of Eq.(7.3).

7.3 Dictionary Learning

As noted in Section 7.2, the dictionaries that are fixed and made by horizontally concatenating the training samples without any "learning" phase involved are

sub-optimal for reconstructive and discriminative tasks [Mairal et al., 2009]. To achieve high classification accuracy we need a sufficient number of training samples from each class which leads to a large dictionary with a lot of atoms and therefore an optimization process of high complexity to estimate sparse codes.

Recent studies reported promising results through the use of dictionary learning methods in the reconstructive task like image restoration [Mairal et al., 2008], discriminative tasks like face recognition [Yang et al., 2010] and object tracking [Taalimi et al., 2015c, Taalimi and Qi, 2015]. The learning process usually leads to a compact dictionary with a fewer number of atoms compared to the fixed dictionary scheme [Mairal et al., 2010a, Aharon et al., 2006, Sani and Vosoughi, 2014]. With unsupervised dictionary learning [Elad and Aharon, 2006, Yang et al., 2010, Amini et al., 2014], the resulting dictionary is adapted to the data by minimizing reconstruction error. The learned dictionary outperforms fixed dictionary methods in reconstructive and discriminative tasks. We define the multi-modal dictionary learning and joint sparsity model as the following optimization problem:

$$\underset{\mathcal{D},\boldsymbol{\Gamma}}{\operatorname{argmin}} \frac{1}{2} \sum_{m \in \mathcal{M}} \|\boldsymbol{y}^m - \boldsymbol{D}^m \boldsymbol{\alpha}^m\|_{\ell_2}^2 + \lambda_1 \Omega(\boldsymbol{\Gamma}) + \frac{\lambda_2}{2} \|\boldsymbol{\Gamma}\|_F^2 \tag{7.6a}$$

$$\mathcal{D} \triangleq \{\boldsymbol{D}^m | m \in \mathcal{M}, \text{ s.t. } \forall j \in \{1, \ldots, p\}, \|\boldsymbol{d}_j^m\|_{\ell_2} \leqslant 1\} \tag{7.6b}$$

where λ_1 and λ_2 are regularization parameters and the $\mathcal{D} = \{\boldsymbol{D}^m | m \in \mathcal{M}\}$ is a compact set of the unit-norm ball. The constraint (7.6b) forces all dictionaries $\boldsymbol{D}^m|_{m=1}^M$ to have unit norm in column level. We solve optimization problem (7.6) by splitting it into parts: first, we obtain multi-modal sparse representation of each class $\boldsymbol{\Gamma}_c = [\boldsymbol{\alpha}_c^1, \ldots, \boldsymbol{\alpha}_c^M]$ while multi-modal dictionaries $\mathcal{D} = \{\boldsymbol{D}^m|_{m=1}^M\}$

are initialized by training samples of all classes, $\boldsymbol{Y}^m = [\boldsymbol{Y}_1^m, \ldots, \boldsymbol{Y}_C^m]$. This is elaborated in Section 7.3.1. Then, from the solution of the first step, we construct sparse representation of all classes in m-th modality $\boldsymbol{\Gamma}^m = [\boldsymbol{\alpha}_1^m, \ldots, \boldsymbol{\alpha}_C^m]$ and use it to update dictionary of m-th modality, $\boldsymbol{D}^m$. We express the details of this step in Section 7.3.2.

7.3.1 Estimate Multi-modal Sparse Representation

We use the alternating direction method of multipliers (ADMM) [Parikh and Boyd, 2013] to obtain multi-modal sparse codes, $\boldsymbol{\Gamma}_c \in \mathbf{R}^{p\times M} = [\boldsymbol{\alpha}_c^1, \ldots, \boldsymbol{\alpha}_c^M]$ of training samples of c-th class, $\boldsymbol{Y}_c = \{\boldsymbol{y}_c^m \mid m \in \mathcal{M}\}$ (Fig. 7.1.b). In this step, the dictionary of m-th modality $\boldsymbol{D}^m$ is initialized by training data of m-th modality from all classes (Fig. 7.1.a). We make sure that the dictionary has at least ϱ_c atoms from c-th class: $\sum_{c=1}^{C} \varrho_c = p$ to guarantee that sufficient representation of each class exists in the dictionary. To obtain multi-modal sparse representations, the optimization problem (7.6) is reorganized as follows:

$$\underset{\boldsymbol{\Gamma}}{\operatorname{argmin}} \frac{1}{2} \sum_{m\in\mathcal{M}} \|\boldsymbol{y}^m - \boldsymbol{D}^m \boldsymbol{\alpha}^m\|_{\ell_2}^2 + \lambda_1 \Omega(\boldsymbol{\Gamma}) + \frac{\lambda_2}{2} \|\boldsymbol{\Gamma}\|_F^2 \tag{7.7}$$

To solve (7.7), let us assume $\boldsymbol{Z} \in \mathbf{R}^{p\times M} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M]$ and $\boldsymbol{U} \in \mathbf{R}^{p\times M} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M]$ and both initialized as zero. We denote the proximal operator associated with the norm Ω as $\mathbf{prox}_{\lambda\Omega}$ that maps its domain, vector $\mathbf{p}$, to the vector $\mathbf{q}$, both in $\mathbf{R}^M$: $\mathbf{prox}_{\lambda\Omega}(\mathbf{p}) \triangleq \operatorname{argmin}_{\mathbf{q}} \frac{1}{2}\|\mathbf{p}-\mathbf{q}\|_2^2 + \lambda\Omega(\mathbf{q})$. Then in iteration k we have:

$$\tilde{\boldsymbol{\Gamma}}^{(k+1)} = \mathbf{prox}_{\alpha f}(\boldsymbol{Z}^{(k)} - \boldsymbol{U}^{(k)}) \tag{7.8a}$$

$$\boldsymbol{Z}^{(k+1)} = \mathbf{prox}_{\alpha\Omega}(\tilde{\boldsymbol{\Gamma}}^{(k+1)} + \boldsymbol{U}^{(k)}) \tag{7.8b}$$

$$\boldsymbol{U}^{(k+1)} = \boldsymbol{U}^{(k)} + \tilde{\boldsymbol{\Gamma}}^{(k+1)} - \boldsymbol{Z}^{(k+1)} \tag{7.8c}$$

where $\tilde{\boldsymbol{\Gamma}}^k$ is the optimization solution of Eq.(7.7) at iteration k. The function f in Eq.(7.8a) is defined same as the Eq.(7.3) and hence it is smooth and differentiable while the function Ω in Eq.(7.8b) represents $\ell_{1,2}$ regularization of Eq.(7.4). The optimization variable $\boldsymbol{Z}^k$ is the solution of minimizing the non-smooth joint sparsity regularization and $\tilde{\boldsymbol{\Gamma}}^k$ and is the solution of minimizing the reconstruction error; the solution is designed so that $\boldsymbol{Z}^k$ and $\tilde{\boldsymbol{\Gamma}}^k$ will eventually converge to each other, ($\boldsymbol{U}^{(k+1)} = \boldsymbol{U}^{(k)}$). The proximal step of Eq.(7.8a) is defined as:

$$\mathbf{prox}_{\alpha f}(\boldsymbol{V}) = \underset{\boldsymbol{\Gamma}}{\operatorname{argmin}} \left(f(\boldsymbol{\Gamma}) + \frac{1}{2\alpha}\|\boldsymbol{\Gamma} - \boldsymbol{V}\|_F^2 \right) \tag{7.9}$$

Since f is smooth, we can calculate its gradient as: $\partial f/\partial \boldsymbol{\Gamma} = -\boldsymbol{D}^\intercal \boldsymbol{Y} + \boldsymbol{D}^\intercal \boldsymbol{D}\boldsymbol{\Gamma}$. Therefore, the solution of the optimization problem (7.9) in iteration $k+1$ is:

$$\boldsymbol{\Gamma}^{k+1} = \left(\boldsymbol{D}^\intercal \boldsymbol{D} + \frac{1}{\alpha}\boldsymbol{I} \right)^{-1} \left(\boldsymbol{D}^\intercal \boldsymbol{Y} + \frac{1}{\alpha}\boldsymbol{V} \right) \tag{7.10}$$

where $\boldsymbol{I}$ is identity matrix. Equation (7.8b) represents the proximal step over $\boldsymbol{Z}$:

$$\mathbf{prox}_{\alpha\Omega}(\boldsymbol{V}) = \underset{\boldsymbol{Z}}{\operatorname{argmin}}\ \Omega(\boldsymbol{Z}) + \frac{1}{2\alpha}\|\boldsymbol{Z} - \boldsymbol{V}\|_F^2 \tag{7.11}$$

Substituting Ω by Eq.(7.4), we obtain:

$$\mathbf{prox}_{\alpha\Omega}(\boldsymbol{V}) = \underset{\boldsymbol{Z}}{\operatorname{argmin}} \sum_{r=1}^{p} \left(\|\boldsymbol{z}_{r\rightarrow}\|_2 + \frac{1}{2\alpha}\|\boldsymbol{z}_{r\rightarrow} - \boldsymbol{v}_{r\rightarrow}\|_{\ell_2}^2 \right) \tag{7.12}$$

where $\boldsymbol{z}_{r\rightarrow}$ and $\boldsymbol{v}_{r\rightarrow}$ are the r-th row of the $\boldsymbol{Z}$ and $\boldsymbol{V}$, respectively. The α in the problem (7.12) and λ in Eq.(7.3) have an inverse relation. The optimization problem (7.12) is solved in p independent optimization, corresponding to p atoms, while each optimization is done on an M-dimensional vectors, $\boldsymbol{z}_{r\rightarrow}$. We solve the proximal step of inducing joint sparsity regularization of Eq. (7.3) using the SPArse Modeling Software (SPAMS) [Jenatton et al., 2010, Mairal et al., 2010b].

7.3.2 Update dictionary atoms

In Section 7.3.1 we obtain multi-modal and joint sparse representations of each class, $\boldsymbol{\Gamma}_c = [\boldsymbol{\alpha}_c^1, \ldots, \boldsymbol{\alpha}_c^M]$ while multi-modal dictionaries are initialized by the training data. In this section, the proposed method is extended to include unsupervised multi-modal dictionary learning, $\boldsymbol{D}^m|_{m=1}^M$, while collaboration between different features are enforced using the joint sparsity model (7.4). As it is shown in Fig. 7.1.c, we construct sparse representation of m-th modality by horizontally concatenating sparse codes of all classes from the same modality: $\boldsymbol{\Gamma}^m = [\boldsymbol{\alpha}_1^m, \ldots, \boldsymbol{\alpha}_C^m]$. The dictionary $\boldsymbol{D}^m = [\boldsymbol{d}_1^m, \ldots, \boldsymbol{d}_p^m]$, will be updated by solving the optimization (7.6) using the Iterative Projection Method proposed in [Rosasco et al., 2009]. Since $\boldsymbol{\Gamma}^m$ is fixed, the problem of (7.6) is changed to (7.13).

$$\underset{\boldsymbol{D}^m}{\operatorname{argmin}} \ \|\boldsymbol{Y}^m - \boldsymbol{D}^m\boldsymbol{\Gamma}^m\|_F^2 \ \text{ s.t } \boldsymbol{D}^m \in \mathcal{D} \tag{7.13}$$

Now, the dictionary is updated atom by atom. The q-th dictionary atom is updating and the problem is rewritten to (7.14).

