HEp-2 Cell Classification via Combining Multi-resolution Co-occurrence Texture and Large Regional Shape Information

Xianbiao Qi, Guoying Zhao, *Senior Member, IEEE*, Chun-Guang Li, *Member, IEEE*, Jun Guo,
Matti Pietikäinen, *Fellow, IEEE*

Abstract—Indirect Immunofluorescence (IIF) Imaging of Human Epithelial type 2 (HEp-2) cell image is an effective evidence to diagnose autoimmune diseases. Recently, computer-aided diagnosis of autoimmune diseases by the IIF HEp-2 cell classification has attracted great attention. However, the HEp-2 cell classification task is quite challenging due to large intra-class and small inter-class variations. In this paper, we propose an effective approach for the automatic HEp-2 cell classification by combining multi-resolution co-occurrence texture and large regional shape information. To be more specific, we propose to: a) capture multi-resolution co-occurrence texture information by a novel Pairwise Rotation Invariant Co-occurrence of Local Gabor Binary Pattern (PRICoLGBP) descriptor, b) depict large regional shape information by using an Improved Fisher Vector (IFV) model with RootSIFT features which are sampled from large image patches in multiple scales, and c) combine both features. We evaluate systematically the proposed approach on the IEEE International Conference on Pattern Recognition (ICPR) 2012, the IEEE International Conference on Image Processing (ICIP) 2013 and the ICPR 2014 contest data sets. The proposed method outperforms the winners of the ICPR 2012 contest using the provided experimental protocol. Our method also greatly improves the winner of the ICIP 2013 contest under four different experimental setups. Using the leave-one-specimen-out evaluation strategy, our method with the introduced two features achieves comparable performance with the winner of the ICPR 2014 contest that combined four features.

Keywords—*HEp-2 Cell Classification, PRICoLGBP, Improved Fisher Vector, Multi-resolution Co-occurrence Texture, Large Regional Shape.*

I. Introduction

Indirect Immunofluorescence (IIF) imaging of Human Epithelial type 2 (HEp-2) cell image is an effective evidence to diagnose autoimmune diseases. At the very beginning, the fluorescence patterns were classified mainly by manual labeling. Unfortunately, the process of human labeling requires high expert knowledge and is quite time consuming. The HEp-2 cell classification task is very challenging due to large intra-class and small inter-class variations. As shown in Fig. 1, the
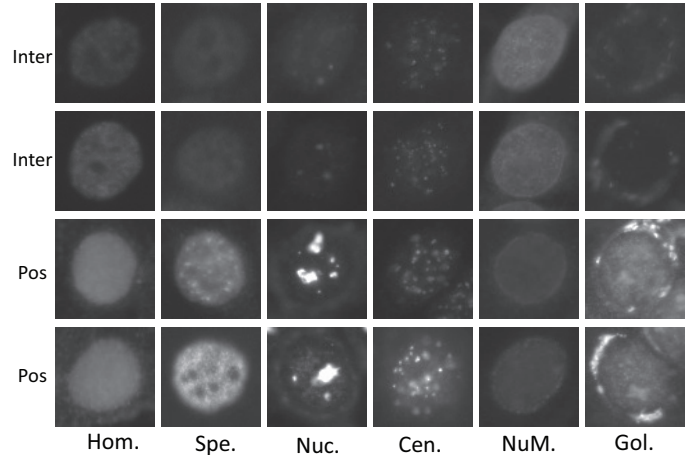
Xianbiao Qi, Guoying Zhao and Matti Pietikäinen are with the Center of Machine Vision Research, University of Oulu, PO Box 4500, FIN-90014, Finland. E-mails: qixianbiao@gmail.com, gyzhao@ee.oulu.fi, mkp@ee.oulu.fi.

Chun-Guang Li and Jun Guo are with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China. E-mails: lichunguang@bupt.edu.cn, guojun@bupt.edu.cn.

Fig. 1: Sample images of all six categories from the ICIP 2013 contest data set with each column from one category. The six categories are "*Homogeneous*", "*Speckled*", "*Nucleolar*", "*Centromere*", "*NuMem*" and "*Golgi*". The first two rows show the "*Intermediate*" samples, and the last two rows show the "*Positive*" samples.

cell images from the same categories (i.e., "*Intermediate*" or "*Positive*") look very different. Moreover, cell images from different categories not only share similar shapes, such as "*Homogeneous*" and "*Speckled*", but also show similar textures, such as "*Nucleolar*" and "*Golgi*". Notice also that the "*Positive*" cells in raw images are perceivable by human eyes; whereas the "*Intermediate*" cells can not be observed clearly.

