Efficient feature extraction and classification for staining patterns of HEP-2 Cells

Xu, Xiang

2016

Xu, X. (2016). Efficient feature extraction and classification for staining patterns of HEP-2 Cells. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/66017>

<https://doi.org/10.32657/10356/66017>

EFFICIENT FEATURE EXTRACTION AND CLASSIFICATION FOR STAINING PATTERNS OF HEP-2 CELLS

XU XIANG

School of Computer Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2016

Abstract

The occurrence of antinuclear antibodies (ANAs) in patient serum has significant relation to autoimmune diseases. The ANAs detection can be accomplished via indirect immunofluorescence (IIF) technique using human epithelial (HEp-2) cell as substrate in laboratory. Identification of IIF slide images is based on human visual inspection, but it suffers from serious problems due to subjective evaluation. Therefore, Computer-Aided Diagnosis (CAD) system for supporting doctors' diagnosis is essential.

In this thesis, our aim is to develop novel methods to automatically classify positive staining patterns of HEp-2 cells. First, we research on Bag-of-Words (BoW) framework which is one of the most successful image representations. To reduce the inevitable information loss caused by coding process, we propose a Linear Local Distance Coding (LLDC) method. The proposed LLDC method transforms original local feature to more discriminative local distance vector by searching for local neighbors of the local feature in the class-specific manifolds. Then we encode and pool the local distance vectors to get salient image representation. Combined with the traditional coding methods, our proposed method achieves higher classification accuracy.

Secondly, we propose a rotation invariant textural feature of Pairwise Local Ternary Patterns with Spatial Rotation Invariant (PLTP-SRI). It is invariant to image rotations, meanwhile it is robust to noise and weak illumination. By adding spatial pyramid structure, our proposed method captures spatial layout information. While the proposed PLTP-SRI feature extracts local feature, the BoW framework builds a global image representation. It is reasonable to combine them together to achieve impressive classification performance, as the combined feature takes the advantages of the two kinds of features in different aspects.

Finally, we design a Co-occurrence Differential Texton (CoDT) feature to represent the local image patches of HEp-2 cells. The CoDT feature reduces the information loss by ignoring the quantization while it utilizes the spatial relations among the differential micro-texton feature. Thus it can increase the discriminative power. We build a generative model to adaptively characterize the CoDT feature space of the training data. Furthermore, we exploit a discriminant representation for the HEp-2 cell images based on the adaptive partitioned feature space. Therefore, the resulting representation is adapted to the classification task. By cooperating with linear Support Vector Machine (SVM) classifier, our proposed framework can exploit the advantages of both generative and discriminative approaches for image classification.

Throughout, we provide evaluations for our proposed methods on two publicly available HEp-2 cells datasets: ICPR2012 dataset from the ICPR'12 *HEp-2 cell classification contest* and ICIP2013 training dataset from the ICIP'13 *Competition on cells classification by fluorescent image analysis*.

Acknowledgments

I am above all grateful to Prof. Lin Feng for providing me patient and insightful guidance during my years as his student. I would express my wholehearted gratitude towards Dr Khai Pang Leong and Carol Ng from Tan Tock Seng Hospital for their assistance. I also thank my family and friends for encouraging me and bearing with me during all these years. Finally, I would like to thank Nanyang Technological University for providing me with an opportunity and platform to pursue higher achievements.

This work was partially supported by three research grants, M408020000 from Nanyang Technological University, MOE2011-T2-2-037 and RG139/14 from Ministry of Education, Singapore.

Contents

Abstract	ii
Acknowledgments	iv
List of Figures	viii
List of Tables	xi
1 Introduction	1
1.1 Background	1
1.2 Motivation and Objective	6
1.3 Experimental Datasets	9
1.3.1 The ICPR2012 Dataset	10
1.3.2 The ICIP2013 Training Dataset	11
1.4 Contributions and Structure of the Thesis	12
2 Related Work	15
2.1 Staining Pattern of HEp-2 Cells	15
2.2 Feature Extraction	18
2.2.1 Low-level Features	19
2.2.1.1 Local Binary Patterns	19
2.2.1.2 Scale Invariant Feature Transform	23
2.2.2 Mid-level Features	28

2.3	Classification	29
2.3.1	Support Vector Machine	29
2.3.2	Nearest Neighbor Classifier	30
3	Linear Local Distance Coding Method	32
3.1	Introduction	32
3.2	Proposed Method	35
3.2.1	Bag-of-Words Representation	35
3.2.1.1	Coding	36
3.2.1.2	Pooling	39
3.2.2	Distance Vector	40
3.2.3	Local Distance Vector	41
3.2.4	The Algorithm Framework	43
3.3	Experiments and Analyses	45
3.3.1	Experiment Setup	45
3.3.2	Experimental Results on the ICPR2012 Dataset	47
3.3.3	Experimental Results on the ICIP2013 Training Dataset	49
3.3.4	Discussion	51
3.4	Summary	53
4	Encoding Rotation Invariant Features in Staining Patterns	56
4.1	Introduction	56
4.2	Proposed Method	58
4.2.1	Pairwise LTPs with Spatial Rotation Invariant	59
4.2.2	Encoding the SIFT Features with BoW Framework	63

4.3	Experiments and Analyses	64
4.3.1	Experiment Setup	64
4.3.2	Experimental Results on the ICPR2012 Dataset	65
4.3.3	Experimental Results on the ICIP2013 Training Dataset	67
4.3.4	Discussion	69
4.4	Summary	72
5	Adaptive Co-occurrence Differential Texton Space for Staining Pattern Classification	73
5.1	Introduction	73
5.2	Proposed Method	75
5.2.1	Co-occurrence Differential Texton	75
5.2.2	Adaptive CoDT Feature Space	79
5.2.3	HEp-2 Cell Image Representation in the Adaptive CoDT Feature Space	80
5.3	Experiments and Analyses	84
5.3.1	Experiment Setup	84
5.3.2	Experimental Results on the ICPR2012 Dataset	85
5.3.3	Experimental Results on the ICIP2013 Training Dataset	86
5.3.4	Discussion	89
5.4	Summary	90
6	Conclusion and Future Work	92
	References	96
	List of Publications	107

List of Figures

1.1	Patient with systemic lupus erythematosus (picture is from Google Images).	2
1.2	Patient with rheumatoid arthritis (picture is from Google Images).	2
1.3	IIF procedure.	4
1.4	Typical HEp-2 cells with different staining patterns.	6
1.5	Sample images from the ICPR2012 dataset and the ICIP2013 training dataset respectively. The rows named “positive” are the patterns with positive intensity, while the “intermediate” rows are the patterns with intermediate intensity.	8
1.6	Samples of the ICPR2012 dataset and the ICIP2013 training dataset with different staining patterns of HEp-2 cells.	11
2.1	Local binary patterns with different (P, R) s: $(4, 1)$, $(8, 1)$ and $(16, 2)$.	20
2.2	The different “uniform” patterns in the case of $P = 8$. The white circle denotes ‘1’ while the dark circle denotes ‘0’.	22
2.3	Construction of $D(x, y, \sigma)$. The scale space images are produced by convolving initial image incrementally with Gaussians. The DoG images are produced by subtracting adjacent Gaussian images.	24
2.4	Keypoints detected in an image. The start point of arrow is the keypoint’s location, the direction indicates the orientation of the local gradient at the keypoint and the length denotes the magnitude of the local gradient.	27
2.5	Keypoint descriptor.	28

3.1	Overview of the feature extraction-coding-pooling scheme.	37
3.2	Distance vector <i>vs.</i> local distance vector. $\mathbf{x}_i$ is a query local feature. Distance vector searches the mapping point $\mathbf{x}_i$ which is determined by the nearest few neighbors in each manifold M^c . Local distance vector retrieves only the local neighborhood in $M = [M^1, M^2, \dots, M^C]$	42
3.3	Overview of the image classification flowchart based on our proposed LLDC method.	44
3.4	Classification accuracy of the LLDC method under $k_{LDV} \in \{2, 5, 10, 15, 20, \dots, 40\}$ on ICPR201 dataset.	52
3.5	Classification accuracy of the LLDC method under $k_{LDV} \in \{2, 5, 10, 20, 30, \dots, 70\}$ on ICIP2013 training dataset.	53
3.6	Classification accuracy of the LLDC method using LLC strategy under $k_{LLC} \in \{2, 5, 10, 20, 30, \dots, 70\}$	54
3.7	Classification accuracy of the LLDC method using LSC strategy under $k_{LSC} \in \{2, 5, 10, 15, 20, \dots, 40\}$	55
4.1	HEp-2 slide images with positive and intermediate fluorescent intensity. .	58
4.2	Difference between LBP and LTP encoding procedures.	61
4.3	An example of the rotation equivalence class. Black and white circles correspond to ‘0’ and ‘1’ respectively. $s(\vartheta)$ is the start point of the binary sequence, where $s(\vartheta) = (x + R\cos(\vartheta), y + R\sin(\vartheta))$	62
4.4	Framework of pairwise LTPs with spatial rotation invariant.	63
4.5	Classification accuracy of BoW representation with various codebook sizes.	71
5.1	The LBP encoding procedure on a 3×3 microtexton.	76
5.2	Comparison between the LBPs for similar local structure and for different local structure.	77

5.3	Illustration of co-occurrence differential texton. (a) an example of differential vector. (b) two pairs of DVs with rotation angles 0° and θ respectively. The number of neighboring pixel of DV is set as $P = 8$	78
5.4	Classification performance of the AdaCoDT method under different number of GMM components.	90

List of Tables

1.1	Composition of the ICPR2012 dataset. Each table item represents the number of cells and the number of images which is in parentheses.	10
1.2	Composition of the ICIP2013 training dataset. Each table item represents the number of cells and the number of images which is in parentheses. . .	12
3.1	Classification performance on the ICPR2012 dataset.	48
3.2	The confusion matrix for the cell level classification on the ICPR2012 dataset.	49
3.3	The confusion matrix for the image level classification on the ICPR2012 dataset.	49
3.4	Classification performance on the ICIP2013 training dataset.	50
3.5	The confusion matrix for the cell level classification on the ICIP2013 training dataset.	51
3.6	The confusion matrix for the image level classification on the ICIP2013 training dataset.	51
4.1	Parameters for comparative algorithms.	65
4.2	Classification performance on the ICPR2012 dataset.	66
4.3	The confusion matrix for the cell level classification on the ICPR2012 dataset.	67
4.4	The confusion matrix for the image level classification on the ICPR2012 dataset.	67

4.5	Classification performance on the ICIP2013 training dataset.	68
4.6	The confusion matrix for the cell level classification on the ICIP2013 training dataset.	68
4.7	The confusion matrix for the image level classification on the ICIP2013 training dataset.	68
4.8	Classification accuracy of PLTP-SRI under various (R, d) s.	70
4.9	Classification accuracy of PLTP-SRI with different thresholds for ternary pattern calculation.	70
4.10	Classification accuracy of PLTP-SRI with different levels of spatial pyramid.	71
5.1	Parameters for comparative algorithms.	85
5.2	Classification performance on the ICPR2012 dataset.	85
5.3	The confusion matrix for the cell level classification on the ICPR2012 dataset.	86
5.4	The confusion matrix for the image level classification on the ICPR2012 dataset.	87
5.5	Classification performance on the ICIP2013 training dataset.	87
5.6	The confusion matrix for the cell level classification on the ICIP2013 training dataset.	88
5.7	The confusion matrix for the image level classification on the ICIP2013 training dataset.	88
5.8	Classification performance of the AdaCoDT method with different parameters (P, R, d)	89

Chapter 1

Introduction

In this chapter, we introduce the background of our study, followed by the research motivations and objectives. Then, we introduce the publicly available datasets used to evaluate our proposed methods. Finally, we summarize our contributions and give the structure of the report in the end.

1.1 Background

Autoimmune diseases (AD) occur when the immune system mistakenly attacks and destroys healthy body tissues, which can be organ-specific or systemic. They are prevalent diseases affecting a large number of human beings. There are more than eighty different types of AD such as rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis and autoimmune hepatitis. The antinuclear antibody (ANA), which refers to a unique group of autoantibodies targeting at the nuclear contents of the cell, has significant relation to specific AD, such as systemic lupus erythematosus (SLE), systemic sclerosis and rheumatoid arthritis. For instance, Figure 1.1 shows the patient with SLE showing inflammation and vasculitis, while Figure 1.2 shows a hand of the patient with rheumatoid arthritis. The identification of ANAs plays an important role in the clinical medicine and clinical immunology.

Figure 1.1: Patient with systemic lupus erythematosus (picture is from Google Images).

Figure 1.2: Patient with rheumatoid arthritis (picture is from Google Images).

Indirect immunofluorescence (IIF) is the original approach for ANA test described by Coons and Kaplan more than 50 years ago [1]. Over time, other methods of detecting ANA have been developed such as solid phase immunoassays, e.g. including enzyme-linked immunosorbent assay (ELISA), and the multiplex platform [2, 3, 4], which is much easier and cheaper compared to IIF. However, the new tests usually show lower sensi-

tivity [5, 6, 7]. In addition, different commercial ELISA kits produce different results [8] and they cannot avoid the problem of false positive results. Recently, the American College of Rheumatology issued a position statement stressing that the IIF method based on HEp-2 cells for ANA assay remains the gold standard [9]. HEp-2 cells, originally considered to originate from a human larynx carcinoma, are now known to have been established from a HeLa cell contamination [10]. They have many advantages, such as large nuclei, sensitiveness, etc.

The technique for performing the IIF test in the Clinical Immunology Laboratory of Tan Tock Seng Hospital, Singapore is shown in Figure 1.3 [11]. Briefly, the process of IIF is as follows: First, diluted patient's sample is incubated with the slides containing fixed HEp-2 cells. After a period, unbound antibodies are washed off and a fluorescein-conjugated anti-human immunoglobulin will be applied. Then, washing after the second incubation, any unbound secondary immunoglobulin will be removed. The ANAs are finally revealed as fluorescent cells under the fluorescence microscope. Both fluorescence intensity and positive staining patterns for each slide image are identified by highly qualified and skillful physicians.

In practice, the intensity of fluorescence cells is reported on a scale of values from 0 to 4+ as follows:

- 4+: bright green fluorescence;
- 3+: less bright green fluorescence;
- 2+: diminished fluorescence yet clearly observable;
- 1+: very subdued fluorescence;
- 0 : negative.

Figure 1.3: IIF procedure.

It is worth noting that negative intensity or level 0 indicates that the patient is normal. Staining patterns are recorded for samples with positive intensity (i.e., from 1+ to 4+), since different patterns are related to specific diseases. There are more than thirty kinds of staining patterns in the world, however, the most frequent staining patterns of HEp-2

cells in clinical practice are as follows [10, 12]:

- *Centromere*: characterized by 40-60 discrete speckles diffused over all the nuclei in the interphase and characteristically found in the condensed nuclear chromatin during mitosis as a bar of closely related speckles;
- *Nucleolar*: characterized by large coarse speckles (less than six) within the nucleoli of interphase cells;
- *Homogeneous*: characterized by a uniform distributed fluorescence of the entire interphase nuclei and fluorescence of the nuclear chromatin during mitotic;
- *Speckled*: there are two sub-categories:
 - *Fine Speckled*: characterized by a fine granular nuclear staining of nuclei in the interphase, which is distributed uniformly;
 - *Coarse Speckled*: characterized by dense, intermediate sized particles in interphase nuclei together with large speckles;
- *Cytoplasmic*: characterized by staining of the cytoplasm exclusive of the nucleus.

The staining patterns mentioned above are shown in Figure 1.4. As a matter of fact, according to different needs of the lab, the categorization of staining patterns can be changed slightly. In some cases, mixed staining patterns reveal when patient's serum contains more than one autoantibody specificity. However, mixed patterns are atypical and occur rarely in clinical diagnosis. In most situations, same pattern is shown in the same slide then it can be classified into common class. In academic community, only the images with the same staining pattern are used to study.

Figure 1.4: Typical HEp-2 cells with different staining patterns.

1.2 Motivation and Objective

Worldwide, there is a strong demand for IIF image analysis due to its effectiveness and high-quality. However, the main drawback of the strategy is the human subjective evaluation since the classification always depends on highly qualified experts' skill. Results are significantly influenced by one's qualification and reading systems, causing high-rate intra- and inter-laboratory variance. The low level of standardization limits the communications between the clinic units and reproducibility of IIF readings. As the demand for applying IIF method in diagnosis of autoimmune diseases is increasing, lacking of resources and adequately trained personnel is becoming another bottleneck [13, 14].

To address aforementioned issues, Computer Aided Diagnostic (CAD) systems are proposed for automatically supporting the IIF diagnosis in many ways. They can aug-

ment the technologists' capabilities, reduce the time and improve the reliability. They free the technologists from mass screening tasks and enable them to concern only about the most involved cases. Moreover, the CAD systems can be used as an education tool to train specialized medical personnel.

Several approaches have been proposed in the recent research for all the major stages of the IIF diagnostic procedure. The main technologies have been investigated in the CAD systems are automated preparation of slides with robotic devices [15], image acquisition [16, 17], image segmentation [18, 19], mitotic cell recognition [20], fluorescence intensity classification [15, 21] and staining pattern classification [14, 22, 23]. Till now, the highest level of automation in the CAD systems for ANA test is the preparation of slides by robotic devices conducting dilution, dispensation and washing operations [20]. In image acquisition stage, images are taken using a fluorescence microscope, equipped with both a mercury vapour lamp and a digital camera to auto-focus; then the digital images are displayed on a computer screen. Images are analyzed and processed to be more suitable to following classification. Fluorescence intensity is classified into negative group and positive group with intensity level based on the intensity-related features. The positive group is further classified into several main staining pattern groups based on the pattern-related features.

While all aspects of the CAD systems contribute to the automation of IIF procedure in one way or another, staining pattern classification is proven to be the most challenging task in the research community due to large intra-class variation and small between-class variation regardless of its importance. To reduce the variability of multiple readings, the levels of fluorescent intensity are always generally classified into three levels named *negative*, *intermediate* (with intensity of 1+) and *positive* (with intensity of 2+ or more), where intermediate and positive is belong to the positive group whose pattern need to be further identified. In the following of this thesis, when we say "positive staining patterns",

ICPR2012 Dataset

ICIP2013 Training Dataset

Figure 1.5: Sample images from the ICPR2012 dataset and the ICIP2013 training dataset respectively. The rows named “positive” are the patterns with positive intensity, while the “intermediate” rows are the patterns with intermediate intensity.

we refer to staining patterns with non-negative fluorescence intensity, which includes positive and intermediate level. As shown in Figure 1.5, the cells with intermediate and positive intensity which are from the same pattern categories have large variations. Particularly, the cells with intermediate intensity can not be seen clearly. Meanwhile, some categories share similar shape and texture. Image representation is crucial for HEP-2 cells analysis. Compared with the signal in general image classification, HEP-2 cells do not contain abundant structural information. In addition, the features between various HEP-2 cells are much more similar than those between different objects or natural scene images. In this thesis, we investigate into the feature extraction and machine learning methods for automatic staining pattern classification of HEP-2 cells.

This PhD work is collaborated with Tan Tock Seng Hospital in Singapore. The objective of our work is to classify the positive staining patterns automatically into different classes and get high accuracy to support doctors' final diagnosis in practice. Designing effective features is a fundamental and critical issue in computer vision. It plays a significant role in a wide range of applications including texture classification, object and scene recognition, image retrieval, face recognition, et al. Suitable features can efficiently present different contents in the image and are essential to obtain satisfying classification performance. Therefore, our work aims to develop discriminative and descriptive features for the HEP-2 cell images.

1.3 Experimental Datasets

In order to evaluate the performance of our proposed methods in this thesis, we use two publicly available HEP-2 cells datasets: ICPR2012 dataset from the ICPR'12 *HEP-2 cell classification contest* dataset and ICIP2013 training dataset from the ICIP'13 *Competition on cells classification by fluorescent image analysis*. Some examples of the datasets are shown in Figure 1.6.

1.3.1 The ICPR2012 Dataset

The ICPR2012 dataset consists of 1455 HEp-2 cells segmented from 28 slide images which are obtained by using a fluorescence microscope of 40-fold magnification, equipped with a 50W mercury vapor lamp and a digital camera using a CCD with square pixel of $6.45\mu m$. The resolution of obtained images is 1388×1038 pixels and the color depth is 24 bits. Each image can be categorized into one of six staining patterns, namely centromere (ce), coarse speckled (cs), cytoplasmic (cy), fine speckled (fs), homogeneous (ho) and nucleolar (nu). Also, fluorescent intensity (i.e., positive and intermediate) is assigned to each image. The cells in the images are manually segmented and annotated by specialists. Then, each cell image and slide image is verified by a medical doctor specialized in immunology with 11 years' experience. According to the experimental protocol of the ICPR'12 contest, the ICPR2012 dataset is divided into a training set with 721 cells from half of the slide images and a test set with 734 cells from rest of the slide images. The composition of the dataset is reported in Table 1.1.