$$\underset{\boldsymbol{d}_q^m}{\operatorname{argmin}}\ Tr(\boldsymbol{D}^{m\intercal}\boldsymbol{D}^m\boldsymbol{\Gamma}^m\boldsymbol{\Gamma}^{m\intercal} - 2\boldsymbol{D}^{m\intercal}\boldsymbol{Y}^m\boldsymbol{\Gamma}^{m\intercal}) \quad \text{s.t } \|\boldsymbol{d}_q^m\|_{\ell_2} \leqslant 1 \tag{7.14}$$

Let $\boldsymbol{\Theta} = \boldsymbol{\Gamma}^m\boldsymbol{\Gamma}^{m\intercal}$, $\boldsymbol{\Upsilon}^m = \boldsymbol{Y}^m\boldsymbol{\Gamma}^{m\intercal}$. The q-th dictionary atom is updated and the problem is reformulated as follows.

$$\underset{\boldsymbol{d}_q^m}{\operatorname{argmin}}\ Tr(\boldsymbol{D}^{m\intercal}\boldsymbol{D}^m\boldsymbol{\Theta}^m - 2\boldsymbol{D}^{m\intercal}\boldsymbol{\Upsilon}^m) \quad \text{s.t } \|\boldsymbol{d}_q^m\|_{\ell_2} \leqslant 1 \tag{7.15}$$

where $\boldsymbol{d}_q^m$ is the q-th column vectors of $\boldsymbol{D}^m$. Let $\Theta_{q,q}^m$ be the element in q-th column and q-th row of $\boldsymbol{\Theta}^m$, $\boldsymbol{\Theta}_q^m$ be the q-th column vectors of $\boldsymbol{\Theta}^m$, and $\boldsymbol{\Upsilon}_q{}^m$ be the q-th column vectors of $\boldsymbol{\Upsilon}^m$. According to the algorithm of dictionary updating proposed in [Mairal et al., 2010a], dictionary atom $\boldsymbol{d}_q^m$ with corresponding $\Theta_{q,q}^m > 0$, is updated and is normalized to have unit l_2-norm as follows:

$$\boldsymbol{d}_q^m = \frac{\boldsymbol{\Upsilon}_q{}^m - \boldsymbol{D}^m\boldsymbol{\Theta}_q{}^m}{\Theta_{q,q}^m + 1/\alpha} \tag{7.16a}$$

$$\Pi_{\mathcal{D}} = \{\boldsymbol{d}_q^m\}_{q=1}^p = \begin{cases} \boldsymbol{d}_q^m & \text{if } \|\boldsymbol{d}_q^m\|_{\ell_2} < 1 \\ \frac{\boldsymbol{d}_q^m}{\|\boldsymbol{d}_q^m\|_{\ell_2}} & \text{otherwise} \end{cases} \tag{7.16b}$$

which will converge after several iterations. Algorithm 3 shows the steps required to learn the multi-modal unsupervised dictionary and the joint sparse modeling using joint sparsity regularization.

Algorithm 3: Multimodal dictionary learning and joint sparse modeling

Input: $\boldsymbol{Y}^m \; \forall m \in \{1 \cdots M\}, iter$

```
begin
    Initialize D^m with samples of m-th modality of all classes.
    foreach k = 1 to iter do
        Fix D^m|_{m=1}^M and estimate Γ^m|_{m=1}^M of each class
        foreach Class c ∈ {1, ..., C} do
            Obtain multi-modal Γ_c = [α_c^1, ..., α_c^M] using joint sparse
            modeling (Section 7.3.1).
        foreach Modality m ∈ {1, ..., M} do
            Construct Γ^m = [α_1^m, ..., α_C^m].
            Update dictionary D^m (Section 7.3.2)
```

7.3.3 Sparse Codes Pooling

Patch based approach of calculating features and corresponding sparse codes result in obtaining local texture features, but we also need the spatial information for each image by aggregating the information of local patches. A naive solution is to concatenate the features of all patches in each image but this results in a long vector of sparse codes, which has two main problems. Firstly, the neighboring patch information is lost and secondly, the final size of the feature vector varies depending on the number of patches for each image. We describe how this issue is addressed in the proposed method in Fig. 7.1.*d* and Fig. 7.3.

Fig. 7.2 shows the Spatial Pyramid Matching (SPM) [Lazebnik et al., 2006] method that divides the image into 1, 4 and 16 non-overlapping regions (21 regions in total) and performs max-pooling on the sparse codes in each region to finally produce a feature vector of size $(1+4+16)\times p$, where p is the number of atoms in the dictionary. A limitation of this approach is that the image is blindly divided

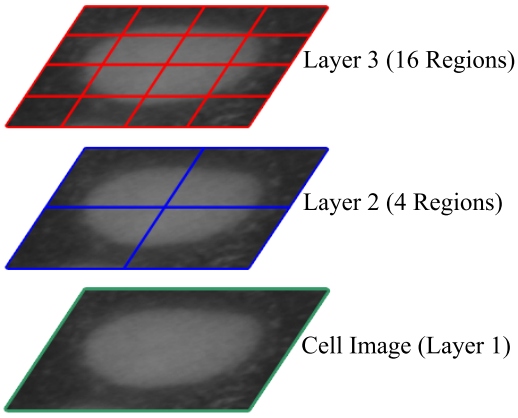


Figure 7.2: SPM method.

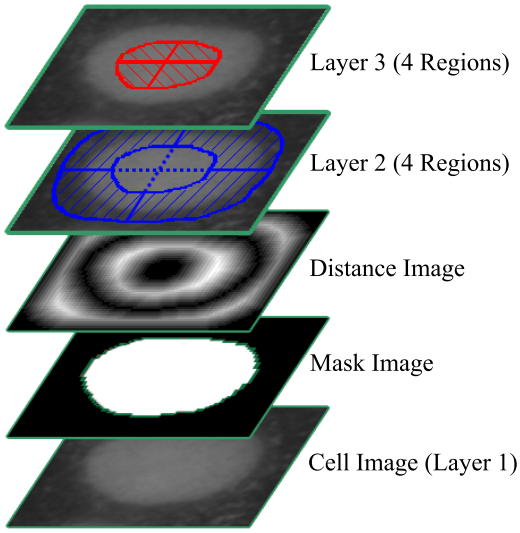


Figure 7.3: Proposed SCP method.

into different layers without taking into account the underlying information in the image. As evident from Fig 7.2, the information pertaining to the cell boundary and inside cells are totally different but the SPM combines them nevertheless. Moreover, the SPM results in a large regions (e.g. 21 regions) and concatenating them all, produces a long feature vector for classification.

To alleviate these limitations, we propose a Sparse Codes Pooling (SCP) method which is shown in Fig. 7.3. 'Layer 1' is the whole cell image and the information of all the image patches are pooled. The *distance transform* is applied on the cell mask, which assigns a value to each image pixel with the Euclidean distance to the nearest cell boundary pixel. As can be seen in Fig. 7.3, two boundaries are extracted from the *distance image*, which are shown in blue circles in 'Layer 2' and create the tube-shape region around the cell boundary. This layer is then divided to four regions as in SPM. 'Layer 3' is created by using the inner circle of the 'Layer 2' and also divided to four regions. The pooling strategy is

then applied on the regions and all the feature vectors concatenated.

This approach benefits two main advantages; first, the final feature dimension vector is $9\times p$, which is around 57% lower than $21\times p$ in SPM. Second, the most informative area of the cells are near cell boundaries (e.g. Golgi and Nucleolar Membrane classes) and inner area of cells (e.g. Nucleolar and Speckled classes) as is evident in Fig. 3.5. By focusing on these two important areas, we can obtain more informative and discriminative feature vectors.

By considering the three image layers $l \in \{0, 1, 2\}$, a pooling function $\mathcal{F}$ is applied on the sparse codes $\boldsymbol{h}^l = [s_1^l, s_2^l, \cdots, s_{n_l}^l]$ in each layer, where s_i^l is the sparse codes of image patch i in layer l and n_l is the number of image patches in layer l. The final feature vector for layer l is $\boldsymbol{x}^l$.

$$\boldsymbol{x}^l = \mathcal{F}(\boldsymbol{h}^l) \tag{7.17}$$

The one-hot encoding, mean- and max- pooling functions are studied. In one-hot encoding, just one representative atom from dictionary is selected by having only one non-zero element in the final sparse code vector which is calculated as follows:

$$\boldsymbol{T}_l = \max\{\boldsymbol{h}^l\} \tag{7.18a}$$

$$\boldsymbol{x}_j^l = \begin{cases} 0 & if\ \boldsymbol{h}_{ji}^l < T_l \\ T_l & if\ \boldsymbol{h}_{ji}^l \geqslant T_l \end{cases} \qquad i = \{1, 2, \cdots, n_l\} \tag{7.18b}$$

where T_l is the maximum sparse code of all patches in layer l and $\boldsymbol{x}_j^l$ is the j-th element of final feature vector.

For mean- and max-pooling, the average and maximum values for each row

of $\boldsymbol{h}^l$ is selected. For instance, the max-pooling function is:

$$\boldsymbol{x}_j^l = \max\{\boldsymbol{h}_{ji}^l\}, \quad i = \{1, 2, \cdots, n_l\} \tag{7.19}$$

The effect of pooling function on the final classification performance is discussed in 7.4.4.

7.4 Experiments and Results

7.4.1 Feature Extraction

We extract gradient based features of SIFT with size 128 and SURF with size 64 from each sample in an overlapping patches. The patch size is 12×12 and the distance between patches is 4 pixels. According to the size of the images, the number of patches is different. However, to train the dictionaries, we randomly select 100 patches from each image to get the balanced distribution of patch samples from all the input images.

7.4.2 Evaluation Strategies

The HEp-2 classification problem is divided into two categories, *Cell* and *Specimen Level*. In the *Cell Level* classification, each cell is classified solely without considering other neighboring cells. In contrast, the *Specimen Level* classification focuses on classifying whole specimen image containing many cells. As described in Section 3.3, two HEp-2 datasets are publicly available (ICPR2012 and ICIP2013) where the following experimental scenarios are exploited to evaluate the proposed

method:

i 'Test set' evaluation, which can be done only on ICPR2012, for which the test set is publicly available but not for ICIP2013 for which a test set is not provided.

ii 'Leave-One-Specimen-Out' (LOSO), where all the cells from one specimen image are used for test and the rest of the specimen cells for training. This scenario is applied to both datasets.

iii 'HSM' evaluation method proposed by [Han et al., 2014], 600 cells from each class (300 cells from Golgi class) are randomly selected for training and the rest of the cells are used for the test set. This strategy is only applied on ICIP2013 for comparison with other methods.

It should be noted that the cell masks for both datasets are provided but the masks are inaccurate specifically for the *Specimen Images* in ICIP2013. For instance, some masks contain non-cell areas and 'touching cells' are not accurately divided. Therefore, to get the correct cell masks, the cell extraction method that is described in [Ensafi et al., 2014b] is used, where morphological features are exploited to extract those cells with similar *curvature* (Solidity property) and *area*.

To report the classification results, the MCA (see Section 3.4) is used.

7.4.3 Results

The proposed JMCDL classification method is evaluated and the results are discussed in this section. We compare the proposed algorithm with the state-of-the-

Table 7.1: The MCA accuracy on ICPR2012 dataset by using two evaluation strategies 'Test set' and 'Leave-One-Specimen-Out' (LOSO).