Recently, computer-aided diagnosis of autoimmune diseases by the IIF HEp-2 cell classification has attracted much attention and hence dozens of pattern recognition techniques have been developed towards automatic diagnosis of the autoimmune diseases [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. Briefly, previous works have shown that the texture and shape cues are very effective in the HEp-2 cell classification task [5], [13], [7], [14], [15], [16]. For instance, Theodorakopoulos et al. [5] proposed to combine the morphological and texture features for the HEp-2 cell classification, in which a uniform Local Binary Pattern (LBP) [17], [18] was used to extract the texture feature. Similarly, Manivannan et al. [16] described the texture information by depicting local

patterns under a Bag of Word (BoW) framework [19], in which the local patterns were combined with other three features. In [14], Theodorakopoulos et al. attempted to capture the shape information by a bundle of local gradient descriptors. In [15], Vestergaard et al. depicted the shape information by using the shape index histograms with dount-shaped spatial pooling, in which the introduced shape index histograms were derived from *shape index* [20].

Although the approaches mentioned above have shown improvements on the HEp-2 cell classification, few of them has exploited the multi-resolution characteristic of the texture in HEp-2 cell images. Besides, while the shape information has been used in previous works, the methods to depict the shape information are not strong enough because most of them model the shape by using a histogram which may fail to accurately reveal the first order and the second order statistics of the distribution in shape descriptors.

In this paper, we attempt to exploit the multi-resolution characteristic of the texture in the HEp-2 cell images and the more effective way to model the distribution of shape descriptors for accurate HEp-2 cell classification. To be more specific, our contributions in this paper are highlighted as follows:

- We propose to capture the multi-resolution texture information by a novel Pairwise Rotation Invariant Co-occurrence of Local Gabor Binary Pattern (PRICoL-GBP) descriptor, which is able to well capture the multi-resolution characteristic of the texture in the HEp-2 cell images.
- We propose an effective way to represent the rich shape information by using an Improved Fisher Vector (IFV) model over the RootSIFT features, which are extracted from *larger* image patches in multiple scales.
- We show that the our multi-resolution texture descriptor and the proposed approach to represent the shape information are complementary.

The rest of the paper is organized as follows. We review the related work in Sec. II and present the proposed texture and shape features in Sec. III. We describe the benchmark data sets and evaluation metrics in Sec. IV, show implementation details and comprehensive evaluations in Sec. V, and finally conclude the paper in Sec. VI.

II. RELATED WORK

In this section, we review the related work of the HEp-2 cell classification from the perspective of feature extraction and combination. Roughly, we divide the existing methods into three categories: a) texture based approaches, b) shape based approaches, and c) approaches based on combining both texture and shape.

A. Texture based Approaches

Texture features are very effective for HEp-2 cell classification. In this line, LBP [17], [18] and its variants [21], [22], [23], [24] have been widely used to capture texture feature for HEp-2 cell classification, e.g., [25], [26], [6], [7].

Among LBP based methods, Co-occurrence of adjacent LBP (CoALBP) [26], [6], Gradient-oriented Co-occurrence of LBPs (GoC-LBPs) [7] and Pairwise Rotation Invariant Co-occurrence of LBP (PRICoLBP) [23] were the three of the best performing LBP variants in HEp-2 cell classification. In [6], Nosaka et al. proposed to use CoALBP for the HEp-2 cell classification and performed the best in the contest for HEp-2 cell classification which was held with the International Conference on Pattern Recognition (ICPR) 2012. In this approach, each image was filtered by a Gaussian function to remove noise and manually rotated with nine orientations (to improve the robustness to rotation), CoALBP features were extracted for all images (both the original images and the manually rotated images), and a linear Support Vector Machine (SVM) was adopted for classification. The success of Nosaka's approach can be attributed to: a) the strong discriminative power of the CoALBP in which 10 templates were used to capture the strong spatial layout information, b) the used green channel, which is much stronger than the red and blue channels, and c) the manually rotated images to augment the training sample. However, the discriminative power of the CoALBP is limited because the CoALBP is built upon the co-occurrence of two LBPs with only four neighbors rather than the co-occurrence of two LBPs with eight neighbors which was exploited in PRICoLBP. Moreover, while the manually rotated images are used to augment the training sample for improving the robustness to rotation, it is still not rotation invariant. Besides, only the green channel was used.

In addition to the methods mentioned above, the original LBP [18], Completed LBP (CLBP) [25], and other famous texture features, e.g., Maximum Response Filter Banks (e.g. MR8) [27], Gray-Level Co-occurrence Matrices (GLCM) [28], Wavelet [29], were also used in the HEp-2 cell classification.