Table 1.1: Composition of the ICPR2012 dataset. Each table item represents the number of cells and the number of images which is in parentheses.

Type	Training set	Test set	Total
Centromere	208(3)	149(3)	357(6)
Homogeneous	150(3)	180(2)	330(5)
Nucleolar	102(2)	139(2)	241(4)
Coarse speckled	109(2)	101(3)	210(5)
Fine speckled	94(2)	114(2)	208(4)
Cytoplasmic	58(2)	51(2)	109(4)
Total	721(14)	734(14)	1455(28)

Figure 1.6: Samples of the ICPR2012 dataset and the ICIP2013 training dataset with different staining patterns of HEp-2 cells.

1.3.2 The ICIP2013 Training Dataset

The HEp-2 cell images of ICIP2013 dataset is obtained by using a monochrome high dynamic range cooled microscopy camera which is fitted on a microscope with a plan-Apochromat 20x/0.8 objective lens and an LED illumination source. At least two scientists are involved in the labeling process. In dubious cases, a third expert is asked to adjudicate the conflict between the two scientists. So far, only the training dataset is available. However, the training dataset is big enough to evaluate different methods. The ICIP2013 training dataset contains 13596 cells which are categorized into six classes: ho-

homogeneous (ho), speckled (sp), nucleolar (nu), centromere (ce), nuclear membrane (nm) and golgi (go). The dataset includes two patterns less frequent occurring in the practical clinic as follows [24]:

- *Nuclear membrane*: characterized by a thin membranous fluorescence around the nucleus in the interphase cells;
- *Golgi*: characterized by speckled staining of a polar organelle adjacent to one part of the nucleus and composed of irregular large granules.

Thus, it offers a more realistic evaluation on the automatic classification algorithms. We partition the ICIP2013 training dataset into a training set consisting of 6842 cells from 42 slide images and a test set consisting of 6754 cells from 41 slide images. See Table 1.2 for detailed information about the dataset.

Table 1.2: Composition of the ICIP2013 training dataset. Each table item represents the number of cells and the number of images which is in parentheses.

Type	Training set	Test set	Total
Centromere	1279(16)	1462(16)	2741(32)
Homogeneous	1347(16)	1147(16)	2494(32)
Nucleolar	1273(16)	1325(16)	2598(32)
Speckled	1391(16)	1440(16)	2831(32)
Nuclear membrane	1190(16)	1018(15)	2208(31)
Golgi	362(2)	362(2)	724(4)
Total	6842(42)	6754(41)	13596(83)

1.4 Contributions and Structure of the Thesis

The main contributions of this thesis can be summarized as follows:

- We revised feature extraction-coding-pooling framework which has shown impressive performance in image classification tasks, because it can obtain discriminative and effective image representation. However, the information loss is inevitable in the coding process. Therefore, we proposed a novel Linear Local Distance Coding (LLDC) method to capture more discriminative information. LLDC transforms original local feature to local distance vector by searching for local nearest few neighbors of local feature in the class-specific manifolds. The obtained local distance vector is further encoded and pooled together to get salient image representation. We demonstrated the effectiveness of LLDC method on two public HEp-2 cells datasets: the ICPR2012 dataset and the ICIP2013 training dataset. Experimental results showed that our proposed method has a superior performance to the state-of-the-art coding methods for staining pattern classification of HEp-2 cells.
- We proposed a novel feature extraction scheme for automatic staining pattern classification of HEp-2 cells. Our proposed method constructs a dual spatial pyramid structure on a powerful rotation invariant texture feature, which has the following advantages: (1) invariance under local rotation of the image, (2) robustness against resolution changes, and (3) strong descriptive ability. Incorporated with a linear Support Vector Machine (SVM) classifier, the proposed method demonstrates its effectiveness by testing on two HEp-2 cells datasets. Particularly, it showed superior classification performance than the best performer at the first edition of the contest for staining pattern classification. Furthermore, we improved the classification accuracy by integrating the the proposed method and Bag-of-Words (BoW) representation.
- We proposed a Co-occurrence Differential Texton (CoDT) feature which is designed to represent the local image patches. Then, a generative model was built to

adaptively characterize the CoDT feature space. We further exploited a more discriminant representation for HEP-2 cell images based on the adaptive partitioned feature space, then fed the representation into a linear SVM classifier for identifying the staining patterns. The experimental results verified that our proposed method remarkably shows impressive performance for staining pattern classification.

Overall, this thesis researches on the features suitable for characterizing the staining patterns of HEP-2 cells. We test related algorithms on challenging publicly available datasets, showing the proposed algorithms improve upon the state-of-the-art.

The thesis is organized as follows:

Chapter 1 is an introduction to this work, presenting the clinic background, research objective, experimental datasets and a brief summary of our contributions.

Chapter 2 reviews the preliminary and related works which have been done for classifying staining patterns of HEP-2 cells.

Chapter 3 first introduces the major technological steps of the BoW framework. Then, a novel coding method named LLDC is proposed and the experimental evaluation is reported.

Chapter 4 proposes a rotation invariant textural feature to characterize the staining patterns. It is confirmed that combining with the BoW representation can improve the classification performance.

Chapter 5 proposes a more discriminant representation based on the CoDT features to better fit the feature space of HEP-2 cell images.

Chapter 6 concludes our current studies and discusses the future works.

Chapter 2

Related Work

In this chapter, we review the literatures on analysis and classification of HEP-2 cell images. Our works focus on the efficient features and classifiers for staining pattern classification of HEP-2 cells. Therefore, we introduce some related researches according to our works.

2.1 Staining Pattern of HEP-2 Cells

The interests for staining pattern classification have been aroused for decades. Perner et al. [19] extract textural and statistics features for cell image representation, and combined features with a decision tree classifier. Soda and Iannello [22] develop a multiple expert system based on a set of statistical and spectral features to assign the pattern of single cell. Elbischger et al. [25] extract a set of features, including image statistics, cell shape and textural features, from segmented cells. A Mahalanobis distance classifier is finally chose for cell type classification. Hsieh et al. [26] utilize Learning Vector Quantization (LVQ) classifier based on eight textural features, e.g. standard deviation, Spatial Gray-Level Dependence Matrices (SGLDM), Neighborhood Gray-Tone Difference Matrix (NGTDM), etc., to identify the staining patterns. Foggia et al. [20] describe the peculiarities of mitotic cells by using intensity histogram, grey-level co-occurrence matrix

and Local Binary Patterns (LBP). Then they test five classifiers: Multi-Layer Perceptron (MLP), Naive Bayes classifier, k -Nearest Neighbor (k NN), SVM and AdaBoost. The experimental results indicate that both k NN and AdaBoost provides the best performance. Cordelli and Soda [27] experimentally compare four different methods for converting color images into gray scale ones, which are weighted conversion, green channel, intensity channel and Helmholtz-Kohlrausch (HK) conversion. By considering a heterogeneous set of features, e.g., spectral descriptors, statistical descriptors and morphological descriptors, AdaBoost classifier based on intensity channel achieves the best average performance. To extract features of the slide images without segmentation, we verify the effectiveness of Scale Invariant Feature Transform (SIFT) for representing HEP-2 slide images [28]. Wiliem et al. [29] propose a HEP-2 cell classification system comprised of a dual-region codebook-based descriptor, combined with the Nearest Convex Hull classifier.

Aforementioned works are based on private datasets and/or different experimental protocols. While each of aforementioned works addresses one or two aspects in technical advancement, it is difficult in reproduction of their procedures and comparative study in their performance.

Due to the great impact of staining pattern classification of HEP-2 cells on clinical practice, to make the comparisons among different approaches available, the first edition of the *HEP-2 Cells Classification Contest* hosted by the International Conference on Pattern Recognition 2012 (ICPR'12) with a publicly available HEP-2 cell dataset (ICPR2012 dataset) was released. The contest focused on the comparison of systems able to automatically classify the staining patterns of HEP-2 cells within IIF images. Applying the same database and experimental protocol, different systems can be compared and evaluated based on the benchmark. Nosaka et al. [30] utilize an extension of LBP descriptor, named Co-occurrence of Adjacent Local Binary Patterns (CoALBP), to extract textural features. Using linear SVM, their method won the first prize in HEP-2 cells classification

contest [12]. As the second place winner, Kong et al. adopt Varma's MR8 method [31] to extract statistical intensity features. Then each image is characterized by the frequency histogram of the visual words from a global dictionary trained beforehand. Li et al. [32] present four image descriptors including LBP, Gabor, Discrete Cosine Transform (DCT) and a global appearance statistical descriptor. Applying a multiclass boosting SVM algorithm to integrate these descriptors, they obtained the third prize of the contest. All the participated methods are reported in [12] for reference.

Inspired by the first contest, Shen et al. [33] adopt the BoW framework using intensity order pooling based gradient feature. We further improve the coding methods of the BoW framework to reduce information loss caused by feature quantization and apply a linear SVM for classification [34, 35]. To further enhance the descriptive ability of CoALBP, Nosaka et al. [36] propose a more powerful Rotation Invariant Co-occurrence among Local Binary Patterns (RIC-LBP) which is robust to local image rotations. We [37] design a dual spatial pyramid structure on a rotation invariant texture feature to take resolution variations into account and capture spatial layout information of the HEp-2 cell images. Stoklasa et al. [38] extract five image descriptors: Haralick features, LBP, SIFT, surface descriptor and a granulometry-based descriptor. They use the k -NN classifier, which is based on aggregated distance function combining several features together. Theodorakopoulos et al. [39] fuse the distribution of SIFT features and gradient-oriented co-occurrence of LBPs into a dissimilarity space. Then, they utilize a sparse representation-based classification mechanism for classification. Di Cataldo et al. [40] propose a classification approach based on Subclass Discriminant Analysis (SDA), using the integration of morphological, global and local texture features. Kong et al. [41] propose a supervised discriminative algorithm for learning vocabulary. They synergistically combine K-SVD and Sparse Representation-based Classification algorithm (SRC)

to specifically classify staining patterns of HEp-2 cells. Nanni et al. [42] extract features based on pyramid LBP coupled with a method to handle nonuniform bins. The features are then combined and fed into a SVM classifier. Liu and Wang [43] propose a multi-projection-multi-codebook scheme to create multiple linear projection descriptors from k local regions in the feature space and each projection matrix serves as a different type of local descriptor. A SVM with Hellinger's kernel is chosen for final classification due to its good performance. Yang et al. [44] firstly generate the Spontaneous Activity Patterns (SAP) image patches by using SAP model. Then they employ the Independent Component Analysis (ICA) framework to learn a set of filters from SAP patches and patches extracted from HEp-2 cell images. The obtained responses by filtering the image are used as image representation and are finally fed into a kernel SVM classifier based on the Pearson correlation. Thibault et al. [45] present three statistical matrices named multiple gray level Size Zone Matrix (MSZM), Distance Zone Matrix (DZM) and Orientation Length Zone Matrix (OLZM). They consider three kinds of popular and efficient classification methods, which are the logistic regression, the random forest and the neural network.

Following the first edition of HEp-2 cell classification contest, the IEEE International Conference on Image Processing 2013 (ICIP'13) and ICPR'14 hosted contest for HEp-2 cell classification respectively. Although considerable progress has been made, research in HEp-2 cell image analysis is still in its early stage and there is great potential for improvement.

2.2 Feature Extraction

Feature extraction is a very important issue in pattern recognition and classification. Suitable and discriminative features can efficiently present different contents in the image and offer strong supports for final classification. A large number of features have

been proposed, however, we introduce some important and effective features applied for representing the staining patterns in this section. Roughly, we introduce the features from two aspects, i.e., low-level features and mid-level features.

2.2.1 Low-level Features

We define the features describe image content from the primitive level as low-level features. It is a relative concept to the mid-level features obtained by using Bag-of-Words (BoW) framework.

2.2.1.1 Local Binary Patterns

LBP [46] is initially proposed to describe textural features for a local region. It can be obtained by thresholding the gray value of the circularly symmetric neighbor pixels with that of the center pixel. The neighbors whose difference is positive are set as ‘1’ while others are set as ‘0’. Then, these binary values are converted to a decimal number.

Let I be a grayscale image and $I(x, y)$ be a gray value at location (x, y) in I . Then LBP at location (x, y) is defined as

$$LBP_{P,R}(x, y) = \sum_{i=0}^{P-1} 2^i \text{sign}(I(x_i, y_i) - I(x, y)), \quad (\text{Eq. 2.1})$$

$$\text{sign}(x) = \begin{cases} 1 & , \quad \text{if } x \geq 0 \\ 0 & , \quad \text{otherwise.} \end{cases} \quad (\text{Eq. 2.2})$$

where $I(x_i, y_i)$ is the gray value of P equal spaced pixels on a circle of radius R around center pixel (x, y) and (x_i, y_i) is the location of neighbors given by $(x_i, y_i) = (x + R \cos(2\pi i/P), y + R \sin(2\pi i/P))$. If the neighbors do not fall in the center of pixels, their gray values should be estimated by interpolation [47]. Figure 2.1 illustrates the procedure to obtain $LBP_{P,R}$ value with different (P, R) .

Figure 2.1: Local binary patterns with different (P, R) s: $(4, 1)$, $(8, 1)$ and $(16, 2)$.

The $LBP_{P,R}$ produces 2^P different output values, therefore we can calculate the $LBP_{P,R}$ value for each pixel of an image and build a histogram with 2^P bins as the image descriptor. The original $LBP_{P,R}$ achieves invariance against any monotonic transformation and the scaling of the gray-scale.

To achieve rotation invariance, a unique identifier is assigned to each rotation invariant LBP [47], which is formulated by

$$LBP_{P,R}^{ri}(x, y) = \min\{ROR(LBP_{P,R}(x, y), s) | s = 0, 1, \dots, P - 1\}, \quad (\text{Eq. 2.3})$$

where $ROR(x, s)$ performs circle-right-shift on the binary number x s times. For example, $LBP_{P,R}$ values $10000100b$, $00100001b$ and $01000010b$ all map to the unique value $00100001b$.

Another extension of original LBP called “uniform” patterns [47]. A uniformity measure $U(LBP_{P,R})$ is introduced to denote the times of spatial transitions (bitwise 0/1 changes) in the LBPs. For instance, pattern 00000010*b* and 00111000*b* have $U(LBP_{P,R})$ value of 2; pattern 01100010*b* have $U(LBP_{P,R})$ value of 4. Uniform LBP assigns different label to each “uniform” pattern and gives a unique number to all the “non-uniform” patterns. The uniform LBP has $P(P-1) + 3$ different values. Figure 2.2 shows 58 different “uniform” patterns of $(8, R)$ neighborhood. Each “uniform” pattern has an unique label from 1 to 58 and all “non-uniform” patterns (there are 198 patterns are “non-uniform”) are assigned the same label 59. Let $U_P(i, j)$ be the label for a “uniform” pattern, where i is the number of ‘1’ in the pattern (row number in Figure 2.2) and j is the rotation degree (column number in Figure 2.2) The uniform LBP can be formulated as

$$LBP_{P,R}^{u2}(x, y) = \begin{cases} U_P(i, j) & , \text{ if } U(LBP_{P,R}) \leq 2 \\ P \times (P-1) + 3 & , \text{ otherwise.} \end{cases} \quad (\text{Eq. 2.4})$$

$$U_P(i, j) = \begin{cases} 1 & , \quad i = j = 0, \\ (i-1) \times P + j + 2 & , \quad 1 \leq i \leq P, 0 \leq j \leq P-1, \\ P \times (P-1) + 3 & , \quad i = P, j = 0. \end{cases} \quad (\text{Eq. 2.5})$$

The “uniform” patterns provide stronger ability of discrimination in comparison to including all patterns, because they have different statistical properties [47, 48]. Most of the LBPs in natural images are uniform. The proportion of “non-uniform” patterns is so small, therefore the estimation of their probabilities is unreliable. Meanwhile, “uniform” patterns are more stable and only considering “uniform” patterns makes the distribution estimation more reliable.

To improve the capability of rotation invariance and elevate the discrimination, some rotation invariant patterns with low occurrence frequencies, i.e., “non-uniform” patterns of rotation invariant LBPs, are eliminate. The improved rotation invariant texture feature can be defined as [47]

$$LBP_{P,R}^{riu2}(x, y) = \begin{cases} \frac{\sum_{i=0}^{P-1} \text{sign}(I(x_i, y_i) - I(x, y))}{P+1} & , \text{ if } U(LBP_{P,R}) \leq 2 \\ P+1 & , \text{ otherwise.} \end{cases} \quad (\text{Eq. 2.6})$$

Figure 2.2: The different “uniform” patterns in the case of $P = 8$. The white circle denotes ‘1’ while the dark circle denotes ‘0’.

Each rotation invariant “uniform” pattern is assigned a unique label equal to the number of ‘1’ in the pattern, while all the “non-uniform” patterns are labeled by $P + 1$. Therefore, there are $P+2$ different values for $LBP_{P,R}^{riu2}$ and the final feature is the histogram of $LBP_{P,R}^{riu2}$ accumulated over the entire image.

2.2.1.2 Scale Invariant Feature Transform

SIFT is a local descriptor which is invariant to image translation, scaling and rotation, and partially invariant to occlusion/noise and change in illumination and viewpoint [49]. Lots of algorithms have been proposed recently that successfully utilize SIFT for image classification and object recognition [50, 51, 52, 28]. There are mainly four stages of computation involved in the generation of the SIFT feature described in the following sections.

Scale-space extrema detection The first stage of computation is to identify the locations of candidate interest points which are invariant to orientation and scale change. A continuous function of scale, also named as scale space, is used to search stable features through all possible scales. It is verified that the Gaussian function is the only possible scale-space kernel [53], therefore the scale space of an image I can be generated by

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (\text{Eq. 2.7})$$

where $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp \frac{-(x^2+y^2)}{2\sigma^2}$ is a Gaussian function and $*$ is the convolution operation.

The locations of stable keypoints in scale space are detected by using the extrema of the difference-of-Gaussian (DoG) function convolved with the original image. It can be formulated by

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, r\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, r\sigma) - L(x, y, \sigma). \end{aligned} \quad (\text{Eq. 2.8})$$

Figure 2.3: Construction of $D(x, y, \sigma)$. The scale space images are produced by convolving initial image incrementally with Gaussians. The DoG images are produced by subtracting adjacent Gaussian images.

An efficient way to generate $D(x, y, \sigma)$ is illustrated in Figure 2.3 [49]. Each octave of scale space is divided into t intervals (t should be an integer number) and $r = 2^{1/t}$ is a multiplicative constant. The Gaussian image in next octave is the one in the previous octave downsampled by a factor of 2. It is verified that we have to produce $t+3$ Gaussian images for each octave to insure that the extrema detection can cover the complete octaves. The DoG images are generated by subtracting adjacent blurred images.

To find the local extrema of the DoG images, each sample has to be compared to its 26 neighbors in 3×3 regions of the DoG images at the current scale (eight neighbors), above scale (nine neighbors) and below scale (nine neighbors). Samples which are the maxima or minima among all of their neighbors are identified as the keypoint candidates.

Keypoint localization After finding keypoint candidates, it is necessary to find stabile keypoints. It means that the points with low contrast and poor localization along the

edges will be removed. It can be accomplished by using the Taylor expansion of the DoG image and the location of the extremum, $\hat{\mathbf{x}}$, can be determined by following formula [49]:

$$\hat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D}{\partial \mathbf{x}}, \quad (\text{Eq. 2.9})$$

where D , $\frac{\partial D}{\partial \mathbf{x}}$ and $\frac{\partial^2 D}{\partial \mathbf{x}^2}$ are evaluated at the same selected point and $\mathbf{x} = (x, y, \sigma)^\top$ is the offset from this point. It is worth noting that D is the Taylor expansion up to second order of original $D(x, y, \sigma)$.

The function value at the extremum $D(\hat{\mathbf{x}})$ is calculated by

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D^\top}{\partial \mathbf{x}} \hat{\mathbf{x}}. \quad (\text{Eq. 2.10})$$

To reject the unstable extrema with low contrast, based on the experimental results in [49], those $\hat{\mathbf{x}}$ with $D(\hat{\mathbf{x}}) < 0.03$ is discarded.

To define the extreme points along the edges, a 2×2 Hessian matrix is utilized as

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}, \quad (\text{Eq. 2.11})$$

where D_{xx} , D_{xy} and D_{yy} is the second partial derivative of the DoG image.