ICPR2012 (%)			**Proposed**					**DL-based Methods**			**Other Methods**	
					SCDL							
			JMCDL	JMC	SIFTSURF	SIFT	SURF	Ensafi⋆	SNPB*	Kastaniotis†	Nosaka◇	DiCataldo‡
Test set	*Cell Level*	Positive	**82**	78	76	74	72	81	**82**	70	79	60
		Intermediate	**79**	72	69	67	66	62	59	31	58	35
		Average	**80**	75	73	70	69	72	70	51	69	48
	Specimen Level		**93**	86	86	79	64	86	**93**	86	79	**93**
LOSO	*Cell Level*	Positive	**96**	92	90	86	82	91	92	72	80	95
		Intermediate	**84**	80	77	74	71	72	70	55	60	80
		Average	**90**	86	84	80	77	82	81	64	70	88
	Specimen Level		**93**	88	86	79	64	79	86	79	86	**93**

⋆ [Ensafi et al., 2014a] * [Ensafi et al., 2015] † [Theodorakopoulos et al., 2014]
◇ [Nosaka and Fukui, 2014] ‡ [Di Cataldo et al., 2014]

art HEp-2 cell classification methods that demonstrates the significant influence of enforcing different modalities to have similar sparsity pattern while learning multi-modal dictionaries. We also investigate the effect of proposed SCP pooling strategy on the classification performance.

ICPR2012. Table. 7.1 shows the accuracies on ICPR2012 by using 'Test set' and 'LOSO' evaluation methods for both tasks of *Cell* and *Specimen Level* classifications.

The proposed JMCDL has two major components: 1. dictionary learning method and, 2. joint sparsity regularization. We evaluate the performance of each novel components of the proposed method and the whole system on Table. 7.1. We express the performance of JMCDL once without joint sparsity regularization

to observe the effect of proposed dictionary learning. Since this scenario is equal to have only one feature modality, we call it Single-Cue Dictionary Learning (SCDL) and it includes three scenarios: surf only ('SURF'), sift only ('SIFT') and 'SIFT-SURF' that is made by putting together sift and surf features in one vector. The impact of joint sparsity regularization while dictionary is learned by [Mairal et al., 2009] is reported as JMC and finally, the JMCDL reflects the performance of the whole system of joint dictionary learning and multi-modal sparsity regularization.

We compare classification accuracy of JMCDL with three state-of-the-art HEp-2 classifiers that are based on dictionary learning (DL) in 'DL-based Methods' part of the Table: [Ensafi et al., 2014a] use SIFT, [Ensafi et al., 2015] (SNPB) exploit both SIFT and SURF and [Theodorakopoulos et al., 2014] consider modified version of Local Binary Patterns (LBP) features. We also bring the performance of two state-of-the-art and non-sparse based representation methods to compare with the JMCDL; the winner of the ICPR2012 contest[1] [Nosaka and Fukui, 2014] and [Di Cataldo et al., 2014], that exploit LBP, morphological and textural features.

Table. 7.1 shows that the proposed dictionary learning outperforms other methods. Learning dictionary by elastic-net (JMC column) [Zou and Hastie, 2005] while enforcing multi-modal joint sparse regularization outperforms SCDL on average by 5% and 4% in 'Test set' and 'LOSO' evaluation methods. In 'Test set' evaluation strategy, JMCDL increases the accuracy of SIFT and SURF more than 10% and enhances SIFTSURF around 7%. Also, JMCDL shows superior results comparing to the DL-based and other methods particularly in *Cell Level*,

[1] `http://mivia.unisa.it/hep2contest/index.shtml`

where 80% and 90% accuracies are obtained in 'Test set' and 'LOSO' strategies, respectively. These results are 8% better than other DL-based methods in both evaluation strategies and 11% and 2% above the other methods.

Additionally, a significant achievement is obtained on intermediate intensity level classification, where more than 10% in 'Test set' and 4% accuracies in LOSO strategies are improved.

For the *Specimen Level* classification 93% accuracy is obtained, which is similar to other best performances and this is mostly because the number of *specimen* images is limited to 28. It is expected to achieve better results in comparison with other methods by increasing the number of images, as it happened in ICIP2013 dataset, which we will discuss later in this section.

ICIP2013. Comparison of results for ICIP2013 dataset is shown in Table. 7.2. The 'HSM' and 'LOSO' evaluation strategies are used (see section. 7.4.2) for both *Cell* and *Specimen Level* classification tasks where for the *Cell Level* classification task, the positive and intermediate intensity level images are exploited.

The JMCDL method is compared with SCDL, DL-based and other methods. This is also compared with [Manivannan et al., 2014c], the winner of I3A contest[2] (Pattern Recognition Techniques for Indirect Immunofluorescence Images) hosted by International Conference on Pattern Recognition (ICPR) 2014, which is performed on ICIP2013 dataset.

The performance of proposed dictionary learning using 'SIFTSURF' is promising since it performs slightly better than HSM and it can get close result to the SNPB based on HSM measurement. However, based on LOSO standard, SCDL

[2] `http:\i3a2014.unisa.it`

Table 7.2: The MCA accuracy on ICIP2013 dataset by using two evaluation strategies 'HSM' [Han et al., 2014] and 'Leave-One-Specimen-Out' (LOSO).

ICIP2013 (%)			Proposed					DL-based Methods				Other Methods	
					SCDL								
			JMCDL	JMC	SIFTSURF	SIFT	SURF	Ensafi⋆	SNPB*	Gragnaniello†	manivannan◇	HSM‡	Larsen▽
HSM	*Cell Level*	Positive	**98.5**	96.9	96.1	92.3	84.3	95.8	96.8	-	-	95.5	-
		Intermediate	**93.2**	88.7	87.4	86.8	69.7	87.9	88.8	-	-	80.9	-
		Average	**95.9**	92.8	91.8	89.6	77	91.9	92.8	-	-	88.2	-
LOSO	*Cell Level*	Positive	**87.6**	86.8	86.1	82.8	78.2	83.4	83.8	-	-	-	-
		Intermediate	**77.5**	76.9	76.4	68.4	63.4	71.2	72	-	-	-	-
		Average	**82.6**	81.8	81.3	75.6	70.8	77.3	77.9	81.1	80.3	-	78.7
	Specimen Level		**91.6**	89.2	88	84.3	77.1	88	89.2	86.7	89.9	-	-

⋆ [Ensafi et al., 2014b] * [Ensafi et al., 2015] † [Gragnaniello et al., 2014]
◇ [Manivannan et al., 2014c] ‡ [Han et al., 2014] ▽ [Larsen et al., 2014]

using 'SIFTSURF' outperforms all the state-of-the-art methods. Learning dictionary by elastic-net (JMC column) [Zou and Hastie, 2005] while enforcing multi-modal joint sparse regularization outperforms SCDL. The whole system reported in JMCDL obtains better classification accuracy than SCDL using 'SIFTSURF' on average by 4.1% and 1.3% based on HAM and LOSO, respectively.

Table. 7.2 also shows other DL-based methods, where JMCDL outperform 2.3% from the I3A contest winner [Manivannan et al., 2014c] and 1.5% from [Gragnaniello et al., 2014]. JMCDL achieved more than 5% accuracies better than [Ensafi et al., 2014b, Ensafi et al., 2015], which have used SIFT and SURF features in their methodologies. This comparison clearly shows the effect of multi-modal dictionary learning and joint sparse model, which is applied on a large dataset ICIP2013.

For *Specimen Level* classification, the JMCDL outperforms other state-of-

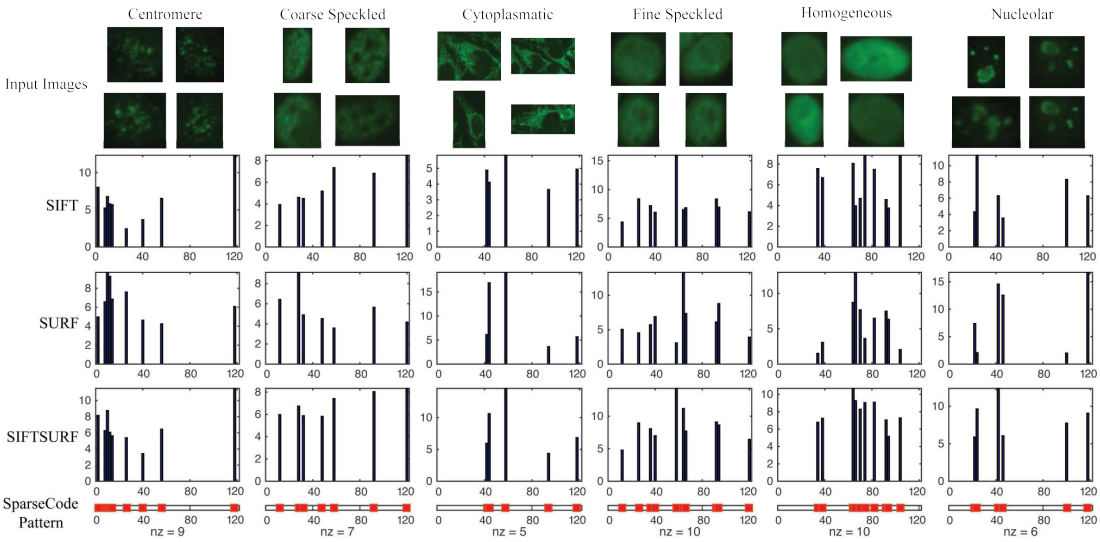


Figure 7.4: Representation coefficients generated by proposed regularization for SIFT, SURF and SIFTSURF features. There are six columns corresponding for six classes.

the-art results by improving 1.7% accuracy in comparison with the I3A contest winner [Manivannan et al., 2014c] and 4.9% from [Gragnaniello et al., 2014].

7.4.4 Sparse Representation With Similar Pattern

The imposed joint sparsity model in this scheme, makes sparse codes more discriminative and hence produces better classification results. The similar patterns are shown in Fig. 7.4, where the first row shows cell sample of the six classes. The sparse representation of each cell class is provided for various features: SIFT, SURF and SIFTSURF. Also, the pattern of the sparse codes imposed by regularization function are presented in the last row. It is evident from Fig. 7.4 that the sparse codes' patterns for different modalities are similar as this is imposed by $\ell_{1,2}$ regularization term on the multimodal sparse codes (see Section 7.2.2).

Table 7.3: The comparison of proposed SCP with SPM strategy by using different pooling functions and using LOSO evaluation method.

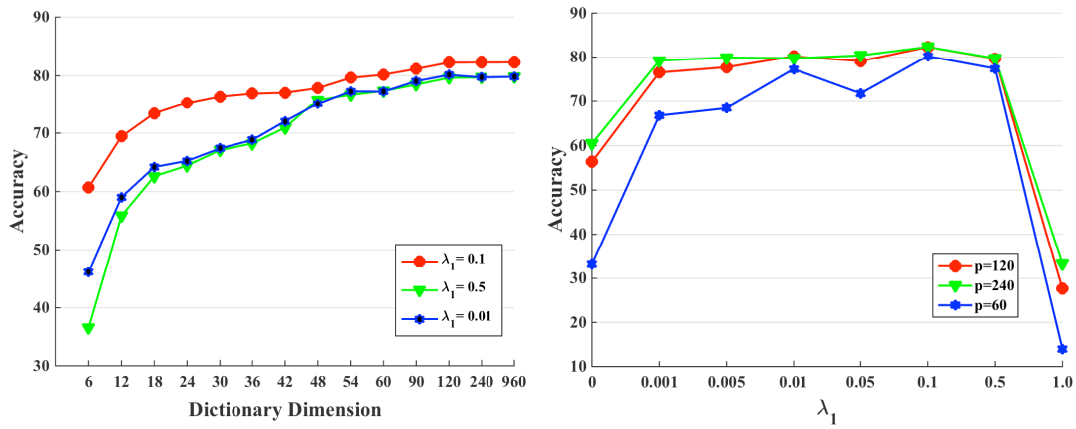
	JMCDL+SCP			JMCDL+SPM			SPM		
	One-hot	Mean	Max	One-hot	Mean	Max	One-hot	Mean	Max
ICPR2012	66.7	84.2	**90.0**	61.3	80.2	86.7	58.1	78.6	82.1
ICIP2013	54.8	76.8	**82.6**	51.4	73.8	78.4	50.5	73.6	77.3

7.4.5 SCP Versus SPM

The effect of proposed SCP pooling strategy is studied and compared with SPM method for two datasets as shown in Table. 7.3. The first two parts of the Table. 7.3 compares the JMCDL with applying SCP and SPM, where the sparse coding and dictionary learning schemes are the same but differs in pooling method. It is evident that the max-pooling strategy outperforms others in both methods however, the combination of JMCDL and SCP obtains better results than other methods. The last part of the Table. 7.3 shows the sparse coding scheme combined with SPM that is used by [Ensafi et al., 2014a] where, JMCDL+SCP outperforms SPM by 7.9% and 5.3% on ICPR2012 and ICIP2013 datasets, respectively.