B. Shape based Approaches

Shape feature is another important aspect in HEp-2 cell classification. In [30], Ponomarev et al. attempted to explicitly represent shape feature by simply counting the distribution of the object number, the average object area and the average hole area. However, this method is very sensitive to the variations in cell shape. In [15], Vestergaard et al. introduced a novel second-order donut-like shape index histogram descriptor for the HEp-2 cell classification and won the merit winer of the HEp-2 cell classification contest which was held with the International Conference on Image Processing (ICIP) 2013. In their method, a two-stage preprocessing was used, in which each image I was firstly augmented with its logarithmic representation $\ln(I)$ and the logarithmic representation was mapped linearly into $[0, 1]$, and then three sorts of features were extracted: a) the proposed donut-like shape index histogram, b) the "intersity" of each image (negative/intermediate/positive) as an integer flat, and c) the morphological features over the provided mask (containing the area of the mask region, eccentricity, major and minor axis length, perimeter). The strategies to depict the shape information mentioned above, however, are not strong enough, because the shape is modeled by using a histogram which may fail to accurately reveal the

first order and the second order statistics of the distribution in shape descriptors.

C. Approaches based on Combining both Texture and Shape

The complementary property of texture and shape feature in the HEp-2 cell classification has also been investigated previously, e.g., [8], [5], [31], [16]. In [8], Kong et al. adopted Varma's MR8 method [27] to extract the texture feature in which each local region was normalized at first and the whole image was represented with the BoW approach over the extracted MR8 features, and used a pyramid histogram of oriented gradients (PHOG) [32] to depict the shape information. The texture and shape histogram were weighted and concatenated, and a k-nearest neighbor (kNN) classifier with χ^2 distance was used for classification. In [31], Shen et al. proposed to combine PRICoLBP [23] and BoW with SIFT feature [33] for the HEp-2 cell classification. The dimension of their used PRICoLBP feature is 5,900, in which 10 templates was used and the dimension of PRICoLBP per template is 590. For BoW with SIFT feature, the codebook is of 1024 codewords and generated with k-means algorithm. The two sources of features were concatenated and a linear SVM with square root normalization [34] was used as the classifier. In [16], Manivannan et al. proposed a method based on combining of four different features and ranked the first place in the HEp-2 cell classification contest [35] which was conjuncted with the ICPR 2014. In their method, each image was rotated to four orientations, multi-scale patches were sampled densely, four types of features were extracted, and Locality-constrained Linear Coding (LLC) was adopted for each type of features and each orientation. In total, 16 histograms were obtained to train 16 linear SVMs. In addition, Theodorakopoulos et al. also investigated the combination of different features, e.g., combining GoC-LBPs [7] and a multivariate distribution of SIFT features [5], combining of the morphological features and a bundle of local gradient descriptors [14]. However, the previous works usually exploit the shape features in small region or represent the shape by using only a histogram. Notice that, a small region might not reveal richer shape information and using simply a histogram might fail to depict the shape information sufficient enough.

In this paper, we attempt to exploit the co-occurrence texture in multi-resolution and depict shape information in large region. Different from the previous texture features (e.g. CoALBP, PRICoLBP), we propose to depict the multi-resolution co-occurrence texture information with the help of the Gabor filters. Different from the previous shape features [30], [15], we propose to capture the shape information over a large region by using the Improved Fisher Vector (IFV) model [36].

III. MULTI-RESOLUTION CO-OCCURRENCE TEXTURE AND LARGE REGIONAL SHAPE INFORMATION

A. Multi-resolution Co-occurrence Texture Information

1) **Local Binary Pattern:** The LBP [18] is considered as a simple and effective texture descriptor. For any pixel A in an image, we can compute its LBP pattern by thresholding the pixel values of its circularly symmetric P neighbors with its pixel value. The LBP of the pixel A is defined as:

$$\text{LBP}_{P,R}(A) = \sum_{i=0}^{P-1} s(V_i - V_c)2^i, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases}$$

where P is the number of the neighbors, R is the radius, V_c is the pixel value of the point A, and V_i is the pixel value of the point A's i-th neighbor. Since the $s(V_i - V_c)$ is invariant to monotonic change of illumination, thus the LBP is gray-scale invariant.

The patterns with very few spatial transitions are considered to depict the fundamental image micro-structures. Such patterns are called as "uniform patterns". Ojala et al. [18] defined a uniformity measure for the uniform patterns, the uniformity measure is $\mathcal{U}(\text{LBP}_{P,R}(A)) \leq N$ (N is usually set to 2). The $\mathcal{U}(\cdot)$ can be calculated as:

$$\mathcal{U}(\text{LBP}_{P,R}(A)) = \sum_{i=0}^{P-1} |s(V_i - V_c) - s(V_{i-1} - V_c)|,$$

where the pixel value of V_{-1} is equivalent to the pixel value of V_{P-1}. For example, "11000000" and "10000001" are uniform patterns, and "10000100" and "10101100" are non-uniform patterns.