To further eliminate the influence of the points localized along the edges, the candidate keypoints which are unable to satisfy following situation will be eliminated:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(\gamma + 1)^2}{\gamma}, \quad (\text{Eq. 2.12})$$

where $Tr(H)$ and $Det(H)$ is respectively the trace and the determinant of matrix H . $\gamma = 10$ is the ratio between the eigenvalue of H with the largest magnitude and the one with smaller magnitude.

Orientation assignment Until now, the accurate keypoints are localized by removing the points with low contrast or along edges. The next step is to assign an orientation to each keypoint.

The gradient magnitude and orientation of the Gaussian smoothed image $L(x, y)$, which is obtained by using the scale σ of the keypoint, can be calculated by

$$m(x, y) = \sqrt{(L(x, y + 1) - L(x, y - 1))^2 + (L(x + 1, y) - L(x - 1, y))^2}, \quad (\text{Eq. 2.13})$$

$$\theta(x, y) = \tan^{-1}\left(\frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)}\right). \quad (\text{Eq. 2.14})$$

The orientations of sample points within a window around the keypoint are stored in one of 36 bins covering the 360° . Each point weighted by its gradient magnitude within a circular window with $\sigma_w = 1.5\sigma$ around the keypoint, is added to the bin corresponding to the point's orientation. The highest peaks in the histogram is the dominant directions of the keypoint. Some other peaks with higher than 80% of the highest one can also be used to create a keypoint with corresponding orientation. At last, a parabola is fitted to the three histogram values closest to each peak for generating the orientation with higher accuracy.

The location and orientation of keypoints in an image are shown in Figure 2.4. By assigning orientation to each keypoint based on local image natures, the capability of invariance to image rotation can be obtained.

Keypoint descriptor The location, scale and orientation has been assigned to each keypoint of an image. The last stage of the SIFT calculation is to create the descriptor, which should be highly distinctive and be partial invariant under differing illumination and viewpoint.

Firstly, the coordinates of the descriptor and the orientations of the local gradient are rotated relative to the orientation of keypoint to achieve orientation invariance. The

Figure 2.4: Keypoints detected in an image. The start point of arrow is the keypoint's location, the direction indicates the orientation of the local gradient at the keypoint and the length denotes the magnitude of the local gradient.

gradient magnitude and orientation are sampled in a region of 16×16 pixels around the keypoint. The magnitudes are weighted by a Gaussian window with a σ that is 1.5 times that of the circular descriptor window. Then, the orientation histograms over 4×4 sample regions are calculated by accumulating the weighted magnitudes with nearly the same direction. Figure 2.5 shows a 4×4 keypoint descriptor array with 8 orientation bins covering 360° in each. The length of each arrow is the sum of the gradient magnitudes of the samples near that orientation in the corresponding region. Since there are 4×4 histogram arrays with 8 orientation bins, which is verified to show the best result [49], the dimension of the feature vector 128.

At last, the feature vector is normalized to unit length and the values in the unit feature vectors which are larger than 0.2 are changed to 0.2. Then the modified feature vectors are normalized again. The final feature vector achieves invariant to illumination with affine changes.

Therefore, the final SIFT features achieve orientation and scale invariance, partial illumination invariance and be stable when noise is added into the image.

Figure 2.5: Keypoint descriptor.

2.2.2 Mid-level Features

BoW framework and spatial pyramid matching (SPM) are two popular examples of mid-level features. The target of BoW framework is to embed low-level descriptors in a representative codebook space. We introduce the key techniques employed in the BoW framework including SPM. First of all, low-level descriptors are firstly extracted at interest points or in dense grids. Then, a pre-defined codebook $\mathbf{B}$ is applied to encode each descriptor using a specific coding scheme. The code is normally a vector with binary or continuous elements depends on coding scheme, which can be referred as mid-level descriptor. Next, the image is divided into increasingly finer spatial subregions. Multiple codes from each subregion are pooled together by averaging or normalizing into a histogram. Finally, the final image representation is generated by concatenating the histograms from all subregions together. There are two modules in the framework, i.e., coding and pooling.

- Coding: Local features of each image can be transformed to a collection of

feature codes using a specific coding method. We compute a set of codes $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N\} \in \mathbb{R}^{M \times N}$ to represent the input local features $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{D \times N}$ by

$$\min_{\mathbf{C}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{c}_i\|_{\ell_2}^2 + \lambda \mathcal{R}(\mathbf{x}_i) \quad (\text{Eq. 2.15})$$

where the first term measures the approximation error and the second one serves as a regularization term. We minimize information loss mainly by adjusting the regularization term.

- **Spatial Pooling:** The pooling procedure transforms mid-level features from an image into a final image representation. A crucial component which has great impact on pooling is SPM. It captures spatial layout information by expressing spatial relations at multiple levels of quantization. The codes within each spatial subregion are summarized by using a specific statistics strategy, such as the average of codes or their maximum.

2.3 Classification

Classification is the final and the most essential part for CAD systems. In this section, we introduce some basic classifiers applied for positive staining pattern classification.

2.3.1 Support Vector Machine

SVM classifier is one of the latest and most successful supervised learning classifiers and has been widely applied for image classification due to its efficiency. Using training labeled samples, a statistical model is constructed and then new samples can be classified according to this trained model. The linear SVM aims at searching for an optimal hyperplane (or hyperplanes) in feature space with a large separating margin and a minimal misclassification rate.

For a binary linear SVM classifier, given training data and its corresponding labels $(x_i, y_i), i = 1, 2, \dots, l$, $x_i \in R^n$, $y_i \in \{-1, +1\}$, the concrete formulation can be defined by

$$\begin{aligned} & \arg \min_{\mathbf{w}, \xi} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^l \xi_i, \\ & s.t. \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad \text{for } i = 1, 2, \dots, l. \end{aligned} \quad (\text{Eq. 2.16})$$

where $C > 0$ is a penalty parameter to allow some misclassification and ξ_i are slack variables. The objective function aims to maximize the margin and the constraints indicate that the training points should be correctly classified by the relaxed decision function $\mathbf{w}^\top \mathbf{x} + b$. To extend binary SVM for multi-class problems, we use the *one-vs-all* approach [54]. We train a single binary SVM classifier per class by treating the cell images of this class as positive samples and those of other classes as negative samples. When classification is performed, all the binary classifiers are run and the classifier with the highest confidence score is chosen.

2.3.2 Nearest Neighbor Classifier

Traditional Nearest Neighbor Classifier (NNC) is one of the most commonly used and the simplest pattern classification methods yet devised [55]. It is a kind of statistic machine learning methods. We have N training sample pairs $(x_i, y_i), i = 1, 2, \dots, N$, where x_i is a set of features and y_i is class label. For a set of features x extracted from an image, it is desired to predict label y by utilizing the information contained in the set of training samples which are labeled correctly. A distance vector $D = d_1, d_2, \dots, d_N$ is firstly calculated by $d_i = \sqrt{(x - x_i)(x - x_i)'}$. The l -th sample is the nearest neighbor of x if $d_l = \min_i d_i$. Therefore x is categorized into the class y_l .

k -Nearest Neighbor algorithm (k NN) is also a simple classifier as a variant of NNC. Based on distance vector D , k nearest neighbors of x with k smallest distances are

obtained. Let the k nearest neighbors be $\{(x'_1, y'_1), \dots, (x'_k, y'_k)\}$, x is assigned to class y according the majority voting among the labels $y'_1, \dots, y'_k$. The only parameter is k which should be chosen carefully. Generally, larger values of k reduce the effect of noise for classification and improve the classification performance, but they make the classes less distinct.

Chapter 3

Linear Local Distance Coding Method

In this chapter, we propose an automated system for identifying staining patterns of HEP-2 cells based on a *feature extraction-coding-pooling* framework, which can produce discriminative and effective image representation. However, the information loss is inevitable in the coding process. Therefore, we propose a Linear Local Distance Coding (LLDC) method to capture more discriminative information. Our proposed LLDC method transforms each original local feature to a more discriminative local distance vector by searching for local nearest few neighbors of the local feature in the class-specific manifolds. The obtained local distance vector is further encoded and pooled together to get salient image representation. The LLDC method is combined with the traditional coding methods to achieve higher classification accuracy.

3.1 Introduction

The most popular image classification framework consists of two major modules: Bag-of-Words (BoW) and Spatial Pyramid Matching (SPM). The framework of SPM based on BoW has been successfully applied to image classification [56, 57] and in recent years

it has been improved for classifying staining patterns of HEP-2 cells [33, 58]. It seems to be suitable for the staining pattern classification task. Within the framework, how to encode each local feature has significant impact on the final classification performance. The traditional and the simplest coding method is Vector Quantization (VQ) [59] (a.k.a. Hard Quantization or Hard Assignment), which assigns a local feature to the closest visual word in the codebook/vocabulary, introducing unrecoverable discriminative information loss. The Soft Assignment (SA) coding method [60, 61, 62] is proposed to reduce information loss by assigning a local feature to different visual words according to its memberships to multiple visual words. Apart from information loss, traditional SPM based on VQ has to use a classifier with nonlinear Mercer kernels, resulting in additional computational complexity and reducing scalability for real application. To alleviate these limitations, Sparse Coding based SPM (ScSPM) [56], Local Coordinate Coding (LCC) [63] and Locality-constrained Linear Coding (LLC) [57] aims at obtaining a nonlinear feature representation which works better with linear classifiers. They search for some weighted coefficients to linearly combine visual words of the codebook to approximate the input low-level descriptor. Salient Coding (SaC) [64] and its extension Group Salient Coding (GSC) [65] is proposed to speed-up while reserving the classification accuracy.

All the improved methods represent images more accurately and achieve impressive image classification performance. However, information loss in feature quantization is still inevitable and affects the performance for good image classification performance. To avoid information loss caused by coding, Naive Bayes Nearest Neighbor (NBNN) method [66] is proposed by retaining all of the feature descriptors. It shows competitive classification performance with coding based methods as it alleviates information loss and keeps the discrimination of input features. However, the NBNN is sensitive to noisy features and easy to be dominated by outlier features. To simultaneously inherit

the advantage of the BoW framework and the NBNN method, Linear Distance Coding (LDC) method [67] has been proposed recently to utilize the discriminative information lost due to the traditional coding schemes. It transforms each local feature to a *distance vector* via calculating neighbors in every class-specific manifold.

In this chapter, we propose a novel LLDC method to increase the accuracy of staining pattern classification. The LLDC method adopts the feature extraction-coding-pooling framework based on *local distance vector* which is a modification of the distance vector. Local distance vector is generated by using only the local neighbors in a merged feature dataset instead of calculating neighbors in every class-specific feature dataset. Therefore it can ignore disturbance from isolated classes. Using image-to-class distance makes it more class-specific as desired for classification. Meanwhile, distance vector in LDC method is obtained by using a linear coding scheme which aggravates the information loss. Local distance vector is only based on Euclidean distance and avoids coding process in distance vector transformation, therefore it is more discriminative. In addition, it is proved that image representations via coding distance patterns are complementary to the ones obtained by the conventional coding schemes [67]. Therefore, we directly concatenate the image representations based on local distance vector and local features to achieve superior performance.

In summary, the main contributions of this study are as threefold: (i) We propose a novel local distance vector based on the image-to-class distance. It is more class-specific than original local feature. Unlike distance vector, it eliminates the need to calculate the distance for each class therefore it can speed up the calculation and achieve better classification performance by ignoring the disturbance from the distant classes. (ii) We propose a LLDC method based on the transformed local distance vector. It takes the advantages of the BoW framework and the NBNN method. It reduces the information

loss caused by traditional coding methods while capturing salient features. (iii) The combination of image representations produced by the LLDC method and the ones produced by the traditional coding methods, can yield superior performance compared with only using single representation. Experiments on two public HEp-2 cells datasets consistently show that the image representation produced by the LLDC framework achieves better performance compared with state-of-the-arts coding methods.

3.2 Proposed Method

In this section, we firstly introduce the BoW framework, including coding and pooling procedure. Then, we present a new coding method based on local distance vector which is a modification of the distance vector. Several methods taking advantage of locality, including LLC [57], LSC [62] and local NBNN [68], achieve improvements over their non-local versions [56, 60, 66]. Our proposed coding method maintains superior discriminative capability and effectiveness of the aforementioned coding methods. It provides better generalization capability by employing the distances between local descriptor and classes to estimate the image membership. Meanwhile, it preserves more discriminative information by avoiding coding process while obtaining image-to-class distance. Furthermore, the LLDC method avoids poor estimates from isolated classes by eliminating the need to calculate distance vector for each class. Hence, the LLDC method can achieve superior image classification performance compared with the other coding schemes.

3.2.1 Bag-of-Words Representation

As we introduced in Section 2.2.2, the target of BoW framework is to embed low-level descriptors in a representative codebook space. The overview is illustrated in Figure 3.1.

There are three main steps for generating BoW representation, which are feature extraction, coding and pooling. Each step within the framework can affect the quality of image representation and the following classification performance.

3.2.1.1 Coding

Here, we introduce three kinds of coding schemes, i.e., VQ, SA, its extension Localized Soft-assignment Coding (LSC) and LLC. Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{D \times N}$ be a set of local descriptors extracted from an image, where D is the descriptor dimensionality and N is the total number of descriptors. Given a codebook with M entities (a.k.a. visual words, atoms), i.e., $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M\} \in \mathbb{R}^{D \times M}$, then the corresponding codes of an image $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\} \in \mathbb{R}^{M \times N}$ can be generated by using various coding schemes.

Vector quantization VQ is the original coding scheme which solves the constrained optimization problem as follows:

$$\begin{aligned} \min_{\mathbf{Y}} \quad & \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{y}_i\|_{\ell_2}^2 \\ \text{s.t.} \quad & \|\mathbf{y}_i\|_{\ell_0} = 1, \|\mathbf{y}_i\|_{\ell_1} = 1, \mathbf{y}_i \geq 0, \forall i \end{aligned} \tag{Eq. 3.1}$$

The constraint $\|\mathbf{y}_i\|_{\ell_0} = 1$ denotes that there is only one non-zero element in $\mathbf{y}_i$ and $\|\mathbf{y}_i\|_{\ell_1} = 1, \mathbf{y}_i \geq 0$ indicates that the weight for each descriptor is equal to 1. In practice, Eq. 3.1 means that each descriptor $\mathbf{x}_i$ is assigned to its closest visual word of the predefined codebook with activation equal to 1. It is also known as Hard assignment or Hard quantization.

Soft assignment coding To relieve the quantization error of VQ, SA coding method [60, 61] assigns a local feature to all of the visual words. The code of corresponding local feature represents the membership of the local feature to each visual word.

Figure 3.1: Overview of the feature extraction-coding-pooling scheme.

The j th coding coefficient represents the degree of membership of a local feature $\mathbf{x}_i$ to the j th visual word:

$$y_{ij} = \frac{\exp(-\beta \|\mathbf{x}_i - \mathbf{b}_j\|_{\ell_2}^2)}{\sum_{k=1}^M \exp(-\beta \|\mathbf{x}_i - \mathbf{b}_k\|_{\ell_2}^2)}, \quad (\text{Eq. 3.2})$$

where β is the smoothing factor controlling the softness of assignment.

The SA coding method is computationally efficient and conceptually simple, which only needs to compute the distance from a local feature to each word. It employs the kernel function of distance as the encoding representation and uses multiple visual words for coding, which can improve the accuracy of probability density estimation. However, it cannot give excellent classification performance as the sparse or local coding methods, probably because it does not take the manifold structure of local features into account. Therefore, LSC [62] is proposed, which only considers the k nearest neighbors while coding a local feature.

LSC computes the j th coding coefficient of a local feature $\mathbf{x}_i$ as follows:

$$y_{ij} = \frac{\exp(-\beta \hat{d}(\mathbf{x}_i, \mathbf{b}_j))}{\sum_{l=1}^M \exp(-\beta \hat{d}(\mathbf{x}_i, \mathbf{b}_l))}, \quad (\text{Eq. 3.3})$$

$$\hat{d}(\mathbf{x}_i, \mathbf{b}_l) = \begin{cases} d(\mathbf{x}_i, \mathbf{b}_l) & , \text{if } \mathbf{b}_l \in \mathcal{N}_k(\mathbf{x}_i) \\ 0 & , \text{otherwise.} \end{cases} \quad (\text{Eq. 3.4})$$

where $\hat{d}(\mathbf{x}_i, \mathbf{b}_j)$ is the local version of $d(\mathbf{x}_i, \mathbf{b}_j)$ and $\mathcal{N}_k(\mathbf{x}_i)$ defines the k nearest neighbors of $\mathbf{x}_i$ in codebook.

The SA coding method improves the accuracy of probability density estimation compared with VQ method, because it employs the kernel function about distance to represent the codes instead of the vector with binary elements used in VQ. Furthermore, it uses multiple visual words while VQ method only considers the closest one visual word.

Locality-constrained linear coding Next, LLC [57] is a novel and practical coding scheme which transforms each input feature into a linear combination of the basis in a given codebook utilizing the locality constraint.

LLC uses the following criteria [57]:

$$\begin{aligned} \min_{\mathbf{Y}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{y}_i\|_{\ell_2}^2 + \lambda \|\mathbf{e}_i \odot \mathbf{y}_i\|_{\ell_2}^2 \\ \text{s.t. } \mathbf{1}^T \mathbf{y}_i = 1, \forall i \end{aligned} \quad (\text{Eq. 3.5})$$

where $\odot$ is element-wise multiplication, $\mathbf{e}_i \in \mathbb{R}^M$ is the locality adaptor that measures the similarity between the input descriptor $\mathbf{x}_i$ and codebook entities, and it is defined as follow:

$$e_{ij} = \exp\left(\frac{\|\mathbf{x}_i - \mathbf{b}_j\|_{\ell_2}}{\sigma}\right), \quad (\text{Eq. 3.6})$$

where σ is used to adjust weight decay speed for locality adaptor.

LLC generates a more accurate image representation and offers an analytical solution. After LLC process, nonlinear feature representations are obtained which can work better with linear classifiers.

3.2.1.2 Pooling

Pooling procedure converts the mid-level descriptor to final image representation by aggregating occurrences of visual words in the input image. To capture the shapes or locating an object, SPM [59] is proposed by dividing the image into increasingly finer spatial subregions and computing histogram of mid-level descriptors (codes for local descriptors) from each subregion. Let $\ell = 1, 2, \dots, L$ denote level of subpartition, such that there are $2^{\ell-1} \times 2^{\ell-1}$ subregions at level ℓ . Then, pooling strategy aggregates the occurrences of visual words for each subregion. The final image representation is generated by concatenating all the pooled features.

We introduce two simple yet efficient pooling schemes, i.e., average pooling and max pooling, which are respectively defined as follows:

$$\mathcal{H}_s^\ell = \frac{1}{\mathcal{N}} \sum \mathbf{y}_i, \mathbf{y}_i \in \mathcal{I}_s^\ell, \quad (\text{Eq. 3.7})$$

$$\mathcal{H}_s^\ell = \max \mathbf{y}_i, \mathbf{y}_i \in \mathcal{I}_s^\ell, \quad (\text{Eq. 3.8})$$

where $\mathcal{I}_s^\ell$ is the s th subregion at level ℓ and $\mathbf{y}_i \in \mathcal{I}_s^\ell$ denotes the encoded features within $\mathcal{I}_s^\ell$. $\mathcal{N}$ is the number of features within $\mathcal{I}_s^\ell$. The “sum” and “max” function is a row-wise manner. The max pooling and average pooling strategy is theoretically analyzed in [69, 70]. In particular, max pooling always performs better than average pooling when using a linear SVM.

3.2.2 Distance Vector

Distance vector is an alternative discriminative pattern of local feature in the *class-specific manifold coordinate* system. Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{D \times N}$ be N D -dimensional local features extracted from an image. It is assumed that the local feature is sampled from a class-specific manifold $M^c = [\mathbf{m}_1^c, \mathbf{m}_2^c, \dots, \mathbf{m}_{n_c}^c]$, which is constructed by clustering local features of the training images from the corresponding class c . Then the distance vector which denotes the distance between a local feature $\mathbf{x}_i$ and class c is computed by

$$d(\mathbf{x}_i, c) = \|\mathbf{x}_i - \mathbf{x}_i^c\|_{\ell_2}^2, \quad (\text{Eq. 3.9})$$

where $\mathbf{x}_i^c$ denotes the mapped point of $\mathbf{x}_i$ in class c . It can be computed by linearly combining its neighboring points in the manifold M^c . The LDC method calculates $\mathbf{x}_i^c$ as follow:

$$\begin{aligned} \min_{\mathbf{u}_i^c} \quad & \|\mathbf{x}_i - M^c \mathbf{u}_i^c\|_{\ell_2}^2, \\ \text{s.t.} \quad & u_{ij}^c = 0, \quad \text{if } \mathbf{m}_j^c \notin \mathcal{N}_i^k \\ & \mathbf{1}^T \mathbf{u}_i^c = 1, \forall i \end{aligned} \quad (\text{Eq. 3.10})$$

where $\mathbf{u}_i^c = [u_{i1}^c, u_{i2}^c, \dots, u_{in_c}^c]$ is the linear coefficients of $\mathbf{x}_i$ on the manifold M^c and $\mathcal{N}_i^k$ denotes the set of k nearest neighbors of $\mathbf{x}_i$ on M^c . Then, Eq. 3.9 can be rewritten as

$$d_i^c = d(\mathbf{x}_i, c) = \|\mathbf{x}_i - \mathbf{x}_i^c\|_{\ell_2}^2 = \|\mathbf{x}_i - M^c \mathbf{u}_i^c\|_{\ell_2}^2. \quad (\text{Eq. 3.11})$$

Each local feature of an image is transformed to its distance vector $\mathbf{d}_i = [d_i^1, d_i^2, \dots, d_i^C]$, where C is the class number.