7.4.6 Parameter Study

In this section, two main parameters of the proposed method are analyzed. In particular, the dimension of the dictionary p plays a significant role where a larger number of atoms with much higher feature vector dimension creates an *over complete* dictionary. Such *over complete* dictionary is biologically inspired from human cortex and often gives better classification accuracy [Rehn and Sommer, 2007]. On the other hand, calculating the *over complete* dictionaries are



(a) (b)

Figure 7.5: The accuracy of ICPR2012 positive test set versus different dictionary atoms (a) and λ_1 values (b).

computationally expensive. Fig. 7.5.*a* shows the classification performance with different dictionary dimensions. It is obvious from Fig. 7.5.*a* that the performance keeps improving with the increase of the dictionary dimension until the dictionary dimension reaches 120 where the best performance is obtained.

The other most impactful parameter is regularization coefficient λ_1 in equation 7.6. Fig. 7.5.*b* shows the classification performance versus the regularization parameter. When the λ_1 is near zero, the reconstruction error influences more and provides non-sparse codes. By increasing λ_1 value, the sparsity of the weights helps increase the accuracy. However, by further increasing the λ_1 value, the sparseness of the codes dominates the reconstruction error that reduces the classification accuracy. This study shows that the $\lambda_1 = 0.1$ performs better than the other values.

7.5 Conclusion

The HEp-2 cell classification task is studied in the sparsity scheme and a method is proposed to learn the multi-modal dictionaries while obtaining multi-modal sparse representation of each class using joint sparsity model. The imposed joint sparsity enabled the algorithm to fuse information at feature-level by forcing their sparse codes to have similar basis. This is done using $\ell_{1,2}$ regularization that enforces high amount of correlation between different modalities of each cell class. In other words, we know a priori that the modality configuration (here, SIFT and SURF) induces a strong group structure that is encoded in the optimization using $\ell_{1,2}$ regularization (joint sparsity). That is because joint sparsity gives a strong statistical co-occurrence structure: if a sample belongs to the c-th class most of its modalities should have the same label, so knowing the label of one source can act as a strong prior for inferring the label of others.

Additionally, the SCP pooling strategy method is proposed to intelligently combine the sparse codes of the image patches by estimating the type of underlying information according to boundary and inner-side of the cells. JMCDL obtained better performance in comparison with other state-of-the-are results in both *Cell* and *Specimen Level* HEp-2 classification.

Chapter 8

Conclusions and Future Works

This thesis presents new methods and solutions for HEp-2 cell classification problem with the aim of diagnosing ADs. The dictionary learning and sparse coding scheme with their challenges including preparing informative image patches, estimating dictionary size, distributed and joint multimodal dictionary learning are investigated and the proposed new methods have been shown to improve the final classification result.

First, instead of traditional image patches we proposed a novel superpixel method that requires fewer patches while resulting in increased information in each superpixel. Then, we addressed the challenge of finding an optimal dictionary dimension, which is very crucial to the performance of the learning procedure. To this end, a non-parametric Bayesian approach is proposed to automatically obtain the optimal dictionary size. Furthermore, we reduced the computational cost and memory requirements of the dictionary learning stage, using an adaptive distributed learning method. This proposed method efficiently divides the

dictionary learning procedure into different nodes of a network. Finally, a feature fusion method is proposed to obtain informative and discriminative features for HEp-2 cell classification problem.

A summary of the accomplishments are presented below together with suggestions for possible future directions for the research.

8.1 Feature Selection and Sparse Coding Representation

The BoW model on sparse coding scheme is proposed to the HEp-2 cell classification problem, where the effects of integrating the SIFT and SURF features and the optimal number of iterations for dictionary learning are investigated. Experiments on the datasets show that superior classification accuracy is obtained across cell, specimen, positive intensity, and intermediate intensity level images. The effect of pooling strategies including max-, average-pooling and one-hot encoding is also extensively studied. This is evident from the study that the max-pooling strategy outperforms other pooling strategies. Moreover, a novel algorithm is proposed to extract the cells from their inaccurate masks. This stage is essentially important for specimen level classification where all the cells in one specimen represent the final pattern category of the image.

Beyond the SIFT and SURF features, the effect of other features to the final performance could be an initial future work. Specifically, these features are extracted from the image patches which are dependent on the size and shape of the patches. A challenging problem would be to develop an automatic method

to initialize the size and shape of the patches to gain more discriminative and representative features. Moreover, developing a cell segmentation approach can be an interesting problem to extract cells from the specimen images, rather than working on provided inaccurate cell masks.

8.2 Learning Dictionary Dimension

The size of the dictionary in the sparse coding scheme is highly correlated with the classification performance and is essential to be chosen wisely. To avoid choosing the dictionary size manually, a non-parametric Bayesian approach is proposed. With the help of IBP, the intuitively infinite sized dictionary is initialized and iteratively decreases the dictionary size such that to minimize the cost function. Experiments show the classification performance is improved with the dictionary dimension much lower than their equivalent methods that used manual dictionary sizes. The lower dictionary dimension also decreases the computational time in the test stage.

Although the testing stage performs fast because of the low dimensional dictionary, the training stage is computationally complex. A possible future work could be to facilitate this process. Parallelizing the procedure or applying greedy algorithms could be beneficial. Additionally, it is evident from the plot of the dictionary size versus the number of iterations that the dictionary dimension increases in first steps and finally decreases to its steady state. This trend of convergence seems interesting and can be mathematically investigated.

8.3 Adaptive Distributed Dictionary Learning

The proposed adaptive distributed dictionary learning method benefits from lower computational cost with lower number of tuning parameters which is an important advantage in solving classification problems. Moreover, learning the combination weights adaptively is an important contribution for the proposed method which makes it capable of adjusting itself for different datasets according to the input images. The ADDL method divides the dictionary in a network to N nodes, where each node is responsible of updating itself according to the input data. Moreover, we propose to combine the information of neighboring nodes in an adaptive way that enables the nodes to learn about the usefulness of the information received from their neighbors.

This approach is a foundation for big data analysis where the information is available on the nodes of a computer cluster or cloud. The way that each node communicate with the neighboring node can reduce the burden of information in each cluster by ignoring misleading information.

8.4 Superpixels as Image Patches

In the patch-based image processing scheme, unlike the overlapping fixed-size image patches, a superpixel method is proposed to capture the regions with high level of information. Because the boundaries of superpixels align with the high gradient pixels, the superpixels are then dilated to capture this important information of the images. By applying the proposed method, the number of patches reduces exponentially in contrast to previous patch based techniques and the information in

the patches are maximized. In this method, the number of patches (superpixels) is dramatically decreased, which reduces the dictionary learning computational complexity.

In the superpixel algorithm, the intensity and gradient values are used to group the pixels. A natural next step of this work is to modify the superpixel algorithm by adding more characteristics of images. For example, a variety of engineered features including SIFT, HoG, LBP, etc. can also be added not only forcing the algorithm to capture high gradient information but also gathering the high-level characteristics in one superpixel.

8.5 Feature Fusion and Joint Sparsity Model

A multi-modal dictionary learning is proposed by applying label consistency constraint to learn discriminative and reconstructive sparse codes. The algorithm fuses the input features to obtain similar patterns of the sparse codes for all the modalities in each class which is called joint sparse coding. The similarity patterns of obtained sparse codes make the final classification procedure more accurate.

A natural extension of this work is to add more complex and well-designed features to the system. One of the challenges on HEp-2 cells classification is on the illuminance variance, where a feature which is robust on illumination changes can be beneficial. Recently, Spatial Shape Index Descriptor (SSID) and Local Orientation Adaptive Descriptor (LOAD) features are proposed [Qi et al., 2016]. These features could increase the amount of informative knowledge of the images leading to more representative sparse codes.

8.6 Other Databases

The extension of the proposed methods in this thesis can also applied to other databases including handwriting, texture, natural scenes, etc. by some modifications. The feature selection methods should be modified to capture relevant features however, the classification method can remain the same. Specifically, in image patch analysis methods, the superpixel method can help to capture image patches with high amount of information. These experiments are remained for future work.

Bibliography

[Achanta et al., 2012] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282. 89, 90

[Aharon et al., 2006] Aharon, M., Elad, M., and Bruckstein, A. (2006). k -svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322. 110

[Amini et al., 2014] Amini, S., Sadeghi, M., Joneidi, M., Babaie-Zadeh, M., and Jutten, C. (2014). Outlier-aware dictionary learning for sparse representation. In *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*, pages 1–6. 110

[Amiri and Haykin, 2014] Amiri, A. and Haykin, S. (2014). Improved sparse coding under the influence of perceptual attention. *Neural computation*, 26(2):377–420. 33

[Bach et al., 2012] Bach, F., Jenatton, R., Mairal, J., and Obozinski, G. (2012). Optimization with sparsity-inducing penalties. *Found. Trends Mach. Learn.*,

4(1):1–106. 109

[Bahrampour et al., 2014] Bahrampour, S., Ray, A., Nasrabadi, N., and Jenkins, K. (2014). Quality-based multimodal classification using tree-structured sparsity. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 4114–4121. 104

[Bay et al., 2008] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359. 31

[Bel Haj Ali et al., 2012] Bel Haj Ali, W., Giampaglia, D., Barlaud, M., Piro, P., Nock, R., and Pourcher, T. (2012). Classification of biological cells using bio-inspired descriptors. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3353–3357. 14

[Bertsekas, 1997] Bertsekas, D. P. (1997). A new class of incremental gradient methods for least squares problems. *SIAM Journal on Optimization*, 7(4):913–926. 74

[Candès et al., 2006] Candès, E. J., Romberg, J. K., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223. 20

[Chen et al., 2015a] Chen, J., Richard, C., and Sayed, A. H. (2015a). Diffusion LMS over multitask networks. *Signal Processing, IEEE Transactions on*, 63(11):2733–2748. 76, 79

[Chen and Sayed, 2012] Chen, J. and Sayed, A. H. (2012). Distributed pareto-optimal solutions via diffusion adaptation. In *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP)*, pages 648–651. 74

[Chen et al., 2014] Chen, J., Towfic, Z. J., and Sayed, A. H. (2014). Online dictionary learning over distributed models. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 3874–3878. IEEE. 73, 77

[Chen et al., 2015b] Chen, J., Towfic, Z. J., and Sayed, A. H. (2015b). Dictionary learning over distributed models. *Signal Processing, IEEE Transactions on*, 63(4):1001–1016. 70, 73, 77