By generating image representation based on distance vector, the LDC method captures discriminative information and avoids the case where the discriminative features are dominated by outlier or noisy features. Therefore, using the linear SVM, the LDC method shows impressive image classification performance. However, distance vector is obtained by utilizing the approximate fast solution of the LLC coding method which inherently induces information loss. Meanwhile, distance vector treats every class equally because it is produced through calculating the distance from local feature to each class. Such operation easily brings in the uncorrelated information of classes which are far from query local feature, and consequently arouses unnecessary interference. Therefore, distance vector can be improved further to perform better in image classification tasks.

3.2.3 Local Distance Vector

It is verified that using the distance between local feature and classes (*i.e.*, image-to-class distance) can provide better generation capability. We propose a novel distance pattern, called local distance vector, to define the distance from local feature to a specific class. Local distance vector eliminates the need to search for the nearest few neighbors in every class-specific manifold to generate distance vectors. Instead, it merges all the class-specific manifolds together to form a single dataset, *i.e.*, $M = [M^1, M^2, \dots, M^C] = \{\mathbf{m}_i\}_{i=1}^n$, where $\mathbf{m}_i$ is called “anchor points” [71] and n is the total number of points.

Figure 3.2: Distance vector *vs.* local distance vector. $\mathbf{x}_i$ is a query local feature. Distance vector searches the mapping point $\mathbf{x}_i$ which is determined by the nearest few neighbors in each manifold M^c . Local distance vector retrieves only the local neighborhood in $M = [M^1, M^2, \dots, M^C]$.

To obtain the class-specific distance, we search for k nearest neighbors of a local feature $\mathbf{x}_i$ in M , denoted as $NN(\mathbf{x}_i, k) = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k\} \in M$. Each neighbor $\mathbf{p}_i$ has a label $Class\{\mathbf{p}_i\}$ identifying it belongs to which class. We define the distance from $\mathbf{x}_i$ to those classes found in the k nearest neighbors as follow:

$$\hat{d}_i^c = \min_{\{\mathbf{p}_j | Class(\mathbf{p}_j) = c\}} \|\mathbf{x}_i - \mathbf{p}_j\|_{\ell_2}^2. \quad (\text{Eq. 3.12})$$

The difference between distance vector and local distance vector is shown in Figure 3.2. Our proposed local distance vector is less influenced by isolated classes since it only calculate distance vector for some classes which are close to the query feature. On the contrary, distance vector has to calculate the distance between the local feature and each class; it is inevitable to bring in some irrelative information from distant classes.

For those classes that are not found in the k nearest neighbors, we use the distance to the $k + 1$ nearest neighbors of $\mathbf{x}_i$ to estimate the class-specific distance. And the local distance vector of the local feature $\mathbf{x}_i$ is denoted as $\hat{\mathbf{d}}_i = [\hat{d}_i^1, \hat{d}_i^2, \dots, \hat{d}_i^C]$. The local distance vectors of an image is described in Algorithm 1.

Algorithm 1 Local Distance Vector

Require:

Local features $\{\mathbf{x}_{i=1}^N\}$ of a input image I ;
 the merged dataset M .

Ensure:

Local distance vectors $\hat{\mathbf{d}}_i, i = 1, 2, \dots, N$.

```

for  $i \leftarrow 1; i \leq N; i \leftarrow i + 1$  do
     $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{k+1}\} \leftarrow NN(\mathbf{x}_i, k + 1)$ 
    if category  $c$  is found in the  $k$  nearest neighbors of  $\mathbf{x}_i$  then
         $\hat{d}_i^c = \min_{\{\mathbf{p}_j | \text{Class}(\mathbf{p}_j) = c\}} \|\mathbf{x}_i - \mathbf{p}_j\|_{\ell_2}^2$ .
    else if category  $c$  is not found in the  $k$  nearest neighbors then
         $\hat{d}_i^c = \|\mathbf{x}_i - \mathbf{p}_{k+1}\|_{\ell_2}^2$ .
    Obtain  $\hat{\mathbf{d}}_i = [\hat{d}_i^1, \hat{d}_i^2, \dots, \hat{d}_i^C]$  for the local descriptor  $\mathbf{x}_i$ .
    
```

Unlike the original local features, local distance vector is more class-specific as desired for classification. Such class-specific distance captures the underlying manifold structure of the local features [67]. Meanwhile, it is obtained by using its nearest few neighbors avoiding coding process and ignoring some irrelative classes far from the local feature. Thus it gains stronger discriminative capability, and more robustness to noise and outlier features. Local distance vector obtains another advantage inherited from distance pattern, that is all local distance vectors within the same class are more similar in the distance feature space due to the class-specific characteristic. Therefore, it can cooperate better with following pooling procedure. Furthermore, the calculation of local distance vector is significant faster than that of distance vector because it is produced by searching for nearest neighbors within a merged reference dataset.

3.2.4 The Algorithm Framework

Our proposed LLDC method utilizes local distance vector to generate discriminative and effective image features, then adopts coding-pooling framework to obtain robust image representation. To evaluate the effectiveness and generalization of the proposed local

Figure 3.3: Overview of the image classification flowchart based on our proposed LLDC method.

distance transformation, we apply two different linear coding method respectively, *i.e.*, LLC [57] and LSC [62], to encode local distance vectors due to their high efficiency and prominent performance.

Within the proposed LLDC framework, the local distance vectors are firstly transformed from local features. Then the local distance vectors and the original local features are separately encoded and pooled to generated two image representations. Finally, we directly concatenate them to extract more discriminative and descriptive image representation while they are complementary to each other.

The overview of the LLDC framework is shown in Figure 3.3 including following steps:

- (1) The local features, $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{D \times N}$, are extracted from every image;
- (2) The local distance vectors, $\hat{\mathbf{d}} = \{\hat{\mathbf{d}}_i\}_{i=1}^N$, are transformed from local features one by one following Algorithm 1;
- (3) Local distance vectors are encoded by using LSC or LLC coding scheme based on a pre-trained codebook $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M\} \in \mathbb{R}^{D \times M}$. The resulted codes are $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\} \in \mathbb{R}^{M \times N}$;
- (4) Max-pooling is performed on the codes within each spatial subregion $\mathcal{I}^\ell$ as follow:

$$\mathcal{V}_d^\ell = \max(\mathbf{y}_k | \mathbf{y}_k \in \mathcal{I}^\ell) \quad (\text{Eq. 3.13})$$

where $\max$ is performed element-wisely for the involved vectors in each subregion and $\ell = 1, 2, \dots, \mathcal{L}$ is the numbering of subregions;

- (5) The image representation based on local distance vector can be generated by concentrating all the pooled features from every subregion, *i.e.*, $\mathcal{V}_{\hat{d}} = [\mathcal{V}_{\hat{d}}^1; \mathcal{V}_{\hat{d}}^2; \dots; \mathcal{V}_{\hat{d}}^{\mathcal{L}}]$. And the representation is normalized by

$$\mathcal{V}_{\hat{d}} = \mathcal{V}_{\hat{d}} / \|\mathcal{V}_{\hat{d}}\|_{\ell_2} \quad (\text{Eq. 3.14})$$

- (6) The original local features are also aggregated under the coding-pooling framework through step (3) to (5) to get the image representation $\mathcal{V}$;
- (7) The final image representation obtained by combining aforementioned two image representations $\mathcal{V}_{\hat{d}}$ and $\mathcal{V}$ is fed into a linear SVM classifier for classifying the staining patterns of HEP-2 cells.

3.3 Experiments and Analyses

To evaluate the performance of the proposed LLDC method for staining pattern classification, we use two publicly available HEP-2 cells datasets as described in Section 1.3.

3.3.1 Experiment Setup

We firstly extract dense SIFT features as the local feature. SIFT features are invariant to scaling and rotation, and partially invariant to illumination change, viewpoint change and noise. These properties are advantageous in staining pattern classification as cell images are unaligned and have high within class variabilities. In our experiments, SIFT features are extracted at single-scale from densely located patches of gray-level images. The patches are centered at every 6 pixels and with a fixed size of 18×18 pixels.

To obtain local distance vectors, the number of anchor points $\{\mathbf{m}_i^c\}$ for each class manifold M^c are fixed to 1024, then the size of the merged M for our proposed local distance vectors transformation is 6144×128 . For the SIFT features and the corresponding local distance vectors, all the codebooks in coding process contain 1024 visual words learned from training samples by using k-means clustering method. One of the most important parameters for our proposed LLDC method is k_{LDV} that defines the neighborhood of a local feature in local distance vector transformation. In the following coding process, the number of neighbors in the LLC method(i.e., k_{LLC}) is another parameter which can influence the classification performance. We also adopt the LSC method to encode the local distance vector, therefore the impact of neighbor size (i.e., k_{LSC}) will be discussed while the smoothing factor β is fixed as 10. We will study the influence of these parameters for staining pattern classification in Section 3.3.4.

After coding process, we partition each cell image into three increasingly finer subregions, which is 1×1 , 2×2 and 4×4 . We apply max-pooling strategy to pool the codes from each spatial subregion. The obtained features within all the subregions are concatenated to generate final image representation. Then we employ a linear SVM classifier for classification. In our experiments, we use the LIBLINEAR package [72], thanks to its efficiency in implementation. The linear SVM is trained based on the training set by 10-fold cross validation strategy and tested using the test set. The training set is randomly partitioned into 10 equal sized subset. A single subset is chosen as the validation data for testing the linear SVM and the remaining nine subsets are used for training. Each subset has to be used once as the validation data. Therefore, aforementioned procedure is repeated 10 times. The penalty parameter of the linear SVM is set as $C = 10$. Actually, the classification performance is almost steady for different values of penalty parameter.

The experimental results are reported at the cell level and the image level respectively. At the cell level, let tp_i , tn_i , fp_i and fn_i respectively denote the *true positives*, *true*

negatives, *false positives* and *false negatives* for an individual staining pattern class c_i . In our experiments, we use the performance measures *accuracy* and *sensitivity* at the cell level which are formulated as

$$\text{accuracy} = \sum \frac{tp_i}{tp_i + tn_i + fp_i + fn_i}, \quad (\text{Eq. 3.15})$$

$$\text{sensitivity} = \frac{\sum \frac{tp_i}{tp_i + fn_i}}{S}, \quad (\text{Eq. 3.16})$$

where S is the number of staining pattern classes.

At the image level, the prediction for staining pattern of each image is decided by the most frequently assigned pattern of the cells within that image. In our experiments, we use $\text{accuracy} = \frac{\text{\#correctly classified slide images}}{\text{\#slide images}}$ as the classification accuracy at the image level, where $\#$ means “the number of”.

3.3.2 Experimental Results on the ICPR2012 Dataset

We first test performance of the proposed LLDC method on the ICPR2012 dataset following the experimental protocol of the *HEp-2 cells classification contest* by dividing the cell images into a training set and a test set. The subdivision is performed while maintaining approximately the same image pattern distribution over the two sets [12]. To validate the efficiency of our proposed method for staining pattern classification, we compare four different image representation: the original SIFT based BoW image representation (LLC/LSC-sift), the distance vector based image representation (LLC/LSC-(sift+dv)), our proposed image representation using local distance vector (LLC/LSC-ldv) and the proposed concatenated image representation (LLC/LSC-(sift+ldv)). Table 3.1 gives the comparison results of the classification performance at the cell level and at the image level. It can be observed that, the proposed LLDC method outperforms all the other methods. It is worth noting that the LLDC method outperforms CoALBP [30] which is

the winner of the contest with 70.4% of classification accuracy and 68.4% of sensitivity (at the cell level). Furthermore, the performance obtained by LLC/LSC-(sift+ldv) is better than that obtained by LLC/LSC-sift and LLC/LSC-ldv respectively. In particular, the classification performance achieved by LSC-(sift+ldv) is better than that achieved by LLC-(sift+ldv).

Table 3.1: Classification performance on the ICPR2012 dataset.

Algorithm	Cell level	Cell level	Image level
	Classification accuracy	Sensitivity	Classification accuracy
LLC-(sift+ldv) ^a	70.9%	71.6%	78.6%
LLC-(sift+dv) ^b	67.7%	69.2%	78.6%
LLC-ldv ^c	67.4%	68.9%	78.6%
LLC-sift ^d	66.4%	68.1%	78.6%
LSC-(sift+ldv) ^d	71.7%	72.9%	85.7%
LSC-(sift+dv) ^f	69.3%	70.6%	78.6%
LSC-ldv ^g	69.1%	70.6%	78.6%
LSC-sift ^h	68.9%	70.6%	78.6%

^a The LLDC method based on the concatenated representations of LLC-sift and LLC-ldv.

^b The LDC method based on the concatenated representations of LLC-sift and LLC-dv.

^c The LLC method based on local distance vector. ^d The LLC method based on SIFT features.

^e The LLDC method based on the concatenated representations of LSC-sift and LSC-ldv.

^f The LDC method based on the concatenated representations of LSC-sift and LSC-dv.

^g The LSC method based on local distance vector. ^h The LSC method based on SIFT features.

Table 3.2 shows the confusion matrix at the cell level by the proposed LLDC method using the LSC strategy on the concatenated image representation. The entry (corresponding to row i and column j) in the confusion matrix represents the percentage of cells from class i assigned to class j . It is obvious that cytoplasmic, centromere and homogeneous patterns are classified more accurately than the others. Particularly, cytoplasmic can achieve 100% of classification accuracy. Compared to the cytoplasmic pattern with distinguishable shape and centromere pattern with clear fluorescent dots, fine speckled pattern and homogeneous pattern has similar characteristic and is hard to find discriminative features to separate.

Table 3.2: The confusion matrix for the cell level classification on the ICPR2012 dataset.

	ce	cs	cy	fs	ho	nu
ce	84.2%	3.4%	0.6%	0.0%	1.3%	10.5%
cs	6.9%	72.3%	3.8%	12.1%	4.9%	0.0%
cy	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%
fs	22.8%	1.7%	2.6%	39.1%	32.9%	0.9%
ho	6.7%	4.5%	0.0%	11.1%	76.0%	1.7%
nu	20.1%	0.0%	5.6%	0.7%	7.9%	65.7%

To evaluate the classification performance at the image level, we report the corresponding confusion matrix in Table 3.3. Similarly, the table represents the percentage of images of class i identified to class j with respect to the total number of images in the test set. Our proposed LLDC method obtains 85.7% of the image-level classification accuracy. Centromere, cytoplasmic, homogeneous, and nucleolar patterns achieve 100% of classification accuracy. The most frequent mistake is existed between fine speckled and homogeneous pattern, which is the common mistake at the cell level.

Table 3.3: The confusion matrix for the image level classification on the ICPR2012 dataset.

	ce	cs	cy	fs	ho	nu
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%	50.0%	50.0%	0.0%
ho	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
nu	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%

3.3.3 Experimental Results on the ICIP2013 Training Dataset

Based on the ICIP2013 training dataset, the classification performance of different algorithms at the cell level and the image level is shown in Table 3.4. Our proposed LLDC method achieves the best performance. Particularly, LLC-(sift+ldv) can achieve better

classification performance than LSC-(sift+ldv). In summary, LLC-related features perform better for the ICIP2013 training dataset while LSC-related features is more suitable for the ICPR2012 dataset.

Table 3.5 shows the confusion matrix at the cell level by the proposed LLDC method using the LLC strategy on the concatenated image representation. Nuclear membrane pattern gets the highest classification accuracy rate, followed by homogeneous pattern as they have distinguished characteristic compared with other patterns. Golgi pattern is often mistaken for nucleolar pattern, because some golgi pattern have large speckles within the nucleoli while some only have several cluster of irregular granules, which is just similar to nucleolar pattern. Table 3.6 illustrates the confusion matrix at the image level. The proposed LLDC method obtains the classification accuracy of 90.2% at the image level, which means that 37 slide images are correctly identified while there are 41 slide images in the test set. Nucleolar and nuclear membrane patterns particularly obtain 100% of the image-level accuracy. It is evident that golgi pattern is wrongly classified as nucleolar, which is very common at the cell level.

Table 3.4: Classification performance on the ICIP2013 training dataset.

Algorithm	Cell level	Cell level	Image level
	Classification accuracy	Sensitivity	Classification accuracy
LLC-(sift+ldv)	79.1%	76.2%	90.2%
LLC-(sift+dv)	76.1%	73.6%	87.8%
LLC-ldv	75.1%	72.3%	87.8%
LLC-sift	75.8%	73.3%	87.8%
LSC-(sift+ldv)	77.6%	74.5%	90.2%
LSC-(sift+dv)	75.3%	72.8%	87.8%
LSC-ldv	74.6%	72.1%	87.8%
LSC-sift	75.6%	72.7%	87.8%

Table 3.5: The confusion matrix for the cell level classification on the ICIP2013 training dataset.

	ho	sp	nu	ce	nm	go
ho	84.6%	11.4%	2.2%	0.0%	1.6%	0.2%
sp	8.8%	73.9%	7.8%	8.4%	0.8%	0.3%
nu	1.4%	5.1%	80.5%	6.6%	3.2%	3.2%
ce	1.7%	15.9%	2.4%	79.9%	0.1%	0.0%
nm	3.7%	4.5%	0.6%	0.5%	87.0%	3.7%
go	7.7%	1.4%	35.6%	0.3%	3.9%	51.1%

Table 3.6: The confusion matrix for the image level classification on the ICIP2013 training dataset.

	ho	sp	nu	ce	nm	go
ho	87.5%	12.5%	0.0%	0.0%	0.0%	0.0%
sp	0.0%	87.5%	0.0%	12.5%	0.0%	0.0%
nu	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%
ce	0.0%	12.5%	0.0%	87.5%	0.0%	0.0%
nm	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
go	0.0%	0.0%	50.0%	0.0%	0.0%	50.0%

3.3.4 Discussion

To provide a more comprehensive analysis of the proposed LLDC method, we further evaluate its performance with respect to the number of nearest neighbors for calculating local distance vector and the coding process respectively. It should be noted that, the classification performance evaluated in this section is classification accuracy at the cell level.

Neighbor Number k_{LDV} on Calculating Local Distance Vector: in our proposed method, we firstly introduce a merged manifold M for all the classes. Secondly, we transform the original local features to local distance vectors by searching the nearest k_{LDV} neighbors around the local feature without regard to classes isolated from the local feature. Figure 3.4 and Figure 3.5 show classification accuracy under various values of

Figure 3.4: Classification accuracy of the LLDC method under $k_{LDV} \in \{2, 5, 10, 15, 20, \dots, 40\}$ on ICPR201 dataset.

k_{LDV} for ICPR2012 dataset and ICIP2013 training dataset respectively. Obviously, the proposed LLDC method achieves the best classification performance when $k_{LDV} = 35$ while using LSC coding scheme for ICPR2012 dataset. For ICIP2012 training dataset, $k_{LDV} = 50$ is the best choice while using LLC coding scheme.

Neighbor Number k_{LLC} on LLC Method: we investigate the effect on classification performance under various neighbor number, k_{LLC} , in approximated LLC coding scheme. Figure 3.6 shows the performance using $k_{LLC} \in \{2, 5, 10, 20, 30, \dots, 70\}$. As can be seen, the best classification accuracy is achieved when $k_{LLC} = 5$ and $k_{LLC} = 60$ for ICPR2012 dataset and ICIP2013 training dataset respectively.

Neighbor Number k_{LSC} on LSC Method: with respect to LSC coding strategy, only k_{LSC} nearest neighbors of a local feature are considered in coding procedure. We discuss

Figure 3.5: Classification accuracy of the LLDC method under $k_{LDV} \in \{2, 5, 10, 20, 30, \dots, 70\}$ on ICIP2013 training dataset.

the impact of different k_{LSC} for staining pattern classification performance. Figure 3.7 shows the classification accuracy under $k_{LSC} \in \{2, 5, 10, 15, 20, \dots, 40\}$. Obviously, $k_{LSC} = 10$ is the best choice for ICPR2012 dataset while $k_{LSC} = 30$ is the best for ICIP2013 training dataset.