[Cotsapas and Hafler, 2013] Cotsapas, C. and Hafler, D. A. (2013). Immune-mediated disease genetics: the shared basis of pathogenesis. *Trends in immunology*, 34(1):22–26. 9

[Di Cataldo et al., 2014] Di Cataldo, S., Bottino, A., Islam, I. U., Vieira, T. F., and Ficarra, E. (2014). Subclass discriminant analysis of morphological and textural features for hep-2 staining pattern classification. *Pattern Recognition*, 47(7):2389–2399. 15, 46, 58, 64, 82, 120, 121

[Elad and Aharon, 2006] Elad, M. and Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745. 110

[Ensafi et al., 2014a] Ensafi, S., Lu, S., Kassim, A. A., and Tan, C. L. (2014a). Automatic cad system for hep-2 cell image classification. In *Pattern Recognition*

(ICPR), 2014 22nd International Conference on, pages 3321–3326. IEEE. 25, 31, 37, 46, 58, 64, 66, 82, 120, 121, 125

[Ensafi et al., 2014b] Ensafi, S., Lu, S., Kassim, A. A., and Tan, C. L. (2014b). A bag of words based approach for classification of hep-2 cell images. In *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*, pages 29–32. IEEE. 25, 26, 34, 37, 51, 58, 64, 65, 66, 84, 85, 97, 103, 119, 123

[Ensafi et al., 2015] Ensafi, S., Lu, S., Kassim, A. A., and Tan, C. L. (2015). Sparse non-parametric bayesian model for hep-2 cell image classification. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 679–682. 82, 84, 85, 97, 120, 121, 123

[Faraki et al., 2014] Faraki, M., Harandi, M. T., Wiliem, A., and Lovell, B. C. (2014). Fisher tensors for classifying human epithelial cells. *Pattern Recognition*, 47(7):2348–2359. 14, 46

[Felzenszwalb and Huttenlocher, 2004] Felzenszwalb, P. F. and Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181. 89

[Foggia et al., 2014] Foggia, P., Percannella, G., Saggese, A., and Vento, M. (2014). Pattern recognition in stained hep-2 cells: Where are we now? *Pattern Recognition*, 47(7):2305–2314. 12, 27, 37, 81, 86, 92, 94

[Foggia et al., 2010] Foggia, P., Percannella, G., Soda, P., and Vento, M. (2010). Early experiences in mitotic cells recognition on hep-2 slides. In *Computer-*

based medical systems (CBMS), 2010 IEEE 23rd international symposium on, pages 38–43. IEEE. 11

[Foggia et al., 2013] Foggia, P., Percannella, G., Soda, P., and Vento, M. (2013). Benchmarking hep-2 cells classification methods. *Medical Imaging, IEEE Transactions on*, 32(10):1878–1889. xiv, 11, 13, 14, 18, 37, 46, 49, 64, 81, 86, 95

[Friedman, 1997] Friedman, J. (1997). On bias, variance, 0/1—loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*, 1(1):55–77. 13

[Gao et al., 2014] Gao, Z., Zhang, J., Zhou, L., and Wang, L. (2014). Hep-2 cell image classification with convolutional neural networks. In *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*, pages 24–28. 17, 18

[Gershman and Blei, 2012] Gershman, S. J. and Blei, D. M. (2012). A tutorial on bayesian nonparametric models. *Journal of Mathematical Psychology*, 56(1):1–12. 59

[Ghosh and Chaudhary, 2012] Ghosh, S. and Chaudhary, V. (2012). Feature analysis for automatic classification of hep-2 florescence patterns: Computer-aided diagnosis of auto-immune diseases. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 174–177. IEEE. 13

[González-Buitrago and González, 2006] González-Buitrago, J. M. and González, C. (2006). Present and future of the autoimmunity laboratory. *Clinica chimica acta*, 365(1):50–57. 1, 10

[Gragnaniello et al., 2014] Gragnaniello, D., Sansone, C., and Verdoliva, L. (2014). Biologically-inspired dense local descriptor for indirect immunofluorescence image classification. In *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*, pages 1–5. 84, 97, 123, 124

[Griffiths and Ghahramani, 2011] Griffiths, T. L. and Ghahramani, Z. (2011). The indian buffet process: An introduction and review. *The Journal of Machine Learning Research*, 12:1185–1224. 59, 61

[Guo et al., 2016] Guo, R., Liu, L., Wang, W., Taalimi, A., Zhang, C., and Qi, H. (2016). Deep tree-structured face: A unified representation for multi-task facial biometrics. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. 16

[Han et al., 2014] Han, X.-H., Wang, J., Xu, G., and Chen, Y.-W. (2014). High-order statistics of microtexton for hep-2 staining pattern classification. *Biomedical Engineering, IEEE Transactions on*, 61(8):2223–2234. xi, xiv, xv, 12, 49, 51, 64, 65, 66, 81, 84, 93, 95, 96, 97, 98, 100, 119, 123

[Harris and Stephens, 1988] Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK. 31

[Hiemann et al., 2007] Hiemann, R., Hilger, N., Michel, J., Nitschke, J., Boehm, A., Anderer, U., Weigert, M., and Sack, U. (2007). Automatic analysis of

immunofluorescence patterns of hep-2 cells. *Annals of the New York Academy of Sciences*, 1109(1):358–371. 11

[Hiemann et al., 2006] Hiemann, R., Hilger, N., Sack, U., and Weigert, M. (2006). Objective quality evaluation of fluorescence images to optimize automatic image acquisition. *Cytometry Part A*, 69(3):182–184. 10

[Huang et al., 2012] Huang, Y.-C., Hsieh, T.-Y., Chang, C.-Y., Cheng, W.-T., Lin, Y.-C., and Huang, Y.-L. (2012). Hep-2 cell images classification based on textural and statistic features using self-organizing map. In *Intelligent Information and Database Systems*, pages 529–538. Springer. 11

[Huang et al., 2008a] Huang, Y.-L., Chung, C.-W., Hsieh, T.-Y., and Jao, Y.-L. (2008a). Outline detection for the hep-2 cell in indirect immunofluorescence images using watershed segmentation. In *Sensor Networks, Ubiquitous and Trustworthy Computing, 2008. SUTC'08. IEEE International Conference on*, pages 423–427. IEEE. 11

[Huang et al., 2008b] Huang, Y.-L., Jao, Y.-L., Hsieh, T.-Y., and Chung, C.-W. (2008b). Adaptive automatic segmentation of hep-2 cells in indirect immunofluorescence images. In *Sensor Networks, Ubiquitous and Trustworthy Computing, 2008. SUTC'08. IEEE International Conference on*, pages 418–422. IEEE. 11

[Jenatton et al., 2010] Jenatton, R., Mairal, J., Bach, F. R., and Obozinski, G. R. (2010). Proximal methods for sparse hierarchical dictionary learning. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 487–494. 113

[Jiang et al., 2013a] Jiang, Z., Lin, Z., and Davis, L. (2013a). Label consistent k-svd: Learning a discriminative dictionary for recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(11):2651–2664. 103

[Jiang et al., 2013b] Jiang, Z., Lin, Z., and Davis, L. S. (2013b). Label consistent k-svd: learning a discriminative dictionary for recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(11):2651–2664. 17, 104

[Juan and Gwun, 2009] Juan, L. and Gwun, O. (2009). A comparison of sift, pca-sift and surf. *International Journal of Image Processing (IJIP)*, 3(4):143–152. 31

[Khorsandi et al., 2015a] Khorsandi, R., Taalimi, A., and Abdel-Mottaleb, M. (2015a). Robust biometrics recognition using joint weighted dictionary learning and smoothed l0 norm. In *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*, pages 1–6. 17

[Khorsandi et al., 2015b] Khorsandi, R., Taalimi, A., Abdel-Mottaleb, M., and Qi, H. (2015b). Joint weighted dictionary learning and classifier training for robust biometric recognition. In *Global Conference on Signal and Information Processing (GlobalSIP), 2015 IEEE*. 87

[Knowles et al., 2011] Knowles, D., Ghahramani, Z., et al. (2011). Nonparametric bayesian sparse factor models with application to gene expression modeling. *The Annals of Applied Statistics*, 5(2B):1534–1552. 57, 59, 62

[Kong et al., 2014] Kong, X., Li, K., Cao, J., Yang, Q., and Wenyin, L. (2014). Hep-2 cell pattern classification with discriminative dictionary learning. *Pattern Recognition*, 47(7):2379 – 2388. 17, 46

[Larsen et al., 2014] Larsen, A., Vestergaard, J., and Larsen, R. (2014). Hep-2 cell classification using shape index histograms with donut-shaped spatial pooling. *Medical Imaging, IEEE Transactions on*, 33(7):1573–1580. 18, 84, 92, 97, 123

[Lazebnik et al., 2006] Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, volume 2, pages 2169–2178. IEEE. 25, 69, 105, 115

[Lee et al., 2006] Lee, H., Battle, A., Raina, R., and Ng, A. Y. (2006). Efficient sparse coding algorithms. In *Advances in neural information processing systems*, pages 801–808. 62

[Levinshtein et al., 2009] Levinshtein, A., Stere, A., Kutulakos, K., Fleet, D., Dickinson, S., and Siddiqi, K. (2009). Turbopixels: Fast superpixels using geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2290–2297. 89

[Li et al., 2012] Li, K., Yin, J., Lu, Z., Kong, X., Zhang, R., and Liu, W. (2012). Multiclass boosting svm using different texture features in hep-2 cell staining pattern classification. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 170–173. 14

[Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110. 31

[Lucchi et al., 2010] Lucchi, A., Smith, K., Achanta, R., Lepetit, V., and Fua, P. (2010). A fully automated approach to segmentation of irregularly shaped cellular structures in em images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 463–471. Springer. 87

[Mairal et al., 2012] Mairal, J., Bach, F., and Ponce, J. (2012). Task-driven dictionary learning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(4):791–804. 17, 104

[Mairal et al., 2009] Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009). Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696. ACM. 110, 121

[Mairal et al., 2010a] Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2010a). Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.*, 11:19–60. 110, 114

[Mairal et al., 2008] Mairal, J., Elad, M., and Sapiro, G. (2008). Sparse representation for color image restoration. *Image Processing, IEEE Transactions on*, 17(1):53–69. 110

[Mairal et al., 2010b] Mairal, J., Jenatton, R., Bach, F. R., and Obozinski, G. R. (2010b). Network flow algorithms for structured sparsity. In Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R., and Culotta, A., editors, *Advances*

in Neural Information Processing Systems 23, pages 1558–1566. Curran Associates, Inc. 113

[Mallat and Zhang, 1993] Mallat, S. G. and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415. 23

[Manivannan et al., 2014a] Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., and McKenna, S. (2014a). Hep-2 cell classification using multi-resolution local patterns and ensemble svms. In *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*, pages 37–40. 17, 18, 84, 92, 97

[Manivannan et al., 2014b] Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., and McKenna, S. (2014b). Hep-2 specimen classification using multi-resolution local patterns and svm. In *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*, pages 41–44. 97

[Manivannan et al., 2014c] Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., and McKenna, S. J. (2014c). Hep-2 cell classification using multi-resolution local patterns and ensemble svms. In *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*, pages 37–40. IEEE. 122, 123, 124

[Marée et al., 2013] Marée, R., Wehenkel, L., and Geurts, P. (2013). Extremely randomized trees and random subwindows for image classification, annotation,

and retrieval. In Criminisi, A. and Shotton, J., editors, *Decision Forests for Computer Vision and Medical Image Analysis*, Advances in Computer Vision and Pattern Recognition, pages 125–141. Springer London. 13