3.4 Summary

In this chapter, we have presented a promising framework for automatic staining pattern classification of HEp-2 cells to support the diagnosis of specific autoimmune diseases. The LLDC framework can extract more discriminative information and consequently gives better HEp-2 cells classification performance than many existing coding methods. The LLDC method is based on local distance vector which captures discriminative in-

Figure 3.6: Classification accuracy of the LLDC method using LLC strategy under $k_{LLC} \in \{2, 5, 10, 20, 30, \dots, 70\}$.

formation via image-to-class distance. Furthermore, local distance vector improves the classification performance by making adjustments only to the classes found in the local k_{LDV} nearest neighbors around the local features. It can avoid disturbance from isolated classes. Additionally, the distance patterns and the original local features are proven to be complementary to each other. Therefore their concatenation can achieve better classification performance. Experimental results on the ICPR2012 dataset and the ICIP2013 training dataset validate that the proposed LLDC framework can provide superior performance for staining pattern classification, compared with the some improved coding methods.

Compared with traditional coding methods, the LLDC framework is time consuming as it needs to transform original local features to local distance vectors one by one and it is an integration of two kinds of image representations. In the future, we plan to design

Figure 3.7: Classification accuracy of the LLDC method using LSC strategy under $k_{LSC} \in \{2, 5, 10, 15, 20, \dots, 40\}$.

a new model to reduce the algorithm's complexity while improving the accuracy.

Chapter 4

Encoding Rotation Invariant Features in Staining Patterns

In this chapter, we propose two kinds of rotation invariant descriptors to characterize the staining patterns in HEP-2 cells, i.e., Pairwise Local Ternary Patterns with Spatial Rotation Invariant (PLTP-SRI) feature and the BoW representation based on dense SIFT. The descriptors are concatenated to achieve the advantages of both highly descriptive ability and robustness against rotations. Incorporated with a linear SVM classifier, our proposed approach demonstrates its effectiveness by testing on two HEP-2 cells datasets: the ICPR2012 dataset and the ICIP2013 training dataset. Particularly, it shows superior classification performance than the best performer at the first edition of the HEP-2 cell classification contest.

4.1 Introduction

The characterization of rotation invariant is significant for representing staining patterns of HEP-2 cell images, because the cells are not characterized by principal directions which are useful for aligning the HEP-2 cell images. To improve the classification performance, we propose two kinds of rotation invariant descriptors to characterize HEP-2 cells that are highly discriminative and descriptive with respect to their staining patterns.

We firstly propose a rotation invariant textural feature of PLTP-SRI. The Local Binary Pattern (LBP) operator had become well known as a simple and effective textural descriptor. In recent years, LBP-related features have been applied in staining pattern classification of HEp-2 cells. For example, Nosaka et al. [30] utilize an extension of LBP descriptor, named Co-occurrence of Adjacent LBP (CoALBP), to extract textural features. CoALBP has a high descriptive ability and retains the advantages by considering the spatial relations among adjacent LBPs. The method won the first prize in HEp-2 cells classification contest. To further improve classification performance of CoALBP, Nosaka et al. [36] propose Rotation Invariant Co-occurrence among adjacent LBP (RIC-LBP) which is highly descriptive and invariant to image rotation. Our group [37] construct a dual spatial pyramid structure on a rotation invariant texture feature to take resolution variations into account and capture spatial layout information of the HEp-2 cell images. Theodorakopoulos et al. [39] propose a new descriptor based on Gradient-oriented Co-occurrences of LBPs (GoC-LBPs). Then it is fused with the distribution of SIFT features into a dissimilarity space. Nanni et al. [42] utilize a pyramid multi-scale representation coupled with a multi-resolution LBP to represent the HEp-2 cell images.

However, the intensity gradients of our HEp-2 cells are weak, especially in the intermediate intensity cells, as shown in Figure 4.1. LBP related features are sensitive to noise and smooth weak illumination gradients since they threshold at the gray value of the center pixel [73]. To solve the problem, we replace the binary patterns by three-value patterns which is more efficient than LBP for such a specific classification task. Furthermore, we propose a spatial pyramid structure based on patch-level rotation invariant Local Ternary Pattern (LTP) to capture spatial layout information.

Secondly, we apply a BoW framework for representing the HEp-2 cell images. The BoW framework is one of the most successful image classification framework [56, 57]. Our previous work verified its effectiveness in HEp-2 cell classification [34, 35]. It presents an

Figure 4.1: HEP-2 slide images with positive and intermediate fluorescent intensity.

image in terms of a set of visual words, selected from a trained beforehand codebook. By utilizing a BoW model, we can largely avoid the affection of position and orientation of images, and get a high discrimination rate. To enhance the discrimination power, we further add a spatial pyramid structure to retain spatial information.

As an optimization, the aforementioned two descriptors are integrated into a discriminative and descriptive image representation. Both features are respectively effective for capturing informative characteristics of the staining patterns in their own ways. While our proposed PLTP-SRI feature extracts local feature, BoW builds a global image representation. It is reasonable to extract multiple features for compensation. The combined feature can take the advantages of the two kinds of features from different aspects. Furthermore, we demonstrate the validity of the proposed feature by experimental results consistently.

4.2 Proposed Method

In this section, we focus on the innovation of key techniques for our proposed classification framework for identifying staining patterns of HEP-2 cells.

4.2.1 Pairwise LTPs with Spatial Rotation Invariant

Histogram-based features describe an image as an orderless collection of “patterns” occurrence frequency, ignoring spatial layout information. This seriously limits descriptive ability especially for shape of objects in the image. Inspired by the SPM [59], we propose to construct a spatial pyramid structure on the feature space of the HEp-2 cell image.

Firstly, histogram-based features are extracted from small overlapped patches within an image. Then, the image is partitioned into increasingly finer spatial subregions over the feature space. Let $\ell = 1, 2, \dots, L$ denote level of subpartition, such that there are $2^{\ell-1} \times 2^{\ell-1}$ subregions at level ℓ . At level ℓ , the features within each subregion are combined together as

$$\mathcal{H}_i^\ell = \mathcal{F}(\mathbf{h} | h_j \in \mathcal{I}_i^\ell), \quad (\text{Eq. 4.1})$$

where $\mathcal{I}_i^\ell$ is the i th subregion at level ℓ and $\mathcal{H}_i^\ell$ is the corresponding image feature vector in $\mathcal{I}_i^\ell$. $\mathbf{h} = [h_1, h_2, \dots, h_N]^T \in \mathbb{R}^{N \times Q}$ is the patch-level features and $h_i \in \mathcal{I}_i^\ell$ denotes the features within $\mathcal{I}_i^\ell$. $\mathcal{F}(\cdot)$ is a specific statistics method aggregating occurrences of histogram-based features. In this thesis we adopt the *max-pooling* strategy:

$$\mathcal{H}_{ik}^\ell = \max(|h_{jk}| | h_j \in \mathcal{I}_i^\ell), \quad (\text{Eq. 4.2})$$

where the “max” function is a row-wise manner, $\mathcal{H}_{ik}^\ell$ is the k -th element of $\mathcal{H}_i^\ell$ and h_{jk} is the k -th element of h_j .

Within the spatial pyramid structure, we extract a new rotation invariant textural feature. As aforementioned, LBP is a simple yet effective textural feature. However, the LBP tends to be sensitive to noise and smooth weak intensity gradients, because it thresholds at the gray value of the central pixel [73]. Therefore, the LBPs are extended

to LTP defined as

$$LTP_{P,R}(x, y) = \sum_{i=0}^{P-1} 2^i t(I(x_i, y_i) - I(x, y), th), \quad (\text{Eq. 4.3})$$

$$t(x, th) = \begin{cases} 1 & , & x \geq th \\ 0 & , & |x| < th \\ -1 & , & x \leq -th \end{cases} \quad (\text{Eq. 4.4})$$

where $I(x_i, y_i)$ is the gray value of P equal spaced pixels on a circle of R around (x, y) , $(x_i, y_i) = (x + R \cos(2\pi i/P), y + R \sin(2\pi i/P))$ is the neighbors location and th is a user-specified threshold value. Usually, each ternary pattern is slipped into a positive pattern and a negative pattern as

$$t_p(x, th) = \begin{cases} 1 & , & x \geq th, \\ 0 & , & x < th. \end{cases} \quad (\text{Eq. 4.5})$$

$$t_n(x, th) = \begin{cases} 1 & , & x \leq -th, \\ 0 & , & x > -th. \end{cases} \quad (\text{Eq. 4.6})$$

The difference between the LBP and LTP encoding procedures is shown in Figure 4.2. The computational complexity for a LBP is mainly based on the number of neighbor pixels P . The complexity for a LTP is almost double that for a LBP. Furthermore, the computational times of the LBP and LTP increase proportionally to the pixel count of the image. In the following procedure, operations are implemented based on positive and negative pattern respectively. LTP partially solve aforementioned problems of LBP by encoding the small pixel difference into a separate state [74] and adding the threshold value. Meanwhile it combines the positive and negative halves making it more discriminative. Following operation is implemented on the positive and negative pattern respectively.

To achieve rotation invariance, we assign a rotation invariant value to each LTP pair which is defined by

$$P_{\vartheta}(\mathbf{x}, \Delta\mathbf{x}_{\vartheta}) = (LTP_{\vartheta}(\mathbf{x}), LTP_{\vartheta}(\mathbf{x} + \Delta\mathbf{x}_{\vartheta})), \quad (\text{Eq. 4.7})$$

Figure 4.2: Difference between LBP and LTP encoding procedures.

where $\mathbf{x} = (x, y)$ is the position vector in image I and $\Delta\mathbf{x}_\vartheta = (d \cos \vartheta, d \sin \vartheta)$ is a replacement vector between a LTP pair based on the rotation angle ϑ . It is noted that one LBP has two patterns, i.e., t_p and t_n . Therefore, the rotation invariant values are assigned to t_p and t_n respectively. $LTP_\vartheta(\mathbf{x})$ is the LTP at position $\mathbf{x}$ with the rotation angle ϑ , which can be rewritten as

$$LTP_\vartheta(\mathbf{x}) = \sum_{i=0}^{P-1} 2^i t(I(\mathbf{x} + \Delta\mathbf{r}_{i,\vartheta}) - I(\mathbf{x}), th), \quad (\text{Eq. 4.8})$$

where $I(\mathbf{x} + \Delta\mathbf{r}_{i,\vartheta})$ is the gray value of P neighboring pixels around center pixel with respect to ϑ and $\Delta\mathbf{r}_{i,\vartheta} = (R \cos(2\pi i/P + \vartheta), R \sin(2\pi i/P + \vartheta))$ is a replacement vector from the center pixel to neighboring pixels in a LTP.

Then, the same value is obtained by $P_\vartheta(\mathbf{x}, \Delta\mathbf{x}_\vartheta)(\vartheta = 0, \pi/4, \pi/2, 3\pi/4, \pi)$ since their LTP pairs are rotational equivalent. We show that the pairwise LTPs can achieve rotation invariance in Figure 4.3. For the rotation equivalence class ‘A’, all the LTP pairs obtain the same value as each of them is equal to the others in terms of rotation; the class ‘B’ is also the same. Particularly, the pairwise LTPs of ‘B’ can be obtained accordingly

Figure 4.3: An example of the rotation equivalence class. Black and white circles correspond to ‘0’ and ‘1’ respectively. $s(\vartheta)$ is the start point of the binary sequence, where $s(\vartheta) = (x + R\cos(\vartheta), y + R\sin(\vartheta))$.

from that of ‘A’ by rotating 180 degrees. Therefore, we define that the pairwise LTPs in Figure 4.3 have the same rotation invariant value, that is $P_{\vartheta} = P_{\vartheta+\pi}$.

We calculate the histogram $h_{R,d}$ of rotation invariant values for every ϑ and (R, d) from every patch within an image. Based on the experiments, we choose $P = 4$ for enough accuracy with an affordable costs in computation and memory. The variation of computational cost and memory affected by the choice of parameters R and d is minor. To improve discrimination, the patch-level rotation invariant textural feature $h = \{h_{R,d}\}$ is obtained by combining $h_{R,d}$ with various (R, d) . This framework can be illustrated in Figure 4.4. Firstly, the image is converted to grayscale image. Then, the grayscale image is partitioned into equal sized patches. The pairwise LTPs with rotation invariant are extracted from each patch. Next, the grayscale image is divided into a sequence of

Figure 4.4: Framework of pairwise LTPs with spatial rotation invariant.

increasingly finer grids over the feature space. Within each grid, the extracted features are integrated using the max-pooling strategy. Finally, all the pooled features from the grids are concatenated together for final classification. Our proposed PLTP-SRI feature is rotation invariance. Meanwhile, it obtains strong descriptive and discriminative power.

4.2.2 Encoding the SIFT Features with BoW Framework

In this study, we exploit the BoW framework to represent the staining patterns of HEp-2 cells. The BoW framework has been introduced with detail in Section 3.2.1. Briefly, we firstly divide the HEp-2 cell image into small overlapped patches. Secondly, we extract the SIFT features which are invariant to scaling and rotation, and partially invariant to illumination change, viewpoint change and noise. These properties are advantageous in staining pattern classification as most of HEp-2 cell images present large orientation variations and have high intra-class variability. The patch-level SIFT features are extracted from patches within an image, i.e., $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$. N is the number of patches of the image and the dimension of SIFT features is 128.

Thirdly, all the patch-level SIFT features of the images for training are clustered to generate a codebook $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M] \in \mathbb{R}^{D \times M}$. Generally, a larger number of visual words can achieve a higher classification accuracy but needs higher computational cost.

Then, by using a specific coding method, SIFT features of each image can be transformed to a collection of feature codes, i.e., $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N\} \in \mathbb{R}^{M \times N}$. Finally,

the image is divided into increasingly finer spatial regions. Multiple codes from each subregion are pooled together by the max-pooling strategy to retain spatial information of the visual words' location. The final representation is generated by concatenating the histograms from all subregions together.

4.3 Experiments and Analyses

In order to evaluate the efficiency of our proposed algorithm, we use two publicly available HEp-2 cells datasets as described in Section 1.3.

4.3.1 Experiment Setup

In preprocessing, HEp-2 cell images are converted to grayscale images. Our automatic classification system firstly extracts two kinds of image descriptors: PLTP-SRI and SIFT-BoW. With respect to the parameters of the number of LTP neighbor pixels P , we choose $P = 4$. The dimension of PLTP-SRI is $S \cdot 2^P(2^P + 1)$, where S is a constant related to the following spatial pyramid structure, so the higher P needs much more memory and time to calculate. We extract PLTP-SRI from each patch of the image with different parameters. The scale of LTP radius, R , the interval of LTP pairs, d , the threshold values, th , and the spatial pyramid scheme, ℓ , are four important parameters to be considered. We will study the influence of these parameters for staining pattern classification in Section 4.3.4.

Dense SIFT features are extracted at single-scale from densely located patches of grayscale images. The patches are centered at every six pixels and with a fixed size of 18×18 pixels. The codebook is generated by using k-means clustering based on the patch-level SIFT features of all the training images. Based on the pre-trained codebook, SIFT features are quantized to codes by some specific coding algorithm as mentioned in Section 3.2.1. LSC algorithm is chosen in our experiments due to its computationally

Table 4.1: Parameters for comparative algorithms.

Algorithm	Parameters ($\mathbf{P}, \mathbf{R}$) or ($\mathbf{P}, \mathbf{R}, \mathbf{d}$)
Our algorithm	(4,1,2),(4,2,4),(4,4,8)
CoALBP	(4,1,2),(4,2,4),(4,4,8)
RIC-LBP	(4,1,2),(4,2,4),(4,4,8)
CLBP	(8,1),(12,2),(16,3)
LBP	(8,1),(12,2),(16,3)

efficiency and superior performance for HEp-2 cells classification. With respect to the spatial pyramid structure on BoW framework, we choose three levels $\ell = (1, 2, 3)$, which is a common choice in the academic field. The images are divided into three increasing finer subregions, i.e., $2^0 \times 2^0$, $2^1 \times 2^1$ and $2^2 \times 2^2$. In each subregion, we employ max-pooling strategy to illustrate the characteristic in the corresponding feature space.

To evaluate the performance of the proposed algorithm, we compare it with CoALBP, which is the best performing algorithm in the ICPR'12 contest [12], RIC-LBP [75], completed LBP (CLBP) [76] and multi-resolution LBP [47]. Table 4.1 shows the parameters used for each algorithm in our experiments. Additionally, we evaluate SIFT-BoW and PLTP-SRI which are components of our proposed algorithm.

We report classification accuracy at the cell level and at the image level to get comprehensive assessment. In our experiments, we use the performance measures *accuracy* and *sensitivity* at the cell level. At the image level, the classification accuracy is the percentage of slide images correctly classified.

4.3.2 Experimental Results on the ICPR2012 Dataset

We firstly test performance of our automatic staining pattern classification system following the experimental protocol of the ICPR'12 HEp-2 cells classification contest by dividing the ICPR2012 dataset into one training set and one test set. A linear SVM is

Table 4.2: Classification performance on the ICPR2012 dataset.

Algorithm	Cell level accuracy	Sensitivity	Image level accuracy
Our algorithm	75.9%	76.5%	85.7%
PLTP-SRI	70.2%	71.2%	78.6%
CoALBP [12]	68.7%	70.4%	85.7%
SIFT-BoW	68.9%	70.6%	78.6%
RIC-LBP [36]	67.5%	67.6%	71.4%
CLBP	61.2%	61.1%	71.4%
LBP	58.9%	59.2%	64.3%

trained by 10-fold cross validation based on training set. Table 4.2 shows the classification performance of each algorithm. Our proposed algorithm outperforms all the other algorithms at the cell level, achieving 75.9% of classification rate on individual cells. It is worth noting that our proposed PLTP-SRI feature and the combined feature of PLTP-SRI and SIFT-BoW can respectively outperform CoALBP [30] which is the winner of the contest with accuracy of 68.7%.

Table 4.3 illustrates the confusion matrices presenting the classification performance for each staining pattern at the cell level. It is obvious that cytoplasmic, centromere and homogeneous patterns are classified more accurately than the other patterns. More particularly, cytoplasmic can achieve 100% of classification accuracy. However, the fine speckled pattern is hard to categorize, with 37.7% of classification accuracy. It is easily mis-classified as homogeneous pattern. We find that the differences between fine speckled and homogeneous are subtle, therefore it is difficult to find discriminative features to separate them.

With respect to the classification performance at the image level, we use the majority voting scheme. As can be seen in the Table 4.2, our proposed algorithm and CoALBP achieves the highest classification accuracy of 85.7%. Table 4.4 shows the corresponding

confusion matrix. We can see that the element in the confusion matrix at the image level is similar to that of the confusion matrix at the cell level. It is obviously that one coarse speckled image is wrongly classified into fine speckled class while one fine speckled image is identified as centromere by mistake.

Table 4.3: The confusion matrix for the cell level classification on the ICPR2012 dataset.

	ce	cs	cy	fs	ho	nu
ce	92.0%	0.0%	0.0%	0.7%	0.0%	7.3%
cs	7.9%	70.3%	1.0%	19.8%	1.0%	0.0%
cy	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%
fs	29.0%	1.8%	0.9%	37.7%	30.6%	0.0%
ho	6.7%	2.2%	0.0%	7.2%	82.2%	1.7%
nu	16.5%	2.2%	0.0%	0.0%	4.3%	77.0%

Table 4.4: The confusion matrix for the image level classification on the ICPR2012 dataset.

	ce	cs	cy	fs	ho	nu
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	50.0%	0.0%	0.0%	50.0%	0.0%	0.0%
ho	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
nu	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%

4.3.3 Experimental Results on the ICIP2013 Training Dataset

We evaluate our algorithm on the ICIP2013 training dataset following the way similar to experimental protocol of the ICIP'13 contest. The parameters of each algorithm are the same with those for ICPR2012 dataset as shown in Table 4.1. The classification performance of different algorithms are shown in Table 4.5. Our proposed algorithm achieves the best cell level performance again. Table 4.6 shows the confusion matrix of

Table 4.5: Classification performance on the ICIP2013 training dataset.

Algorithm	Cell level accuracy	Sensitivity	Image level accuracy
Our algorithm	77.1%	74.4%	87.8%
SIFT-BoW	75.6%	72.7%	87.8%
PLTP-SRI	74.6%	73.6%	87.8%
CoALBP	67.1%	65.8%	75.6%
RIC-LBP	66.4%	64.4%	75.6%
CLBP	63.5%	61.2%	73.2%
LBP	60.7%	54.5%	65.9%

Table 4.6: The confusion matrix for the cell level classification on the ICIP2013 training dataset.

	ho	sp	nu	ce	nm	go
ho	78.1%	14.1%	5.1%	0.0%	2.6%	0.1%
sp	4.7%	68.8%	13.4%	12.2%	0.7%	0.2%
nu	1.5%	5.2%	80.6%	7.3%	2.8%	2.6%
ce	0.7%	16.1%	1.9%	78.8%	1.1%	1.4%
nm	3.2%	1.9%	0.9%	0.7%	90.1%	3.2%
go	6.1%	0.8%	37.8%	1.1%	3.9%	50.3%

Table 4.7: The confusion matrix for the image level classification on the ICIP2013 training dataset.

	ho	sp	nu	ce	nm	go
ho	87.5%	12.5%	0.0%	0.0%	0.0%	0.0%
sp	0.0%	75.0%	12.5%	12.5%	0.0%	0.0%
nu	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%
ce	0.0%	12.5%	0.0%	87.5%	0.0%	0.0%
nm	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
go	0.0%	0.0%	50.0%	0.0%	0.0%	50.0%

our proposed algorithm at the cell level. Nuclear membrane pattern gets the highest classification accuracy rate of 90.1%, followed by nucleolar as they have distinguished characteristic compared with other patterns. Speckled and golgi patterns are hard to be

recognized by comparison. Speckled pattern is easily to be categorized into centromere pattern, as they both have a uniform diffuse fluorescence of the entire interphase nuclei. It is sometimes mis-classified into nucleolar pattern, because some nucleolar patterns which are characterized by weakly clustered small granules are similar to speckled pattern. Golgi pattern is often mistaken for nucleolar pattern, because some golgi pattern have large speckles within the nucleoli while some only have several cluster of irregular granules, which is just similar to nucleolar pattern. On the other hand, golgi pattern only contains cells from 4 slide images, which influences its classification performance. Table 4.7 illustrates the confusion matrix at the image level. The proposed algorithm obtains the classification accuracy of 87.8% at image level, which indicates that 36 images are correctly identified while there are 41 images in the test set. Both nucleolar and nuclear membrane patterns particularly obtain 100% image level accuracy. The mistakes happened in the image level classification are just similar to that in the cell level classification.

4.3.4 Discussion

To analyze our proposed algorithm more comprehensive, we further evaluate its performance with respect to four important parameters, that is the scale of LTP radius, R , the interval of LTP pairs, d , the threshold value, th , and the spatial pyramid structure, ℓ . In addition, we investigate into the classification performance affected by the codebook size of the BoW framework. In the following, we use the classification accuracy at the cell level as measurement.

The scale of LTP radius and the interval of LTP pairs: Table 4.8 shows the classification performance of various (R, d) s with fixed $P = 4$. Thus, we choose the combination of LTPs with parameters $(R, d) = (1, 2), (2, 4), (4, 8)$ for both the ICPR2012 dataset and the ICIP2013 training dataset.

Table 4.8: Classification accuracy of PLTP-SRI under various (R, d) s.

(R, d)	Accuracy(ICPR2012)	Accuracy(ICIP2013)
(1,2)	65.6%	69.2%
(2,4)	63.5%	67.8%
(4,8)	61.7%	60.6%
(1,2),(2,4)	69.2%	71.1%
(1,2),(4,8)	65.6%	69.9%
(2,4),(4,8)	62.0%	67.0%
(1,2),(2,4),(4,8)	70.2%	74.6%

Table 4.9: Classification accuracy of PLTP-SRI with different thresholds for ternary pattern calculation.

th	0	1	2	3	4	5
Acc (ICPR2012)	70.2%	67.2%	63.8%	59.1%	57.8%	51.6%
Acc (ICIP2013)	71.5%	73.1%	74.6%	71.0%	70.8%	66.3%

The threshold value: Table 4.9 summarizes the performance of proposed PLTP-SRI algorithm with different thresholding values on the ICPR2012 dataset and the ICIP2013 training dataset respectively. The histograms of the HEp-2 cell images are narrow and centered toward the low side of the gray scale. The difference between the center pixels and the neighbor pixels is very small. Through the optimum value, the extracted feature tends to be less discriminative when the threshold is increasing. It is observed that for the ICPR2012 dataset $th = 0$ is showing better performance, while for the ICIP2013 training dataset $th = 2$ is a better choice.

The spatial pyramid structure of PLTP-SRI: with respect to the spatial pyramid scheme, we evaluate the classification accuracy under various levels as shown in Table 4.10. Obviously, we choose three levels $\ell = (1, 2, 3)$, which means that images are partitioned into three increasing finer subregions, i.e., 1×1 , 2×2 and 4×4 . In each sub-region, we use max-pooling strategy to illustrate the characteristic in the corresponding feature space.

Table 4.10: Classification accuracy of PLTP-SRI with different levels of spatial pyramid.

th	1	2	3	4
Acc (ICPR2012)	61.3%	68.0%	70.2%	66.5%
Acc (ICIP2013)	64.8%	70.2%	74.6%	72.7%

The codebook size of BoW framework: To evaluate the performance of codebook with various sizes, we chose five increasing sizes: 256, 512, 1024, 2048 and 4096. Figure 4.5 represents the impact of codebook size for classification. It can be seen that, the classification accuracy improved with the increment of size from 256 to 1024 but it tends to be consistent for larger size from 1024 to 4096. Noted that due to the limited PC memory, we cannot generate the codebook with the size of 4096 for the ICIP2013 training dataset. Considering both efficiency and accuracy, 1024 is a preferred choice.

Figure 4.5: Classification accuracy of BoW representation with various codebook sizes.

4.4 Summary

In this chapter, we have proposed a novel framework for automatic staining pattern classification of HEp-2 cells. The characteristic of rotation invariant is essential for HEp-2 cell classification, since most of the HEp-2 cells in the image present large orientation variances. To this purpose, we incorporate two kinds of rotation invariant descriptors to represent HEp-2 cells. The first is a rotation invariant textural feature based on ternary patterns and considering spatial layout information. It is highly descriptive, robust and rotation invariant. Additionally, we exploit the BoW framework to encode dense SIFT features. To incorporate spatial information, spatial pyramid structure is introduced to further improve the robustness to rotational variances.

Both features are respectively effective for capturing informative characteristics of the staining patterns in their own ways. While our proposed PLTP-SRI extracts local feature, BoW builds a global image representation. They are fused into a powerful image representation by taking the advantages of the two kinds of descriptors in different aspects. The final classification is performed by a linear SVM classifier.

The proposed algorithm was fully evaluated on two HEp-2 cell datasets: the ICPR2012 dataset and the ICIP2013 training dataset. The effectiveness of the proposed algorithm is proven by the experimental results. In particular, it significantly outperforms the winner of the ICPR'12 contest.

Chapter 5

Adaptive Co-occurrence Differential Texton Space for Staining Pattern Classification

In this chapter, we design a Co-occurrence Differential Texton (CoDT) feature to represent the local image patches of HEp-2 cell image, and a generative model is built to adaptively characterize the CoDT feature space. We further exploit a more discriminant representation for the HEp-2 cell images based on the adaptive partitioned feature space, and then feed the representation into a linear SVM classifier to identify the staining patterns. Two benchmark datasets are used for evaluation on the classification performance of our proposed method.

5.1 Introduction

It is verified that powerful discrimination can be achieved by exploiting the distributions of gray values over extremely compact neighborhoods (starting from as small as 3×3 pixels) [47, 77]. LBP related features have been applied successfully in the staining pattern classification [42, 75]. The main idea is to characterize each microstructure into a binary series by thresholding the gray value of the neighboring pixels by that of the

center. However, some important information is lost since the LBP represents the local structures with only two quantized levels, i.e., 0 and 1 [78]. In this study, we propose a CoDT feature without quantizing over micro-structure to characterize the HEP-2 cells. Furthermore, we capture the spatial relations among the differential micro-texton features to improve the discriminative power of features.

Next, the BoW framework, which is very influential in image classification, has been proposed as a set of visual words selected from a codebook learned beforehand [56, 57]. As we mentioned in Chapter 3, it has been applied in classification of staining patterns and has shown impressive performance [33, 34, 35, 79]. The BoW framework seems suitable for the purpose of staining pattern classification. However, the BoW approach suffers from its own problems to which we would like to explore for the solution: (i) Information loss in feature quantization process is inevitable and affects the performance for good image classification [66]; (ii) The computational cost of histogram generation depends directly on the size of codebook. Since better performance is always obtained with larger vocabulary, the computational cost of the BoW framework is high. (iii) The reason for such a histogram representation has to be optimal for the specific classification task is in question [80].

To handle these problems, we make use of the *Fisher Kernel* (FK) principal [81, 82] for staining pattern classification to avoid aforementioned problems of the BoW framework. It characterizes an image by a gradient vector derived from a generative model of the training data, thus the resulting representation is adapted to the classification task. Meanwhile, with the same size of vocabulary, the FK based method can extract much larger image representation than the BoW representation. Hence, it can provide excellent performance with simple linear classifier. For staining pattern classification of HEP-2 cells, we apply a Gaussian Mixture Model (GMM) to adaptively approximate the distribution of our proposed CoDT features. Then we can obtain a high discriminative

and powerful descriptive HEP-2 cell image representation, which is intrinsically adapted to the classification task, based on the adaptive CoDT feature space.

Our major contributions are threefold: (i) A local feature, CoDT, is defined and extracted directly using the differential vectors of micro-texton and its neighborhoods. The information loss caused by quantization of the LBP related features is reduced, and the spatial information is captured simultaneously. Therefore our proposed CoDT feature provides powerful discriminative capability; (ii) The CoDT feature space is adaptively characterized as a generative model, GMM. Thus the parameters can be adjusted from the training cell images which are better fitting the CoDT feature space; (iii) The final image representation, which is derived from the generative model, can cooperate with a simple linear SVM classifier for identifying staining pattern of HEP-2 cells. Our proposed framework (AdaCoDT) can exploit the advantages of both generative and discriminative approaches for image classification. Experimental results verify that our proposed method can provide much better performance of staining pattern classification than that of the traditional LBP and its extensions. It also improves the classification performance in comparison of the BoW representation.

5.2 Proposed Method

In this section, we firstly propose a CoDT feature to represent the local structure of the HEP-2 cell images. Then we exploit the image representation based on the adaptive CoDT feature space modeled by a GMM.

5.2.1 Co-occurrence Differential Texton

LBP can be obtained by thresholding the gray value of the circularly symmetric surrounding pixels with that of the center pixel within a local patch (micro-texton). We

have given the details on the LBP in Section 2.2.1.1. Here we just give the formula for calculating the original LBP operator. Given a grayscale image I , $I(x, y)$ is a gray value at location (x, y) in I . Mathematically, the LBP at location (x, y) is defined as

$$LBP_{P,R}(x, y) = \sum_{i=1}^P 2^{i-1} \text{sign}(I(x_i, y_i) - I(x, y)), \quad (\text{Eq. 5.1})$$

$$\text{sign}(x) = \begin{cases} 1 & , \text{ if } x \geq 0 \\ 0 & , \text{ otherwise.} \end{cases} \quad (\text{Eq. 5.2})$$

where (x, y) is the location of center pixel and (x_i, y_i) is the neighbors location. The LBP encoding process on a 3×3 microtexton is illustrated in Figure 5.1. The histogram of LBP values over the entire image is exploited as a kind of highly discriminative textural feature.

Figure 5.1: The LBP encoding procedure on a 3×3 microtexton.

Due to the computational simplicity and discriminative capability, LBP and its extended versions have been applied in many fields of computer vision e.g., face recognition [83], texture classification [76] and image retrieval [84]. Recently, some improved LBP features such as CoALBP [30] and RIC-LBP [36] have been applied in HEp-2 cells and shown superior performances compared with the conventional LBP. However, one major drawback of the LBP related methods is that they will lose some discriminant information since they represent the microstructure with only two quantized levels. Thus, the histogram of the corresponding LBP related values is not adequately descriptive for representing the image. We give an example in Figure 5.2. Local patch ‘A’ and local

patch ‘B’ is very different, but their LBP values are the same; ‘B’ and ‘C’ has similar local structure, but their LBP values are different.

Figure 5.2: Comparison between the LBPs for similar local structure and for different local structure.

In order to reserve more discriminant information, we propose to use a *Differential Vector* (DV) to describe the cell images. A DV is a microstructural feature based on the differential domain skipping the quantization procedure, which can be formulated as

$$DV_{P,R}(x, y) = (I(x_1, y_1) - I(x, y), \dots, I(x_P, y_P) - I(x, y)). \quad (\text{Eq. 5.3})$$

where the notations follow Eq. 5.1. To enhance the discriminative power, we further propose a CoDT feature capturing the spatial relation between differential micro-textons. Spatial co-occurrence features can characterize more subtle and complex structure than

a single feature. Thus, CoDT feature can provide more information than individual DV. The CoDT feature consisting of one pair of DVs is illustrated in Figure 5.3 and formulated as

$$CoDT_{P,R,d}(\mathbf{x}) = [DV_{P,R}(\mathbf{x}), DV_{P,R}^\theta(\mathbf{x}_\theta)], \quad (\text{Eq. 5.4})$$

where $\mathbf{x} = (x, y)$ is the position vector in image I . $\mathbf{x}_\theta = \mathbf{x} + \Delta\mathbf{x}_\theta$ and $\Delta\mathbf{x}_\theta = (d \cos \theta, d \sin \theta)$ is a replacement vector between a DV pair with interval d . $DV_{P,R}^\theta(\mathbf{x}_\theta)$ is the DV at position $\mathbf{x}_\theta$ with the rotation angle θ , which can be calculated as

$$DV_{P,R}^\theta(\mathbf{x}_\theta) = DV_{P,R}(\mathbf{x} + \Delta\mathbf{x}_\theta). \quad (\text{Eq. 5.5})$$

In this thesis, we extract four pairs of DVs, that is $\theta = 0^\circ, 45^\circ, \dots, 135^\circ$. While the dimension of DV is P , the final CoDT feature will be a $5P$ -dimensional feature vector, which provides more variant and complex image patterns than the single DV. Extracting more pairs of DVs with $\theta = 180^\circ, 225^\circ, \dots, 315^\circ$ is unnecessary as it will bring in some redundant information and is helpless for improving classification performance.

Figure 5.3: Illustration of co-occurrence differential texton. (a) an example of differential vector. (b) two pairs of DVs with rotation angles 0° and θ respectively. The number of neighboring pixel of DV is set as $P = 8$.

5.2.2 Adaptive CoDT Feature Space

To combine the strengths of generative and discriminative approaches for image classification, we characterize our proposed CoDT features of a HEp-2 cell image by a gradient vector derived from a generative model of the training data to model the generation process of the features, and then we feed the output image representations into a discriminative classifier for the identification of HEp-2 cells.

Let $X = \{x_n, n = 1, 2, \dots, N\}$ be a set of samples from the CoDT feature space of one HEp-2 cell image. The probability density distribution of the CoDT feature, which is used to model the generative process of elements in the feature space, is described by a GMM. Let $\lambda = \{w_t, \mu_t, \Sigma_t, t = 1, 2, \dots, T\}$ denote the parameters of the GMM, where w_t , μ_t and Σ_t respectively denotes the mixture weight, mean vector and covariance matrix of Gaussian t . Then we can formulate

$$\begin{aligned} p(x_n|\lambda) &= \sum_{t=1}^T w_t p_t(x_n|\lambda), \\ \text{s.t. } \sum_{t=1}^T w_t &= 1, \end{aligned} \tag{Eq. 5.6}$$

and $p_t(x_n|\lambda)$ is the Gaussian t defined as

$$p_t(x_n|\lambda) = \frac{\exp\{-\frac{1}{2}(x_n - \mu_t)\Sigma_t^{-1}(x_n - \mu_t)\}}{(2\pi)^{D/2}|\Sigma_t|^{1/2}}, \tag{Eq. 5.7}$$

where D is the dimension of the CoDT feature. Actually, the GMM, which models the generation process of the CoDT features, can be regarded as a probabilistic codebook/vocabulary [85]. The parameters of GMM can be adaptively estimated by the Expectation Maximization (EM) algorithm [86] based on the training CoDT feature space. Briefly, EM algorithm can be implemented by the following two steps:

Expectation step (E step): the posteriori probability for each training data x_n is given by

$$\begin{aligned}\xi_{n,t} &= \frac{w_t p_t(x_n|\lambda)}{\sum_{j=1}^T w_j p_j(x_n|\lambda)} \\ &= \frac{\frac{w_t}{(2\pi)^{D/2}|\Sigma_t|^{1/2}} \exp\{-1/2(x_n - \mu_t)^\top \Sigma_t^{-1}(x_n - \mu_t)\}}{\sum_{j=1}^T \frac{w_j}{(2\pi)^{D/2}|\Sigma_t|^{1/2}} \exp\{-1/2(x_n - \mu_j)^\top \Sigma_j^{-1}(x_n - \mu_j)\}}.\end{aligned}\quad (\text{Eq. 5.8})$$

Maximization step (M step): the means, variances and mixture weights are updated by

$$\mu_t = \frac{\sum_{n=1}^N \xi_{n,t} x_n}{\sum_{n=1}^N \xi_{n,t}}, \quad (\text{Eq. 5.9})$$

$$\Sigma_t = \frac{\sum_{n=1}^N \xi_{n,t} (x_n - \mu_t)(x_n - \mu_t)^\top}{\sum_{n=1}^N \xi_{n,t}}, \quad (\text{Eq. 5.10})$$

$$w_t = \frac{1}{N} \sum_{n=1}^N \xi_{n,t}. \quad (\text{Eq. 5.11})$$

Each Gaussian component can be treated as a micro-texton word of the micro-texton vocabulary while w_t corresponds to the relative frequency of word t , μ_t represents the mean of the word and Σ_t is the variation around the mean. In the followings, the covariance matrices are assumed to be diagonal and denoted by $\sigma_t = \text{diag}(\Sigma_t)$.

5.2.3 HEp-2 Cell Image Representation in the Adaptive CoDT Feature Space

In the previous section, we model the CoDT feature space of HEp-2 cell images as a GMM, and learn the adaptive parameters $\lambda = \{w_t, \mu_t, \Sigma_t, t = 1, 2, \dots, T\}$ of the GMM.

The samples X can be described by the following gradient vector, a.k.a. *score function*:

$$G_\lambda(X) = \nabla_\lambda \log p(X|\lambda). \quad (\text{Eq. 5.12})$$

The gradient describes how the parameters λ should be justified to best fit the input X . To measure the similarity between two HEP-2 cell images, a *Fisher Kernel* (FK) [82] is calculated as

$$K_F(X, Y) = G_\lambda^T(X) F_\lambda^{-1} G_\lambda(Y), \quad (\text{Eq. 5.13})$$

where F_λ is the *Fisher Information Matrix* (FIM) formulated as

$$F_\lambda = \mathbb{E}_X[G_\lambda(X)G_\lambda^T(X)]. \quad (\text{Eq. 5.14})$$

The superscript T means the transpose of G_λ . Fisher information is a measurement about the amount of information that X carries with respect to parameters λ .

As F_λ is symmetric and positive semi-definite, and F_λ^{-1} can be decomposed as $F_\lambda^{-1} = L_\lambda^T L_\lambda$, the FK can be rewritten as

$$K_F(X, Y) = \mathcal{G}_\lambda^T(X) \mathcal{G}_\lambda(Y), \quad (\text{Eq. 5.15})$$

where

$$\mathcal{G}_\lambda(X) = L_\lambda \nabla_\lambda \log p(X|\lambda). \quad (\text{Eq. 5.16})$$

The normalized gradients with respect to the weights w_t , the mean μ_t and covariance Σ_t also correspond respectively to 0-order, 1st-order and 2nd-order statistics.

Let $\zeta(t)$ denote the occupancy probability of the CoDT feature x_n for the t -th Gaussian:

$$\zeta_n(t) = \frac{w_t p_t(x_n|\lambda)}{\sum_{k=1}^T w_k p_k(x_n|\lambda)}. \quad (\text{Eq. 5.17})$$

It can be also regarded as the soft assignment of x_n to the t -th Gaussian.

To avoid enforcing explicitly the constraints in Eq. 5.6, we use a parameter ϵ_t to re-parameterize the weight parameter w_k following the soft-max formalism, which is defined as:

$$w_t = \frac{\exp(\epsilon_t)}{\sum_{k=1}^T \exp(\epsilon_k)}. \quad (\text{Eq. 5.18})$$

The gradients of a single CoDT feature x_n w.r.t the parameters ϵ_t , μ_t and σ_t of the GMM can be formulated as

$$\nabla_{\epsilon_t} \log p(x_n|\lambda) = \zeta_n(t) - w_t, \quad (\text{Eq. 5.19})$$

$$\nabla_{\mu_t^d} \log p(x_n|\lambda) = \zeta_n(t) \left(\frac{x_n^d - \mu_t^d}{(\sigma_t^d)^2} \right), \quad (\text{Eq. 5.20})$$

$$\nabla_{\sigma_t^d} \log p(x_n|\lambda) = \zeta_n(t) \left[\frac{(x_n^d - \mu_t^d)^2}{(\sigma_t^d)^3} - \frac{1}{\sigma_t^d} \right], \quad (\text{Eq. 5.21})$$

where the superscript d denotes the d -th dimension of the input vector.