[Mian et al., 1991] Mian, I., Bradwell, A. R., and Olson, A. J. (1991). Structure, function and properties of antibody binding sites. *Journal of Molecular Biology*, 217(1):133 – 151. 8

[Minaee et al., 2015] Minaee, S., Abdolrashidi, A., and Wang, Y. (2015). Iris recognition using scattering transform and textural features. In *Signal Processing and Signal Processing Education Workshop (SP/SPE), 2015 IEEE*, pages 37–42. 104

[Monajemi et al., 2014] Monajemi, S., Sanei, S., and Ong, S.-H. (2014). Advances in bacteria motility modelling via diffusion adaptation. In *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, pages 2335–2339. 74

[Monajemi et al., 2015] Monajemi, S., Sanei, S., Ong, S.-H., and Sayed, A. H. (2015). Adaptive regularized diffusion adaptation over networks. In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE. 76, 79

[Nedic and Bertsekas, 2001] Nedic, A. and Bertsekas, D. P. (2001). Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization*, 12(1):109–138. 74

[Nedic and Ozdaglar, 2009] Nedic, A. and Ozdaglar, A. (2009). Distributed subgradient methods for multi-agent optimization. *Automatic Control, IEEE Transactions on*, 54(1):48–61. 74

[Nguyen et al., 2011] Nguyen, N., Nasrabadi, N., and Tran, T. (2011). Robust multi-sensor classification via joint sparse representation. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, pages 1–8. 109

[Nosaka and Fukui, 2014] Nosaka, R. and Fukui, K. (2014). Hep-2 cell classification using rotation invariant co-occurrence among local binary patterns. *Pattern Recognition*, 47(7):2428–2436. 14, 15, 45, 46, 64, 82, 120, 121

[Ojala et al., 1996] Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51 – 59. 15

[Otsu, 1975] Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27. 11

[Parikh and Boyd, 2013] Parikh, N. and Boyd, S. (2013). Proximal algorithms. *Foundations and Trends in optimization*, 1(3):123–231. 111

[Patel and Chellappa, 2011] Patel, V. M. and Chellappa, R. (2011). Sparse representations, compressive sensing and dictionaries for pattern recognition. In *Pattern Recognition (ACPR), 2011 First Asian Conference on*, pages 325–329. IEEE. 23

[Pati et al., 1993] Pati, Y. C., Rezaiifar, R., and Krishnaprasad, P. S. (1993). Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44 vol.1. 23

[Perner et al., 2002] Perner, P., Perner, H., and Müller, B. (2002). Mining knowledge for hep-2 cell image classification. *Artificial intelligence in medicine*, 26(1):161–173. 11

[Philbin et al., 2008] Philbin, J., Chum, O., Isard, M., Sivic, J., and Zisserman, A. (2008). Lost in quantization: Improving particular object retrieval in large scale image databases. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 1–8. IEEE. 21

[Ponomarev et al., 2014] Ponomarev, G. V., Arlazarov, V. L., Gelfand, M. S., and Kazanov, M. D. (2014). Ana hep-2 cells image classification using number, size, shape and localization of targeted cell regions. *Pattern Recognition*, 47(7):2360–2366. 13, 46

[Qi et al., 2016] Qi, X., Zhao, G., Chen, J., and Pietikäinen, M. (2016). Exploring illumination robust descriptors for human epithelial type 2 cell classification. *Pattern Recognition*, page in Press. 132

[Ramirez et al., 2010] Ramirez, I., Sprechmann, P., and Sapiro, G. (2010). Classification and clustering via dictionary learning with structured incoherence and

shared features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3501–3508. 103

[Rattani et al., 2007] Rattani, A., Kisku, D., Bicego, M., and Tistarelli, M. (2007). Feature level fusion of face and fingerprint biometrics. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–6. 104

[Rehn and Sommer, 2007] Rehn, M. and Sommer, F. T. (2007). A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *Journal of computational neuroscience*, 22(2):135–146. 3, 125

[Rosasco et al., 2009] Rosasco, L., Verri, A., Santoro, M., Mosci, S., and Villa, S. (2009). Iterative projection methods for structured sparsity regularization. *Technical Report MIT-CSAIL-TR-2009- 050 CBCL-282*. 113

[Ruta and Gabrys, 2000] Ruta, D. and Gabrys, B. (2000). An overview of classifier fusion methods. *Computing and Information systems*, 7(1):1–10. 103

[Sack et al., 2003] Sack, U., Knoechner, S., Warschkau, H., Pigla, U., Emmrich, F., and Kamprad, M. (2003). Computer-assisted classification of hep-2 immunofluorescence patterns in autoimmune diagnostics. *Autoimmunity Reviews*, 2(5):298–304. 11

[Sani and Vosoughi, 2014] Sani, A. and Vosoughi, A. (2014). Resource allocation optimization for distributed vector estimation with digital transmission. In *2014*

48th Asilomar Conference on Signals, Systems and Computers, pages 1463–1467. 110

[Sayed, 2014] Sayed, A. H. (2014). Adaptation, learning, and optimization over networks. *Foundations and Trends in Machine Learning*, 7(4-5):311–801. 69, 74, 75

[Shekhar et al., 2014] Shekhar, S., Patel, V., Nasrabadi, N., and Chellappa, R. (2014). Joint sparse representation for robust multimodal biometrics recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(1):113–126. 103, 104

[Shen et al., 2014] Shen, L., Lin, J., Wu, S., and Yu, S. (2014). Hep-2 image classification using intensity order pooling based features and bag of words. *Pattern Recognition*, 47(7):2419–2427. 46, 58, 64

[Shi and Tomasi, 1994] Shi, J. and Tomasi, C. (1994). Good features to track. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 593–600. IEEE. 31

[Snell et al., 2012] Snell, V., Christmas, W., and Kittler, J. (2012). Texture and shape in fluorescence pattern identification for auto-immune disease diagnosis. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3750–3753. 13

[Soda and Iannello, 2006] Soda, P. and Iannello, G. (2006). A multi-expert system to classify fluorescent intensity in antinuclear autoantibodies testing. In

Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on, pages 219–224. IEEE. 11

[Soda and Iannello, 2009] Soda, P. and Iannello, G. (2009). Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis. *Information Technology in Biomedicine, IEEE Transactions on*, 13(3):322–329. 11

[Soda et al., 2009] Soda, P., Iannello, G., and Vento, M. (2009). A multiple expert system for classifying fluorescent intensity in antinuclear autoantibodies analysis. *Pattern Analysis and Applications*, 12(3):215–226. 11

[Soda et al., 2006] Soda, P., Rigon, A., Afeltra, A., and Iannello, G. (2006). Automatic acquisition of immunofluorescence images: Algorithms and evaluation. In *Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on*, pages 386–390. IEEE. 10

[Sriram et al., 2014] Sriram, A., Ensafi, S., Roohi, S. F., and Kassim, A. A. (2014). Classification of human epithelial type-2 cells using hierarchical segregation. In *Control Automation Robotics Vision (ICARCV), 2014 13th International Conference on*, pages 323–328. 64

[Stoklasa et al., 2014] Stoklasa, R., Majtner, T., and Svoboda, D. (2014). Efficient k-nn based hep-2 cells classifier. *Pattern Recognition*, 47(7):2409 – 2418. 14, 46

[Storch, 2000] Storch, W. B. (2000). *Immunofluorescence in clinical immunology: a primer and atlas*. Springer. 10

[Strandmark et al., 2012] Strandmark, P., Ulen, J., and Kahl, F. (2012). Hep-2 staining pattern classification. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 33–36. 13

[Taalimi et al., 2015a] Taalimi, A., Ensafi, S., Qi, H., Lu, S., Kassim, A. A., and Tan, C. L. (2015a). Multimodal dictionary learning and joint sparse representation for hep-2 cell classification. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, pages 308–315. Springer. 104

[Taalimi et al., 2015b] Taalimi, A., Khorsandi, R., and Qi, H. (2015b). Online multi-modal task-driven dictionary learning and robust joint sparse representation for visual tracking. In *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on.* 87

[Taalimi et al., 2015c] Taalimi, A., Khorsandi, S., and Qi, H. (2015c). Online multi-modal task-driven dictionary learning and robust joint sparse representation for visual tracking. *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 0:539–546. 110

[Taalimi and Qi, 2015] Taalimi, A. and Qi, H. (2015). Robust multi-object tracking using confident detections and safe tracklets. In *Image Processing (ICIP), 2015 IEEE International Conference on.* 110

[Theodorakopoulos et al., 2012] Theodorakopoulos, I., Kastaniotis, D., Economou, G., and Fotopoulos, S. (2012). Hep-2 cells classification via fusion of morphological and textural features. In *Bioinformatics &*

Bioengineering (BIBE), 2012 IEEE 12th International Conference on, pages 689–694. IEEE. 14

[Theodorakopoulos et al., 2014] Theodorakopoulos, I., Kastaniotis, D., Economou, G., and Fotopoulos, S. (2014). Hep-2 cells classification via sparse representation of textural features fused into dissimilarity space. *Pattern Recognition*, 47(7):2367 – 2378. 45, 46, 82, 120, 121

[Thibault and Angulo, 2012] Thibault, G. and Angulo, J. (2012). Efficient statistical/morphological cell texture characterization and classification. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 2440–2443. 17, 18

[Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288. 22

[Tikhonov, 1963] Tikhonov, A. (1963). Solution of incorrectly formulated problems and the regularization method. In *Soviet Math. Dokl.*, volume 5, pages 1035–1038. 32

[Towfic et al., 2014] Towfic, Z. J., Chen, J., and Sayed, A. H. (2014). Dictionary learning over large distributed models via dual-admm strategies. In *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*, pages 1–6. IEEE. 77

[Tropp, 2004] Tropp, J. A. (2004). Greed is good: algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242. 23

[Tu and Sayed, 2012] Tu, S.-Y. and Sayed, A. H. (2012). Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks. *Signal Processing, IEEE Transactions on*, 60(12):6217–6234. 69, 74

[Wang et al., 2014] Wang, W., He, L., Markham, P., Qi, H., Liu, Y., Cao, Q., and Tolbert, L. (2014). Multiple event detection and recognition through sparse unmixing for high-resolution situational awareness in power grid. *IEEE Trans. on Smart Grid*, 5(4):1654–1664. 87

[Wang et al., 2013] Wang, W., Luo, J., and Qi, H. (2013). Action recognition across cameras via reconstructable paths. In *Distributed Smart Cameras (ICDSC), 2013 Seventh International Conference on*, pages 1–6. 87

[Wang et al., 2016] Wang, W., Taalimi, A., Duan, K., Guo, R., and Qi, H. (2016). Learning patch dependent random kernel forest for person re-identification. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. 86

[Wiliem et al., 2014] Wiliem, A., Sanderson, C., Wong, Y., Hobson, P., Minchin, R. F., and Lovell, B. C. (2014). Automatic classification of human epithelial type 2 cell indirect immunofluorescence images using cell pyramid matching. *Pattern Recognition*, 47(7):2315–2324. 46, 58, 64

[Wiliem et al., 2013] Wiliem, A., Wong, Y., Sanderson, C., Hobson, P., Chen, S., and Lovell, B. C. (2013). Classification of human epithelial type 2 cell indirect immunofluoresence images via codebook based descriptors. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 95–102. 18

[Wright et al., 2009] Wright, J., Yang, A., Ganesh, A., Sastry, S., and Ma, Y. (2009). Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227. 107

[Xiao and Boyd, 2004] Xiao, L. and Boyd, S. (2004). Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78. 74