Then, the normalized gradients are computed by multiplying the square-root inverse of the diagonal FIM. Let f_{ϵ_t} , $f_{\mu_t^d}$ and $f_{\sigma_t^d}$ be the entry on the diagonal of F corresponding to $\nabla_{\epsilon_t} \log p(x_n|\lambda)$, $\nabla_{\mu_t^d} \log p(x_n|\lambda)$ and $\nabla_{\sigma_t^d} \log p(x_n|\lambda)$ respectively, and calculated approximately as $f_{\epsilon_t} = w_t$, $f_{\mu_t^d} = w_t/(\sigma_t^d)^2$ and $f_{\sigma_t^d} = 2w_t/(\sigma_t^d)^2$. Therefore, the corresponding gradients as follows:

$$\mathcal{G}_{\epsilon_t}(X) = \frac{1}{\sqrt{w_t}} \sum_{n=1}^N (\zeta_n(t) - w_t), \quad (\text{Eq. 5.22})$$

$$\mathcal{G}_{\mu_t^d}(X) = \frac{1}{\sqrt{w_t}} \sum_{n=1}^N \zeta_n(t) \left(\frac{x_n^d - \mu_t^d}{\sigma_t^d} \right), \quad (\text{Eq. 5.23})$$

$$\mathcal{G}_{\sigma_t^d}(X) = \frac{1}{\sqrt{w_t}} \sum_{n=1}^N \zeta_n(t) \frac{1}{\sqrt{2}} \left[\frac{(x_n^d - \mu_t^d)^2}{(\sigma_t^d)^2} - 1 \right]. \quad (\text{Eq. 5.24})$$

The Fisher representation is the concatenation of all the gradients for $d = 1, 2, \dots, D$ dimension of the CoDT feature and for T Gaussians. In our cases, we only consider the gradients with respect to the mean and covariance, i.e., $\mathcal{G}_{\mu_t^d}(X)$ and $\mathcal{G}_{\sigma_t^d}(X)$, since the

gradient with respect to the weights is verified that bring little additional information [80]. Therefore the dimension of the resulting representation is $2DT$. The CoDT features are embedded in a higher-dimensional feature space which is more suitable for linear classification.

To avoid dependence on the sample size, we normalize the final image representation by the size of CoDT features from the HEP-2 cell image, N , i.e., $\mathcal{G}_\lambda(X) = \frac{1}{N}\mathcal{G}_\lambda(X)$. After that, two additional normalization steps [87] are conducted in order to improve the results, that are the power normalization and ℓ_2 -normalization.

Power normalization is performed in each dimension as:

$$f(\mathcal{G}_\lambda(X)) = \text{sign}([\mathcal{G}_\lambda(X)]_d)|[\mathcal{G}_\lambda(X)]_d|^\tau, \quad 0 < \tau < 1. \quad (\text{Eq. 5.25})$$

In this study, we choose the power coefficient $\tau = \frac{1}{2}$. The motivation of power normalization is to “unsparisify” the Fisher representation which becomes sparser while the number of Gaussian components of the GMM is increasing.

ℓ_2 -normalization is defined as:

$$f(\mathcal{G}_\lambda(X)) = \frac{\mathcal{G}_\lambda(X)}{\sqrt{\mathcal{G}_\lambda(X)^T \mathcal{G}_\lambda(X)}}. \quad (\text{Eq. 5.26})$$

Our proposed AdaCoDT method has several advantages over the BoW framework [80, 87]. Firstly, it is a generalization of the BoW framework. The resulting representation is not limited to the occurrences of each visual word. It additionally includes the information about the distribution of the CoDT features. It overcomes the information loss raised by the quantization procedure of the BoW framework. Secondly, it defines a kernel from a generative model of the data. Thirdly, it can be generated from a much smaller codebook and therefore it reduces the computational cost compared with the BoW framework. And lastly, with the same size of vocabulary, it is much larger than the BoW representation. Hence, it assures an excellent performance with a simple linear classifier.

5.3 Experiments and Analyses

In this section, we conduct a series of experiments to validate the efficiency of our proposed method for staining pattern classification on two HEP-2 cell datasets: the ICPR2012 dataset and the ICIP2013 training dataset as described in Section 1.3.

5.3.1 Experiment Setup

We quantitatively compare the classification performance achieved by our proposed method, AdaCoDT, with several LBP related features, namely, conventional LBP [47], CoALBP [30] (the winner of ICPR HEP-2 cell classification contest) and RICLBP [36]. As our proposed AdaCoDT method is a generative version of the BoW framework, we also compare it with BoW representation. LSC algorithm is chosen in our experiments due to its computationally efficiency and superior performance for staining pattern classification [34, 35, 62]. In order to make the evaluation comprehensive, we also use the FK based on dense SIFT (FK-SIFT) and the LSC based on CoDT (LSC-CoDT) as comparison. The protocols are used as the ICPR'12 HEP-2 cells classification contest and the parameters are optimized manually via several trials. The parameters for the comparative methods are set as Table 5.1. Two parameters need to be considered while extracting LBP related features, that is, the number of neighbor pixels, P , and the radius, R . The interval between the LBP pair, d , should also be taken into account for CoALBP and RIC-LBP. The number of GMM components T is another parameter to be considered for the FK based methods. With respect to the LSC method, the codebook size is chosen as 1024 due to the trade-off between classification accuracy and computational cost. The parameters of the proposed AdaCoDT method will be discussed in Section 5.3.4.

Table 5.1: Parameters for comparative algorithms.

Algorithm	$(\mathbf{P}, \mathbf{R})$ or $(\mathbf{P}, \mathbf{R}, \mathbf{d})$	$\mathbf{T}$
FK-SIFT	/	128
CoALBP	(4,1,2),(4,2,4),(4,4,8)	/
RIC-LBP	(4,1,2),(4,2,4),(4,4,8)	/
LBP	(8,1),(12,2),(16,3)	/

Table 5.2: Classification performance on the ICPR2012 dataset.

Algorithm	Average accuracy	Sensitivity
AdaCoDT	75.2%	77.1%
Human [12]	73.3%	/
LSC-SIFT	68.9%	70.6%
LSC-CoDT	66.9%	66.5%
FK-SIFT	67.6%	66.4%
CoALBP [12]	68.7%	70.4%
RIC-LBP [36]	67.5%	67.6%
LBP [47]	58.9%	59.2%

5.3.2 Experimental Results on the ICPR2012 Dataset

We first evaluate our proposed AdaCoDT method on the ICPR2012 dataset. We choose linear SVM classifier due to its effectiveness and efficiency. The linear SVM is trained based on the training set by 10-fold cross validation strategy and tested using the test set. Table 5.2 shows the classification performance of each method. The AdaCoDT method outperforms all the other methods, achieving 75.2% of classification accuracy on individual cells. The obtained classification accuracy is even higher than that of a human expert. It is worth noting that the AdaCoDT method significantly outperforms CoALBP [30] which is the winner of the contest.

Table 5.3 illustrates the confusion matrices presenting the classification performance for each staining pattern at the cell level. It is obvious that cytoplasmic, centromere and homogeneous patterns are classified more accurately than the others. More particularly,

Table 5.3: The confusion matrix for the cell level classification on the ICPR2012 dataset.

	ce	cs	cy	fs	ho	nu
ce	85.9%	8.1%	0.0%	0.0%	0.0%	6.0%
cs	4.0%	75.3%	2.9%	17.8%	0.0%	0.0%
cy	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%
fs	20.2%	3.5%	6.2%	52.6%	17.5%	0.0%
ho	8.3%	2.8%	0.6%	11.1%	73.9%	3.3%
nu	2.2%	0.0%	11.5%	3.6%	7.9%	74.8%

cytoplasmic can achieve 100% of classification accuracy. However, the classification accuracy for fine speckled pattern is only 52.6%. Compared to the cytoplasmic pattern with distinguishable shape and centromere pattern with clear fluorescent dots, speckled pattern and homogeneous pattern has similar characteristic and is hard to find discriminative features to separate.

To evaluate the classification performance at the image level, we report the corresponding confusion matrix in Table 5.4. Our proposed AdaCoDT method obtains the classification accuracy of 85.7%. Centromere, cytoplasmic, homogeneous and nucleolar patterns achieve 100% classification accuracy. The most frequent mistake is existed between fine speckled and centromere pattern, which is common mistake at the cell level.

5.3.3 Experimental Results on the ICIP2013 Training Dataset

In this section, we evaluate the AdaCoDT method on the ICIP2013 training dataset by dividing the dataset into training set (6842 cells) and test sets (6754 cells). The classification performance of different methods at the cell level are shown in Table 5.5. Our proposed AdaCoDT method achieves the best cell-level performance again. Although the AdaCoDT method almost achieves the same classification performance with the LSC-SIFT method, it is worth noting that the size of codebook for the BoW framework is

Table 5.4: The confusion matrix for the image level classification on the ICPR2012 dataset.

	ce	cs	cy	fs	ho	nu
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	50.0%	0.0%	0.0%	50.0%	0.0%	0.0%
ho	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
nu	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%

Table 5.5: Classification performance on the ICIP2013 training dataset.

Algorithm	Average accuracy	Sensitivity
AdaCoDT	75.8%	72.9%
LSC-SIFT	75.6%	72.7%
LSC-CoDT	70.6%	69.8%
FK-SIFT	69.7%	68.3%
CoALBP [12]	67.1%	65.8%
RIC-LBP [36]	66.4%	64.4%
LBP [47]	60.7%	54.5%

1024 while the number of GMM components for the AdaCoDT method is only 128. With the same codebook size (for the AdaCoDT method, the number of GMM components can be seen as the codebook size), the AdaCoDT method significantly outperforms the BoW framework. Table 5.6 shows the confusion matrix of the AdaCoDT method at the cell level. Homogeneous pattern gets the highest classification accuracy rate of 89.5%, followed by nuclear membrane as they have distinguished characteristic compared with other patterns. Speckled and golgi patterns are hard to be recognized by comparison. Speckled pattern is easily to be categorized into centromere pattern, as they both have large speckles throughout the interphase nuclei. The classification accuracy for golgi pattern is only 43.6%. The characteristic of golgi pattern is similar to that of nucleolar

Table 5.6: The confusion matrix for the cell level classification on the ICIP2013 training dataset.

	ho	sp	nu	ce	nm	go
ho	89.5%	3.4%	4.1%	0.3%	2.4%	0.3%
sp	11.6%	66.7%	5.6%	15.1%	0.7%	0.3%
nu	0.8%	8.1%	74.2%	11.7%	2.6%	2.6%
ce	0.5%	22.2%	2.4%	74.7%	0.0%	0.2%
nm	1.1%	2.4%	1.1%	0.4%	88.7%	6.3%
go	6.6%	5.0%	38.2%	0.8%	5.8%	43.6%

Table 5.7: The confusion matrix for the image level classification on the ICIP2013 training dataset.

	ho	sp	nu	ce	nm	go
ho	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
sp	12.5%	75.0%	0.0%	12.5%	0.0%	0.0%
nu	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%
ce	0.0%	25.0%	0.0%	75.0%	0.0%	0.0%
nm	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
go	0.0%	0.0%	50.0%	0.0%	0.0%	50.0%

pattern. On the other hand, there are only 362 golgi pattern cells from 4 slide images in the ICIP2013 training dataset, which influences its classification performance. Table 5.7 illustrates the confusion matrix at the image level. The AdaCoDT method obtains the classification accuracy of 87.8% at image level, which means that 36 slide images are correctly identified while there are 41 slide images in the test set. Homogeneous, nucleolar and nuclear membrane patterns particularly obtain 100% image level accuracy. It is evident that golgi pattern is wrongly classified as nucleolar, which is very common at the cell level.

Table 5.8: Classification performance of the AdaCoDT method with different parameters (P, R, d) .

	(8,1,2)	(16,3,6)	(16,4,8)	(16,5,10)	(24,3,6)
Acc (ICPR2012)	53.9%	70.1%	72.9%	71.7%	72.9%
Acc (ICIP2013)	64.0%	73.1%	74.2%	75.8%	73.1%
	(24,4,8)	(24,5,10)	(32,4,8)	(32,5,10)	(32,6,12)
Acc (ICPR2012)	75.2%	73.2%	74.8%	74.1%	70.8%
Acc (ICIP2013)	74.4%	74.4%	73.6%	73.2%	73.0%

5.3.4 Discussion

To further analyze our proposed AdaCoDT method, we evaluate its performance with respect to the CoDT size (P, R, d) and the number of GMM components T . It should be noted that, the classification performance mentioned in this section means *average accuracy* at the cell level.

The CoDT size (P, R, d) : we study the effect of different parameter (P, R, d) for classification performance. As shown in the Table 5.8, the best classification performance is achieved using parameters $(24, 4, 8)$ and $(16, 5, 10)$ for ICPR2012 dataset and ICIP2013 training dataset respectively.

The number of GMM components: we test the performance under various number of GMM components. Figure 5.4 presents the cell-level classification performance of the AdaCoDT method obtained with respect to increasing number of GMM components. As can be seen, the AdaCoDT method with larger number of components can generally achieve better performance for both HEP-2 cells datasets. The classification performance is steady from 128 to 256 for ICPR2012 dataset while the classification accuracy of ICIP2013 training dataset is slightly declining. In our experiments, we choose $T = 256$ for ICPR2012 dataset and $T = 128$ for ICIP2013 training dataset due to the trade-off between accuracy and memory usage.

Figure 5.4: Classification performance of the AdaCoDT method under different number of GMM components.

5.4 Summary

In this chapter, we have presented a promising framework, AdaCoDT, for automatic staining pattern classification of HEP-2 cells to support the diagnosis of specific autoimmune diseases. It is verified that LBP related features can be successfully applied for classifying the staining pattern of HEP-2 cells based on the recent researches on the staining patterns analysis and classification. However, the LBP quantizes the local structures into only two quantized levels by thresholding the neighboring pixel intensity with that of the center one, thus some important information is lost. To reserve more discriminative information, we propose a CoDT feature which directly adpots the differential vectors of micro-texton and its neighborhoods. It further captures the spatial information between neighboring micro-textons, thus it can provide strong discriminative and

descriptive capability.

In addition, the BoW framework, which is one of the most famous and efficient approaches for image categorization, has been applied in classification of HEP-2 staining patterns and has shown impressive performance. Unfortunately, the BoW approach suffers from its own problems which are the inevitable information loss in feature quantization process and the high computational cost to build BoW representation. To handle these problems, we make use of the FK principal for staining pattern classification. It characterizes an image by a gradient vector derived from a generative process of the training data, thus the resulting representation is adapted to the staining pattern classification task. Meanwhile, with the same size of vocabulary, the FK based method can extract much larger image representation than the BoW representation. Hence, it can provide excellent performance with simple linear classifier.

Within the proposed AdaCoDT framework, we first extract the CoDT feature from each HEP-2 cell image. Then, we approximate the distribution of CoDT feature as a GMM which can adaptively partition the CoDT feature space for the classification task. Finally, we obtain a high discriminative and powerful descriptive HEP-2 cell image representation based on the adaptive CoDT feature space using FK principle. We feed the image representation into a linear SVM classifier to predict staining patterns of the HEP-2 cells. Our proposed AdaCoDT method combines the strengths of generative and discriminative approaches, therefore it can achieve excellent classification performance. Experimental results on the ICPR2012 dataset and the ICIP2013 training dataset validate that the AdaCoDT method can provide superior performance for HEP-2 cells classification, compared with the traditional LBP and its extensions. The new feature encoding method also improves the classification performance in comparison of the BoW representation.

Chapter 6

Conclusion and Future Work

In this thesis, we aimed at improving the performance for classifying staining patterns of HEP-2 cells to support the diagnosis of specific autoimmune diseases. Especially, we focused on extracting suitable and effective features for representing the HEP-2 cells w.r.t the staining patterns. We explored three kinds of image descriptors, from low-level local textural descriptor to mid-level coding feature. Experimental evaluations on two publicly available HEP-2 cells datasets validate the effectiveness of our methods.

Firstly, we researched on the BoW framework which had been applied successfully in the staining pattern classification. We improved the coding method and proposed a promising framework, LLDC, for automatic staining pattern classification. Our proposed method is based on local distance vector which can capture discriminative information. Furthermore, it improves the classification performance by making adjustments only to the classes found in the local few nearest neighbors around the local features. To further improve the classification accuracy, two image representations are concatenated together as the distance patterns and the original local features are proven to be complementary to each other.

Then, we integrated the characteristic of rotation invariant into the textural features. We proposed to extract two kinds of rotation invariant descriptors: the PLTP-SRI fea-

ture and the BoW representation based on dense SIFT. Our proposed method takes the advantages of the two kinds of features from different aspects. It has the advantages of invariance under image rotations, meanwhile it has strong discriminative and descriptive ability. Incorporated with a linear SVM classifier, our method demonstrates its effectiveness by experimental results consistently.

Thirdly, we designed CoDT features to represent HEp-2 cell images. The information loss caused by quantization of the LBP related features is reduced, and the spatial information is captured simultaneously. Therefore our proposed CoDT feature provides powerful discriminative capability. Then, we built a generative model to adaptively characterize the CoDT feature space. We further exploited a more discriminant representation for the HEp-2 cell images based on the adaptive partitioned feature space, and fed the representation into a linear SVM classifier for identifying the staining patterns. Our proposed framework (AdaCoDT) can exploit the advantages of both generative and discriminative approaches for image classification.

In recent years, although considerable progress has been made, research in HEp-2 cell image analysis is still in its early stage and there is great potential for improvement. For future work, we suggest some long-term research goals for staining pattern classification.

Firstly, designing corresponding features is a prerequisite for image classification. It is not easy to find effective features for staining patterns of HEp-2 cells as the distinction between different patterns is subtle. It is even difficult for human beings to separate them. In this thesis, we have proposed three kinds of features to characterizing the staining patterns. However, there are still lots of developed features existed such as BRISK [88], ORB [89], etc. Moreover, the mid-level features which are transformed from low-level features should be researched more deeply. On the other hand, multiple features can be heuristically combined into one high dimensional feature to complement each other. In

addition, feature selection is a procedure to select a subset of the most relevant features from the input data which can describe the input data efficiently and provide good prediction results. It can choose the most discriminative features, reduce the measurement and storage requirements, reduce computational complexity and training time, defy the curse of dimensionality to improve classification performance [90]. There are plenty of feature selection methods available in literatures, e.g., filter methods, wrapper methods, embedded methods [91]. We will further study on extracting and selecting effective and reasonable features to satisfy the need of staining pattern classification. Furthermore, we plan to explore the feature extraction and classification for general images.

Secondly, deep learning, also called hierarchical learning, which is a branch of machine learning, has become a hot topic in recent years. Until now, most machine learning methods had exploited shallow structured architectures such as SVM, logistic regression, multilayer perceptrons (MLPs), etc. However, human information processing mechanisms suggest to construct deep architectures to extract complex structure and build internal representation from rich sensory inputs [92]. It is reasonable that the state-of-the-art can be improved if appropriate deep learning algorithms can be developed. There are plenty of deep architectures, such as Deep Neural Networks (DNNs), Deep Belief Network (DBN), etc. They have been successfully applied to speech recognition [93, 94] and computer vision [95, 96]. However, they have not been developed for staining pattern classification. In the future, we will study deeply on this strong learning method and improve it for staining pattern classification.

Last but not least, we will research on the features and classifiers for fluorescent intensity classification. Some researchers have investigated into the fluorescent intensity classification. For example, Soda et al. [15] extract a set of statistical features, which is based on the first-order and the second-order gray-level histograms, from the whole

image. For intensity classification, a multiple expert approach based on three classifiers is adopted. Each classifier specialized in identifying one of three input classes (i.e., positive, intermediate and negative). Rigon et al. [97] present a comprehensive system supporting the ANA test. It includes two systems respectively classifying the fluorescence intensity and staining pattern. For fluorescence intensity, they use the multi expert/-module system and propose two rules to provide the final classification. However, their researches are evaluated by using private datasets and experimental protocols. Next, we are exploring features and classifiers to improve the performance of fluorescence intensity classification. Our final goal is to build an automatic system to support the doctors' diagnosis which should contain two major tasks, i.e., fluorescent intensity classification and positive staining pattern classification.

References

- [1] A. H. Coons and M. H. Kaplan, “Localization of antigen in tissue cells ii. improvements in a method for the detection of antigen by means of fluorescent antibody,” *The Journal of experimental medicine*, vol. 91, no. 1, pp. 1–13, 1950.
- [2] W. Emlen and L. O’Neill, “Clinical significance of antinuclear antibodies. comparison of detection with immunofluorescence and enzyme-linked immunosorbent assays,” *Arthritis & Rheumatism*, vol. 40, no. 9, pp. 1612–1618, 1997.
- [3] S. VOLLSET, “Evaluation of diagnostic tests for antinuclear antibodies in rheumatological practice,” *Scand. J. Immunol*, vol. 52, pp. 309–315, 2000.
- [4] Y. Kumar, A. Bhatia, and R. W. Minz, “Antinuclear antibodies and their detection methods in diagnosis of connective tissue diseases: a journey revisited,” *Diagn Pathol*, vol. 4, no. 1, pp. 1–10, 2009.
- [5] P. Kern, M. Kron, and K. Hiesche, “Measurement of antinuclear antibodies: assessment of different test systems,” *Clinical and diagnostic laboratory immunology*, vol. 7, no. 1, pp. 72–78, 2000.
- [6] A. P. Nifli, G. Notas, M. Mamoulaki, M. Niniraki, V. Ampartzaki, P. A. Theodoropoulos, M. J. Kopnitsky, and E. Castanas, “Comparison of a multiplex, bead-based fluorescent assay and immunofluorescence methods for the detection of ana and anca autoantibodies in human serum,” *Journal of immunological methods*, vol. 311, no. 1, pp. 189–197, 2006.
- [7] P. L. Meroni and P. H. Schur, “Ana screening: an old test with new recommendations,” *Annals of the Rheumatic Diseases*, 2010.

REFERENCES

- [8] E. Tonutti, D. Bassetti, A. Piazza, D. Visentini, M. Poletto, F. Bassetto, P. Caciagli, D. Villalta, R. Tozzoli, and N. Bizzaro, “Diagnostic accuracy of elisa methods as an alternative screening test to indirect immunofluorescence for the detection of antinuclear antibodies. evaluation of five commercial kits,” *Autoimmunity*, vol. 37, no. 2, pp. 171–176, 2004.
- [9] A. C. of Rheumatology, “Position statement: Methodology of testing for antinuclear antibodies,” www.rheumatology.org/practice/clinical/position/ana_position_stmt.pdf, August, 2011.
- [10] A. Bradwell and R. Hughes, *Atlas of Hep-2 patterns and laboratory techniques*. Binding Site, 2007.
- [11] N. ORDER, “Anti-coxsackie virus ifa (iga, igg or igm) biochip mosaic test system test instruction.”
- [12] P. Foggia, G. Percannella, P. Soda, and M. Vento, “Benchmarking hep-2 cells classification methods.” *IEEE transactions on medical imaging*, vol. 32, no. 10, pp. 1878–1889, 2013.
- [13] P. Soda, “Early experiences in the staining pattern classification of hep-2 slides,” in *20th IEEE International Symposium on Computer-Based Medical Systems*, 2007, pp. 219–224.
- [14] R. Hiemann, T. Büttner, T. Krieger, D. Roggenbuck, U. Sack, and K. Conrad, “Challenges of automated screening and differentiation of non-organ specific autoantibodies on hep-2 cells,” *Autoimmunity Reviews*, vol. 9, no. 1, pp. 17–22, 2009.
- [15] P. Soda and G. Iannello, “A multi-expert system to classify fluorescent intensity in antinuclear autoantibodies testing,” in *19th IEEE International Symposium on Computer-Based Medical Systems*, 2006, pp. 219–224.
- [16] R. Hiemann, N. Hilger, U. Sack, and M. Weigert, “Objective quality evaluation of fluorescence images to optimize automatic image acquisition,” *Cytometry Part A*, vol. 69, no. 3, pp. 182–184, 2006.

REFERENCES

- [17] P. Soda, A. Rigon, A. Afeltra, and G. Iannello, “Automatic acquisition of immunofluorescence images: Algorithms and evaluation,” in *19th IEEE International Symposium on Computer-Based Medical Systems*, 2006, pp. 386–390.
- [18] G. Percannella, P. Soda, and M. Vento, “A classification-based approach to segment hep-2 cells,” in *25th international symposium on Computer-based medical systems*, 2012, pp. 1–5.
- [19] P. Perner, H. Perner, and B. Müller, “Mining knowledge for hep-2 cell image classification,” *Artificial intelligence in medicine*, vol. 26, no. 1, pp. 161–173, 2002.
- [20] P. Foggia, G. Percannella, P. Soda, and M. Vento, “Early experiences in mitotic cells recognition on hep-2 slides,” in *23rd international symposium on Computer-based medical systems*, 2010, pp. 38–43.
- [21] P. Soda, G. Iannello, and M. Vento, “A multiple expert system for classifying fluorescent intensity in antinuclear autoantibodies analysis,” *Pattern Analysis and Applications*, vol. 12, no. 3, pp. 215–226, 2009.
- [22] P. Soda and G. Iannello, “Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 3, pp. 322–329, 2009.
- [23] R. Hiemann, N. Hilger, J. Michel, J. Nitschke, A. Boehm, U. Anderer, M. Weigert, and U. Sack, “Automatic analysis of immunofluorescence patterns of hep-2 cells,” *Annals of the New York Academy of Sciences*, vol. 1109, no. 1, pp. 358–371, 2007.
- [24] A. S. Wiik, M. Høier-Madsen, J. Forslid, P. Charles, and J. Meyrowitsch, “Antinuclear antibodies: a contemporary nomenclature using hep-2 cells,” *Journal of autoimmunity*, vol. 35, no. 3, pp. 276–290, 2010.
- [25] P. Elbischger, S. Geerts, K. Sander, G. Ziervogel-Lukas, and P. Sinah, “Algorithmic framework for hep-2 fluorescence pattern classification to aid auto-immune diseases diagnosis,” in *Biomedical Imaging: From Nano to Macro, 2009. ISBI’09. IEEE International Symposium on*. IEEE, 2009, pp. 562–565.

- [26] T.-Y. Hsieh, Y.-C. Huang, C.-W. Chung, and Y.-L. Huang, “Hep-2 cell classification in indirect immunofluorescence images,” in *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*. IEEE, 2009, pp. 1–4.
- [27] E. Cordelli and P. Soda, “Color to grayscale staining pattern representation in iif,” in *24th International Symposium on Computer-Based Medical Systems*, 2011, pp. 1–6.
- [28] X. Xu, F. Lin, C. Ng, and K. P. Leong, “Staining pattern classification of ana-iif based on sift features,” *Journal of Medical Imaging and Health Informatics*, vol. 2, no. 4, pp. 419–424, 2012.
- [29] A. Wiliem, Y. Wong, C. Sanderson, P. Hobson, S. Chen, and B. C. Lovell, “Classification of human epithelial type 2 cell indirect immunofluorescence images via code-book based descriptors,” in *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*. IEEE, 2013, pp. 95–102.
- [30] R. Nosaka, Y. Ohkawa, and K. Fukui, “Feature extraction based on co-occurrence of adjacent local binary patterns,” in *Advances in Image and Video Technology*. Springer, 2012, pp. 82–91.
- [31] M. Varma and A. Zisserman, “A statistical approach to texture classification from single images,” *International Journal of Computer Vision*, vol. 62, no. 1-2, pp. 61–81, 2005.
- [32] K. Li, J. Yin, Z. Lu, X. Kong, R. Zhang, and W. Liu, “Multiclass boosting svm using different texture features in hep-2 cell staining pattern classification,” in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 170–173.
- [33] L. Shen, J. Lin, S. Wu, and S. Yu, “Hep-2 image classification using intensity order pooling based features and bag of words,” *Pattern Recognition*, vol. 47, no. 7, pp. 2419–2427, 2014.

REFERENCES

- [34] X. Xu, F. Lin, C. Ng, and K. P. Leong, “Linear local distance coding for classification of hep-2 staining patterns,” in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*. IEEE, 2014, pp. 393–400.
- [35] X. Xu, F. Lin, C. Ng, and K. P. Leong, “Automated classification for hep-2 cells based on linear local distance coding framework,” *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 1–13, 2015.
- [36] R. Nosaka and K. Fukui, “Hep-2 cell classification using rotation invariant co-occurrence among local binary patterns,” *Pattern Recognition*, vol. 47, no. 7, pp. 2428–2436, 2014.
- [37] X. Xu, F. Lin, C. Ng, and K. P. Leong, “Dual spatial pyramid on rotation invariant texture feature for hep-2 cell classification,” in *The International Joint Conference on Neural Networks (IJCNN). Killarney, Ireland*. IEEE, 2015 (in process).
- [38] R. Stoklasa, T. Majtner, and D. Svoboda, “Efficient k-nn based hep-2 cells classifier,” *Pattern Recognition*, vol. 47, no. 7, pp. 2409–2418, 2014.
- [39] I. Theodorakopoulos, D. Kastaniotis, G. Economou, and S. Fotopoulos, “Hep-2 cells classification via sparse representation of textural features fused into dissimilarity space,” *Pattern Recognition*, vol. 47, no. 7, pp. 2367–2378, 2014.
- [40] S. Di Cataldo, A. Bottino, I. Ul Islam, T. Figueiredo Vieira, and E. Ficarra, “Sub-class discriminant analysis of morphological and textural features for hep-2 staining pattern classification,” *Pattern Recognition*, vol. 47, no. 7, pp. 2389–2399, 2014.
- [41] X. Kong, K. Li, J. Cao, Q. Yang, and L. Wenyin, “Hep-2 cell pattern classification with discriminative dictionary learning,” *Pattern Recognition*, vol. 47, no. 7, pp. 2379–2388, 2014.
- [42] L. Nanni, M. Paci, and S. Brahnam, “Indirect immunofluorescence image classification using texture descriptors,” *Expert Syst Appl*, vol. 41, no. 5, pp. 2463–2471, 2014.

REFERENCES

- [43] L. Liu and L. Wang, “Hep-2 cell image classification with multiple linear descriptors,” *Pattern Recognition*, vol. 47, no. 7, pp. 2400–2408, 2014.
- [44] Y. Yang, A. Wiliem, A. Alavi, B. C. Lovell, and P. Hobson, “Visual learning and classification of human epithelial type 2 cell images through spontaneous activity patterns,” *Pattern Recognition*, vol. 47, no. 7, pp. 2325–2337, 2014.
- [45] G. Thibault, J. Angulo, and F. Meyer, “Advanced statistical matrices for texture characterization: application to cell classification,” *Biomedical Engineering, IEEE Transactions on*, vol. 61, no. 3, pp. 630–637, 2014.
- [46] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [47] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.
- [48] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, “Local binary patterns for still images,” in *Computer Vision Using Local Binary Patterns*. Springer, 2011, pp. 13–47.
- [49] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [50] H. Zhou, Y. Yuan, and C. Shi, “Object tracking using sift features and mean shift,” *Computer vision and image understanding*, vol. 113, no. 3, pp. 345–352, 2009.
- [51] D. Ai, X. Han, X. Ruan, and Y.-W. Chen, “Adaptive color independent components based sift descriptors for image classification,” in *Proceedings of the 2010 20th International Conference on Pattern Recognition*. IEEE Computer Society, 2010, pp. 2436–2439.
- [52] L. Chang, M. M. Duarte, L. E. Sucar, and E. F. Morales, “A bayesian approach for object classification based on clusters of sift local features,” *Expert Systems With Applications*, vol. 39, no. 2, pp. 1679–1686, 2012.

REFERENCES

- [53] T. Lindeberg, “Scale-space theory: A basic tool for analyzing structures at different scales,” *Journal of applied statistics*, vol. 21, no. 1-2, pp. 225–270, 1994.
- [54] R. Rifkin and A. Klautau, “In defense of one-vs-all classification,” *Mach. Learn. Res.*, vol. 5, pp. 101–141, 2004.
- [55] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” *Information Theory, IEEE Transactions on*, vol. 13, no. 1, pp. 21–27, 1967.
- [56] J. Yang, K. Yu, Y. Gong, and T. Huang, “Linear spatial pyramid matching using sparse coding for image classification,” in *Proc. CVPR*, 2009, pp. 1794–1801.
- [57] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, “Locality-constrained linear coding for image classification,” in *Proc. CVPR*, 2010, pp. 3360–3367.
- [58] A. Wiliem, C. Sanderson, Y. Wong, P. Hobson, R. F. Minchin, and B. C. Lovell, “Automatic classification of human epithelial type 2 cell indirect immunofluorescence images using cell pyramid matching,” *Pattern Recognition*, vol. 47, no. 7, pp. 2315–2324, 2014.
- [59] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2. IEEE, 2006, pp. 2169–2178.
- [60] J. C. van Gemert, J.-M. Geusebroek, C. J. Veenman, and A. W. Smeulders, “Kernel codebooks for scene categorization,” in *Proc. ECCV*. Springer, 2008, pp. 696–709.
- [61] J. C. van Gemert, C. J. Veenman, A. W. Smeulders, and J. Geusebroek, “Visual word ambiguity,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1271–1283, 2010.
- [62] L. Liu, L. Wang, and X. Liu, “In defense of soft-assignment coding,” in *Proc. ICCV*, 2011, pp. 2486–2493.
- [63] K. Yu, T. Zhang, and Y. Gong, “Nonlinear learning using local coordinate coding,” in *Proc. NIPS*, 2009, pp. 2223–2231.

- [64] Y. Huang, K. Huang, Y. Yu, and T. Tan, “Salient coding for image classification,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1753–1760.
- [65] Z. Wu, Y. Huang, L. Wang, and T. Tan, “Group encoding of local features in image classification,” in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 1505–1508.
- [66] O. Boiman, E. Shechtman, and M. Irani, “In defense of nearest-neighbor based image classification,” in *Proc. CVPR*, 2008, pp. 1–8.
- [67] Z. Wang, J. Feng, S. Yan, and H. Xi, “Linear distance coding for image classification,” *IEEE Trans. Image Processing*, vol. 22, pp. 537–548, 2013.
- [68] S. McCann and D. Lowe, “Local naive bayes nearest neighbor for image classification,” in *Proc. CVPR*, 2012, pp. 3650–3656.
- [69] Y.-L. Boureau, J. Ponce, and Y. LeCun, “A theoretical analysis of feature pooling in visual recognition,” in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 111–118.
- [70] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, “Learning mid-level features for recognition,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2559–2566.
- [71] K. Yu and T. Zhang, “Improved local coordinate coding using local tangents,” in *Proc. ICML*, 2010, pp. 1215–1222.
- [72] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “Liblinear: A library for large linear classification,” *Mach. Learn. Res.*, vol. 9, pp. 1871–1874, 2008.
- [73] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Trans Image Process*, vol. 19, no. 6, pp. 1635–1650, 2010.

REFERENCES

- [74] J. Ren, X. Jiang, and J. Yuan, “Relaxed local ternary pattern for face recognition,” in *IEEE Conference on Image Processing (ICIP)*, 2013, pp. 3680–3684.
- [75] R. Nosaka, C. H. Suryanto, and K. Fukui, “Rotation invariant co-occurrence among adjacent lbps,” in *Computer Vision-ACCV 2012 Workshops*. Springer, 2013, pp. 15–25.
- [76] Z. Guo and D. Zhang, “A completed modeling of local binary pattern operator for texture classification,” *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1657–1663, 2010.
- [77] M. Varma and A. Zisserman, “Texture classification: Are filter banks necessary?” in *IEEE computer society conference on Computer vision and pattern recognition*, vol. 2, 2003, pp. II–691.
- [78] X.-H. Han, Y.-W. Chen, and G. Xu, “High-order statistics of weber local descriptors for image representation,” *IEEE Transactions on Cybernetics*, 2014.
- [79] G. Iannello, L. Onofri, and P. Soda, “A bag of visual words approach for centromere and cytoplasmic staining pattern classification on hep-2 images,” in *25th international symposium on Computer-based medical systems*, 2012, pp. 1–6.
- [80] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, “Image classification with the fisher vector: Theory and practice,” *International journal of computer vision*, vol. 105, no. 3, pp. 222–245, 2013.
- [81] T. Jaakkola, D. Haussler *et al.*, “Exploiting generative models in discriminative classifiers,” *Advances in neural information processing systems*, pp. 487–493, 1999.
- [82] F. Perronnin and C. Dance, “Fisher kernels on visual vocabularies for image categorization,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [83] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

REFERENCES

- [84] S. Murala and Q. Jonathan Wu, “Local ternary co-occurrence patterns: A new feature descriptor for mri and ct image retrieval,” *Neurocomputing*, vol. 119, pp. 399–412, 2013.
- [85] F. Perronnin, C. Dance, G. Csurka, and M. Bressan, “Adapted vocabularies for generic visual categorization,” in *Computer Vision-ECCV 2006*. Springer, 2006, pp. 464–475.
- [86] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38, 1977.
- [87] F. Perronnin, J. Sánchez, and T. Mensink, “Improving the fisher kernel for large-scale image classification,” in *Computer Vision-ECCV 2010*. Springer, 2010, pp. 143–156.
- [88] S. Leutenegger, M. Chli, and R. Y. Siegwart, “Brisk: Binary robust invariant scalable keypoints,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2548–2555.
- [89] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: an efficient alternative to sift or surf,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2564–2571.
- [90] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [91] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [92] L. Deng and D. Yu, “Deep learning: methods and applications,” *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [93] L. Deng, G. Hinton, and B. Kingsbury, “New types of deep neural network learning for speech recognition and related applications: An overview,” in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 8599–8603.

REFERENCES

- [94] A. Graves, A.-r. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 6645–6649.
- [95] V. Nair and G. E. Hinton, “3d object recognition with deep belief nets,” in *Advances in Neural Information Processing Systems*, 2009, pp. 1339–1347.
- [96] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [97] A. Rigon, F. Buzzulini, P. Soda, L. Onofri, L. Arcarese, G. Iannello, and A. Afeltra, “Novel opportunities in automated classification of antinuclear antibodies on hep-2 cells,” *Autoimmunity Reviews*, vol. 10, no. 10, pp. 647–652, 2011.

Publications

Journal Papers

- (i) **Xiang Xu**, Feng Lin, Carol Ng and Khai Pang Leong, “Staining Pattern Classification of ANA-IIF Based on SIFT Features”, *J. Medical Imaging and Health Informatics*, vol. 2, no. 4, pp. 419 - 424, 2012.
- (ii) **Xiang Xu**, Feng Lin, Carol Ng and Khai Pang Leong, “Automated Classification for HEP-2 Cells Based on Linear Local Distance Coding Framework”, *EURASIP J. on Image and Video Processing*, vol. 2015, no. 1, pp. 1 - 13, 2015.
- (iii) **Xiang Xu**, Feng Lin, Carol Ng and Khai Pang Leong, “Automatic Antinuclear Antibodies Staining Pattern Classification of HEP-2 Cells Using Rotation Invariant Features”, *Expert Systems with Applications* (under review).

Conference Papers

- (i) **Xiang Xu**, Feng Lin, Carol Ng and Khai Pang Leong, “Staining Pattern Classification of ANA-IIF Based on Locality-constrained Linear Coding Using Supervised Fuzzy Codebook”, in *Proceedings of International Workshop on Advanced Image Technology (IWAIT'13)*, Nagoya, Japan, 7-9 January, 2013 (one of the 15 short-listed papers from 208 papers for the Best Paper Award).
- (ii) **Xiang Xu**, Feng Lin, Carol Ng and Khai Pang Leong, “Linear Local Distance Coding for Classification of HEP-2 Staining Patterns”, in *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV'2014)*, Steamboat Springs CO. 24-26 March, 2014.

REFERENCES

- (iii) **Xiang Xu**, Feng Lin, Carol Ng and Khai Pang Leong, “Dual Spatial Pyramid On Rotation Invariant Texture Feature For HEP-2 Cell Classification”, in *Proceedings of The International Joint Conference on Neural Networks (IJCNN’15)*, Killarney, Ireland, 12-17 July, 2015.
- (iv) **Xiang Xu**, Feng Lin, Carol Ng and Khai Pang Leong, “Encoding Rotation Invariant Features in HEP-2 Cell Classification”, in *Proceedings of IEEE International Conference on Image Processing (ICIP’15)*, Quebec City, Canada, 27-30 Sep, 2015 (Top 10% Paper Award).
- (v) **Xiang Xu**, Feng Lin, Carol Ng and Khai Pang Leong, “Adaptive Co-occurrence Differential Texton Space for HEP-2 Cells Classification”, in *Proceedings of The 18th International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI’15)*, Munich, Germany, 5-9 Oct, 2015.