[Yang et al., 2009] Yang, J., Yu, K., Gong, Y., and Huang, T. (2009). Linear spatial pyramid matching using sparse coding for image classification. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 1794–1801. IEEE. 21, 37

[Yang et al., 2014a] Yang, M., Dai, D., Shen, L., and Van Gool, L. (2014a). Latent dictionary learning for sparse representation based classification. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 4138–4145. 103

[Yang et al., 2010] Yang, M., Zhang, L., Yang, J., and Zhang, D. (2010). Metaface learning for sparse representation based face recognition. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1601–1604. 110

[Yang et al., 2014b] Yang, Y., Wiliem, A., Alavi, A., Lovell, B. C., and Hobson, P. (2014b). Visual learning and classification of human epithelial type 2 cell

images through spontaneous activity patterns. *Pattern Recognition*, 47(7):2325 – 2337. 13

[Zhang et al., 2010] Zhang, Y., Jin, R., and Zhou, Z.-H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52. 32

[Zhao and Sayed, 2012] Zhao, X. and Sayed, A. H. (2012). Clustering via diffusion adaptation over networks. In *Cognitive Information Processing (CIP), 2012 3rd International Workshop on*, pages 1–6. IEEE. 79

[Zonoobi and Kassim, 2013] Zonoobi, D. and Kassim, A. A. (2013). On the reconstruction of sequences of sparse signals–the weighted-cs. *Journal of Visual Communication and Image Representation*, 24(2):196–202. 62

[Zonoobi and Kassim, 2014] Zonoobi, D. and Kassim, A. A. (2014). A computationally efficient method for reconstructing sequences of MR images from undersampled k-space data. *Medical image analysis*, 18(6):857–865. 87

[Zonoobi et al., 2011] Zonoobi, D., Kassim, A. A., and Venkatesh, Y. V. (2011). Gini index as sparsity measure for signal reconstruction from compressive samples. *Selected Topics in Signal Processing, IEEE Journal of*, 5(5):927–932. 22, 34

[Zonoobi et al., 2014a] Zonoobi, D., Roohi, S. F., and Kassim, A. A. (2014a). Dependent nonparametric bayesian group dictionary learning for online reconstruction of dynamic mr images. *arXiv preprint arXiv:1408.5667*. 57

[Zonoobi et al., 2014b] Zonoobi, D., Roohi, S. F., and Kassim, A. A. (2014b). Low-rank and sparse matrix decomposition with a-priori knowledge for dynamic 3d mri reconstruction. *arXiv preprint arXiv:1411.6206*. 57

[Zou and Hastie, 2005] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2):301–320. 121, 123

List of Publications

The contents of this thesis are based on the following papers that have been published, accepted, or submitted to the peer-reviewed journals and conferences.

Journal Publications

[1] S. Ensafi, S. Lu, A. A. Kassim, and C. L. Tan, "Accurate HEp-2 Cell Classification Based on Sparse Coding of Superpixels," *Pattern Recognition Letters*, 2016 (In Press).

[2] S. Ensafi, A. Taalimi, S. Lu, A. A. Kassim, H. Qi, and C. L. Tan, "Joint Multi-Cue Dictionary Learning for HEp-2 Cell Classification," *Pattern Recognition*, 2015 (under review).

[3] S. Ensafi, S. Monajemi, S. Lu, A. A. Kassim, C. L. Tan, S. Sanei, and S-H. Ong, "Adaptive Distributed Dictionary Learning for HEp-2 Cell Classification," *Artificial Intelligence in Medicine*, 2016 (under review).

[4] S. Ensafi, S. Lu, A. A. Kassim, and C. L. Tan, "Accurate HEp-2 Cell Classification Based on Sparse Bag of Words Coding," *Computerized Medical Imaging and Graphics*, 2015 (under review).

Conference Publications

[1] A. Taalimi, S. Ensafi, H. Qi, S. Lu, A. A. Kassim, and C. L. Tan, "Multimodal dictionary learning and joint sparse representation for hep-2 cell classification," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 308 - 315, Springer, 2015.

[2] S. Ensafi, S. Lu, A. Kassim, and C. L. Tan, "Sparse non-parametric bayesian model for hep-2 cell image classification," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on,* pp. 679 - 682, April 2015.

[3] S. Ensafi, S. Lu, A. A. Kassim, and C. L. Tan, "Automatic cad system for hep-2 cell image classification," in *Pattern Recognition (ICPR), 2014 22nd International Conference on,* pp. 3321 - 3326, IEEE, 2014.

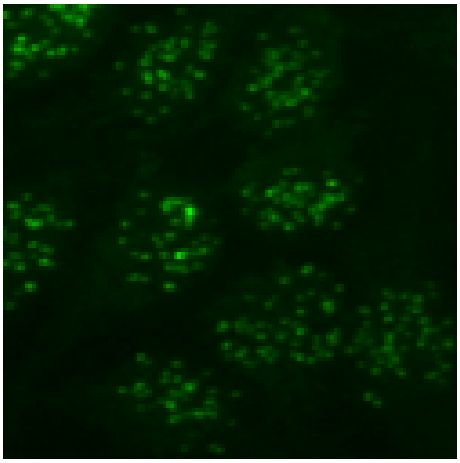
[4] S. Ensafi, S. Lu, A. A. Kassim, and C. L. Tan, "A bag of words based approach for classification of hep-2 cell images," *in Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on,* pp. 29 - 32, IEEE, 2014.

[5] S. Monajemi, S. Ensafi, S. Lu, A. Kassim, C. L. Tan, S. Sanei, and S-H. Ong "Classification of HEp-2 Cells Using Distributed Dictionary Learning," in *Signal Processing (EUSIPCO) 2016 24th European.* IEEE. (Accepted)

[6] A. Siram, S. Ensafi, S. Faghihroohi, and A. A. Kassim, "Classification of human epithelial type-2 cells using hierarchical segregation," in *Control Automation Robotics Vision (ICARCV), 2014 13th International Conference on,* pp. 323 - 328, IEEE, 2014.

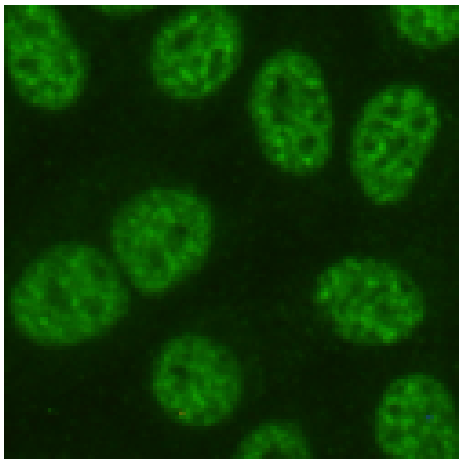
Appendix

ICPR2012 Dataset



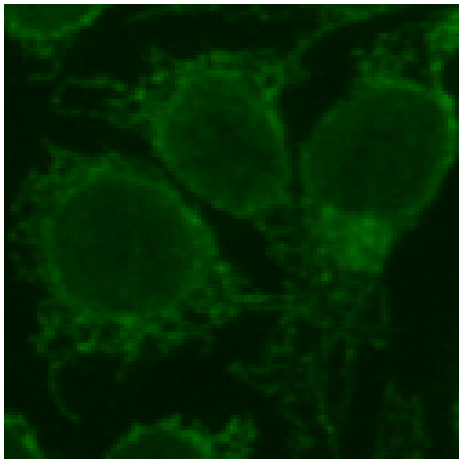
Centromere (Ce): Contains several discrete small spots that are scattered throughout the nuclei area. These speckled can be observed in the nuclear chromatin.

Related AD: CREST Syndrome (calcinosis, Raynaud's, esophageal dysfunction, sclerodactyly and telangiectasia).



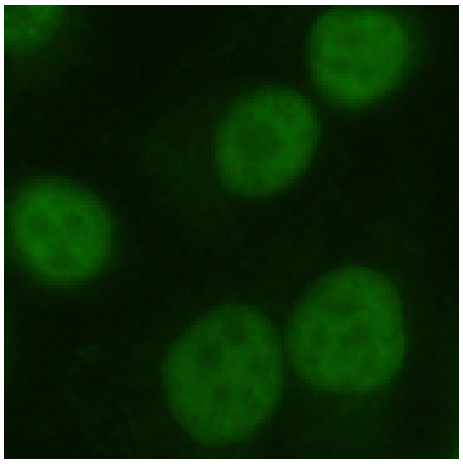
Coarse Speckled (Cs): Contains coarse granular nuclear staining pattern. Some holes are observed on the nuclei area and the pattern is so close to Fine Speckled cell images.

Related AD: Titers>1:160 suggest SLE (Anti-Sm) or mixed connective tissue disease (MCTD) (Anti-RNP).



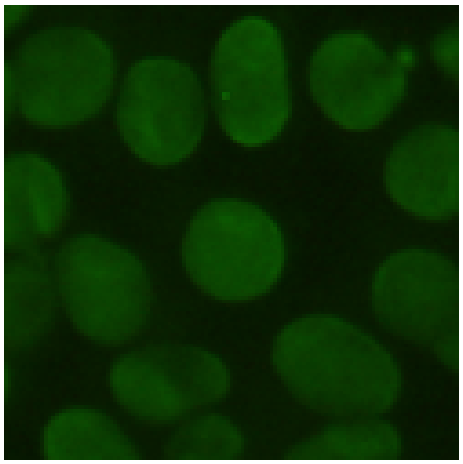
Cytoplasmatic (Cy): contains a very fine dense resembling homogeneous staining which covering part or the cytoplasm.

Related AD: Alcoholic liver disease, rheumatoid arthritis, psoriasis and normal individuals.



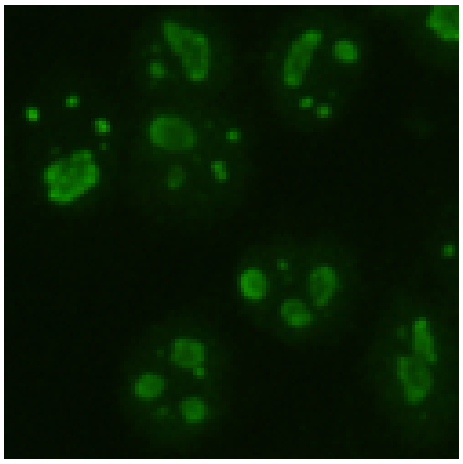
Fine Speckled (Fs): Contains fine granular nuclear staining pattern. In most of the cell images of this type, the patterns are observed blurry.

Related AD: Sjögren's (Anti-SSB), Sjögren's sicca complex (Anti-SSA), SLE, sub-acute cutaneous lupus erythematosus and scleroderma.



Homogeneous (H): The staining pattern is spread out over the interphase nuclei. These type of cell images have smooth diffused characteristics all over the nuclei area.

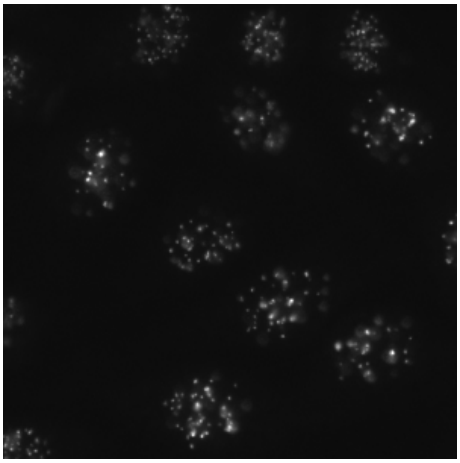
Related AD: Titers>1:160 are highly suggestive of systemic lupus erythematosus (SLE) or other connective tissue diseases.

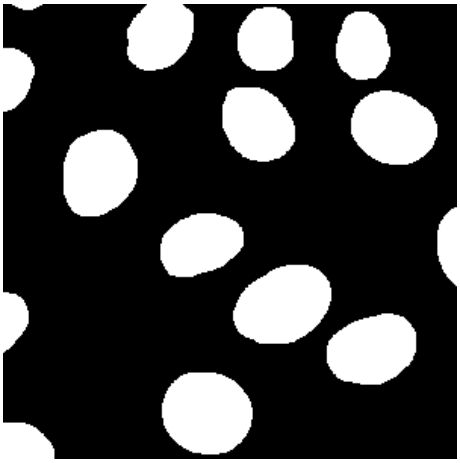


Nucleolar (N): Small compact particles can be observed in the cells' nucleoli. Without these granules, this pattern is close to Homogeneous class. At most six granules can be observed.

Related AD: Polymyositis/scleroderma, polymyositis and scleroderma.

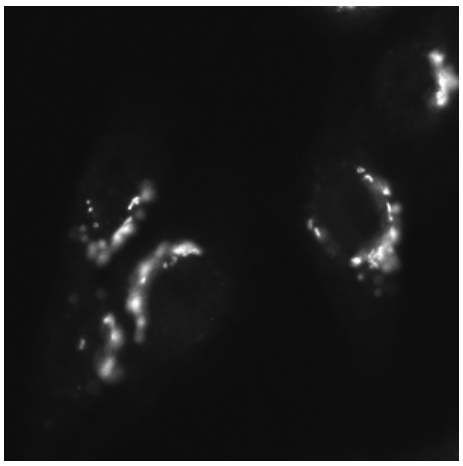
ICIP2013 Dataset

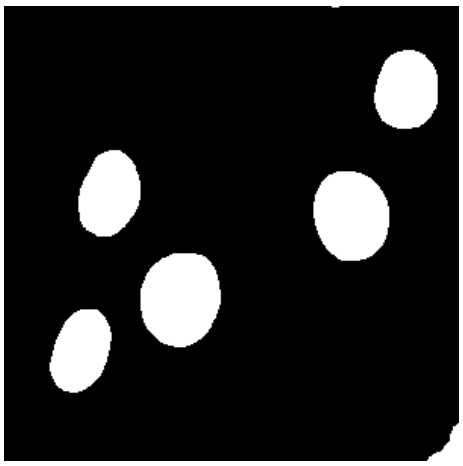




Centromere (Ce): Discrete small spots that are observed in the nuclear chromatin in a scattering fashion.

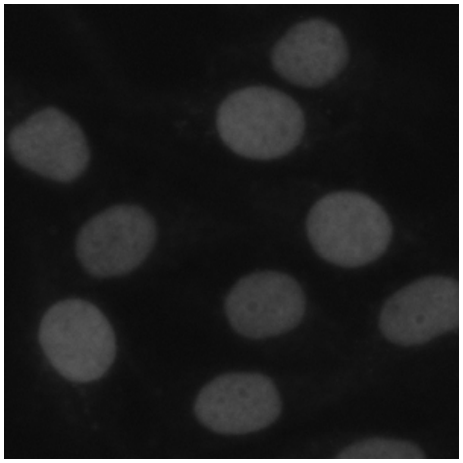
Related AD: CREST Syndrome (calcinosis, Raynaud's, esophageal dysfunction, sclerodactyly and telangiectasia).

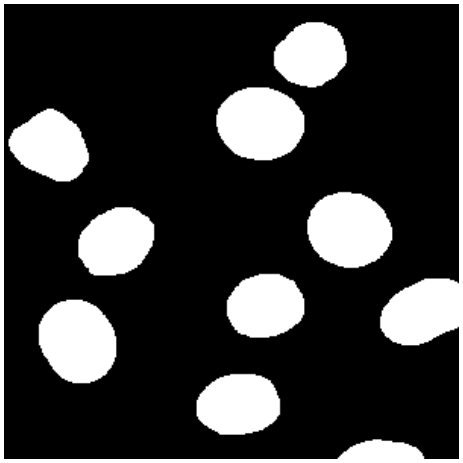




Golgi (G): Composed of irregular large granules which are stained adjacent to the boundaries of nucleus and around chromosomal material.

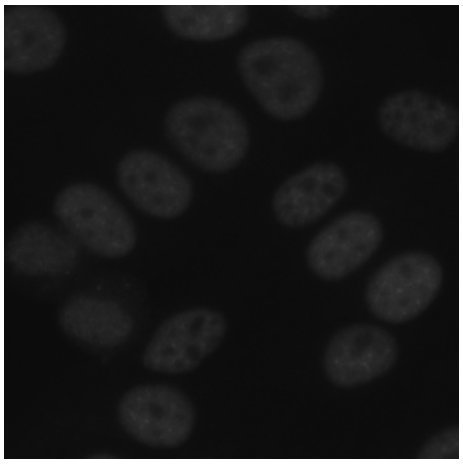
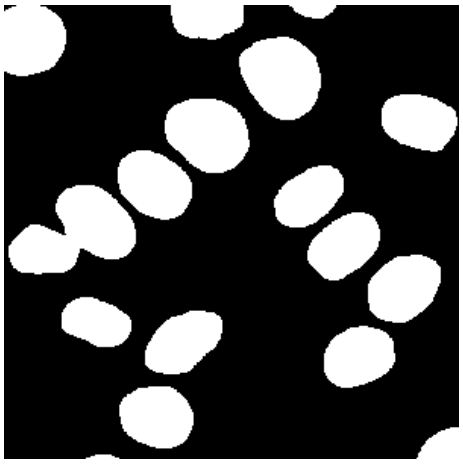
Related AD: SLE and Sjögren's syndrome.





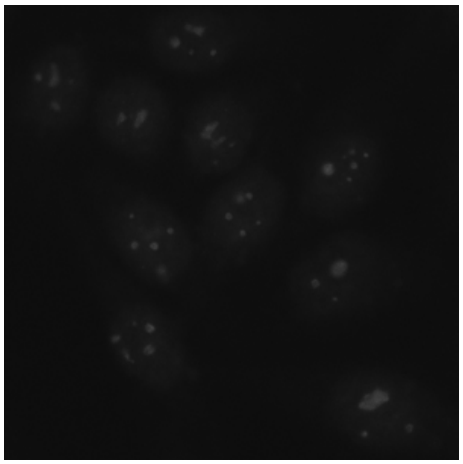
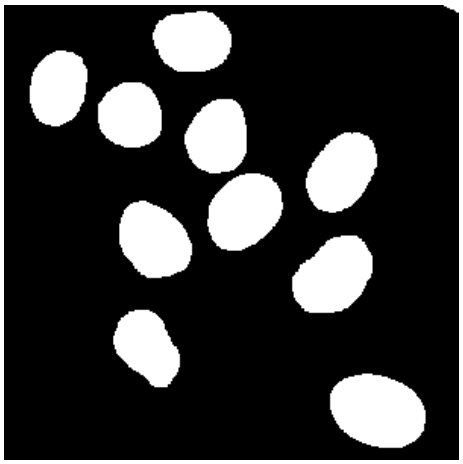
Homogeneous (H): The staining pattern is spread out over the interphase nuclei. These type of cell images have smooth diffused characteristics all over the nuclei area.

Related AD: Titers>1:160 are highly suggestive of systemic lupus erythematosus (SLE) or other connective tissue diseases.

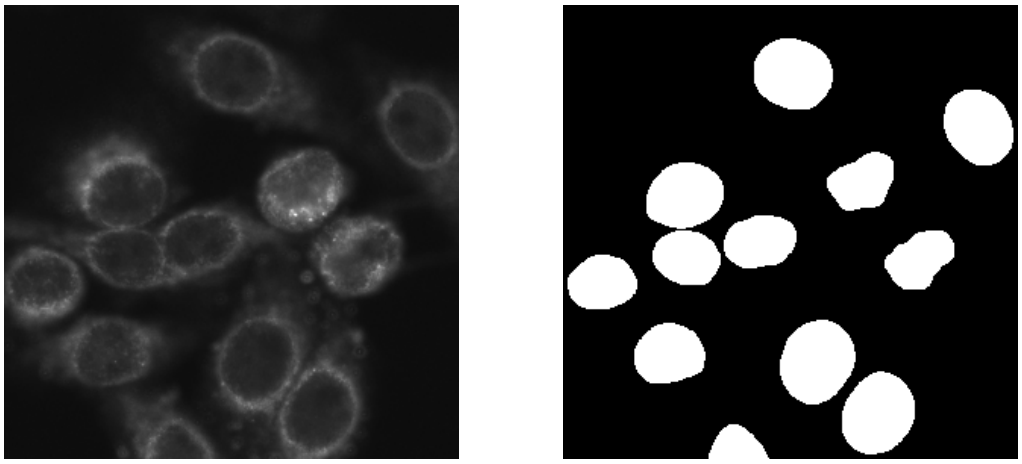
Mitotic Spindle (MP): staining only of the triangular or "banana-shaped" pole area of the mitotic spindle in the metaphase cells. This pattern is rare.

Related AD: Infectious mononucleosis, Hashimotos's disease, thyrotoxicosis and other chronic diseases.

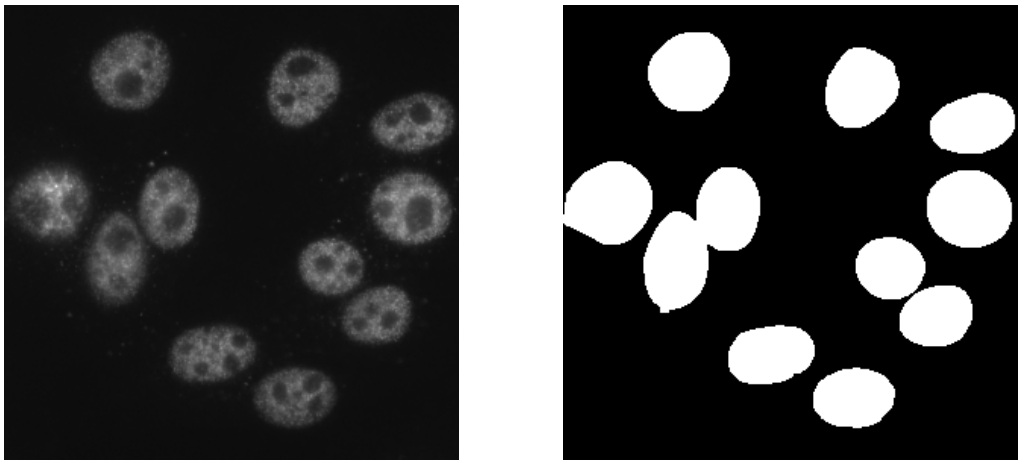
Nucleolar (N): Clustered particles can be observed in the cells' nucleoli. Without these granules, this pattern is close to Homogeneous class. At most six granules can be observed.

Related AD: Polymyositis/scleroderma, polymyositis and scleroderma.



Nuclear Membrane (NM): Contains a tube-like of smooth homogeneous fluorescence in the interphase cells.

Related AD: Smooth staining of the entire nuclear membrane with cytoplasmic staining of the mitochondrial organelles.



Speckled (S): Contains two subcategories of fine- and coarse-speckled. Various sized speckled can be observed densely distributed throughout nucleoplasm.

Related AD: Systemic Sclerosis with diffuse cutaneous involvement or scleroderma.