Rotation invariant LBP (LBP^{ri}) and rotation invariant uniform LBP (LBP^{ru}) are also introduced in [18], where the LBP^{ri} can be defined as:

$$\text{LBP}_{P,R}^{ri}(A) = \min\{\text{ROR}(\text{LBP}_{P,R}(A), i) \mid \forall i \in [0, P-1]\},$$

where $ROR(x, i)$ performs a circularly bit-wise right shift for i times. The $\text{LBP}_{n,r}^{ru}(A)$ is defined as

$$\text{LBP}_{P,R}^{ru}(A) = \begin{cases} \sum_{i=0}^{P-1} s(V_i - V_c), & \mathcal{U}(\text{LBP}(A)) \leq 2 \\ P + 1, & \text{otherwise}, \end{cases}$$

The $\text{LBP}_{8,1}$ has 256 patterns in total, in which 58 patterns are uniform and the rest 198 patterns are non-uniform. Usually, the 198 non-uniform patterns are summarized into one pattern. Thus, usually, 59 patterns are used for the uniform LBP. The rotation invariant uniform $\text{LBP}_{8,1}$ includes ten patterns.

2) **Single-Resolution Texture Information:** Pairwise rotation invariant co-occurrence LBPs (PRICoLBP) is recently introduced by Qi et al. [23] for texture related tasks. As shown in Fig. 2(b), the PRICoLBP is built on the two adjacent LBP points. To calculate the PRICoLBP pattern for any given point A, it contains the following two key steps:

- According to the gradient and the normal orientation of the point A (Normal orientation is the direction that is orthogonal to the gradient orientation.) and the predefined templates as shown in Fig. 2(a), the position of point B_i can be uniquely determined. The gradient orientation can be calculated as $\theta(A) = \arctan(\frac{dy(A)}{dx(A)})$.
- With a pair A and B_i, pairwise rotation invariant encoding was used to encode the co-occurrence of two LBPs.

In practice, the gradient magnitudes of the points A and B_i were used to weight their co-pattern.

In the first step, given a reference point A, the template position of the point B_i, can be determined according to the following equation:

$$B_i = a_i G(A) + b_i N(A) + A, \tag{1}$$

where a_i and b_i are pre-defined coefficients for template i, and $G(A)$ and $N(A)$ are the gradient and the normal directions of the point A. In practice, ten pairs are used for $[a_i, b_i]$ as shown in Fig. 2(a), one pair corresponds to one template.
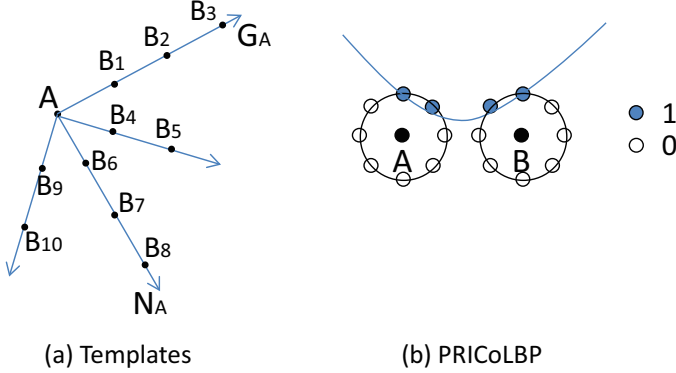

(a) Templates   (b) PRICoLBP

Fig. 2: Pairwise Rotation Invariant Co-occurrence of Local Binary Pattern. The left-hand panel shows ten templates built on the reference point A. G_A is the gradient orientation of the point A, N_A is the normal direction of the point A. The right hand shows one co-occurrence pair of LBPs. The PRICoLBP captures stronger structures than the single LBP.

When the point pair A and B_i are determined, a pairwise rotation invariant encoding strategy is used to encode the pair. Denote $\text{LBP}_{P,R}^u(B, i)$ as the uniform LBP of point B by using i-th index as the start point of the binary sequence. The PRICoLBP can be defined as follows:

$$\text{PRICoLBP}_{P,R}(A, B_i) = [\text{LBP}_{P,R}^{ru}(A), \text{LBP}_{P,R}^u(B_i, i(A))]_{co}, \tag{2}$$

where $i(A) \in \{0, P-1\}$ is an index, which can be determined by minimizing the binary sequence of the point A. $[\cdot, \cdot]_{co}$ is a co-occurrence operator firstly introduced in [28]. Suppose $\text{LBP}_{P,R}^{ru}(A)$ has M patterns, and $\text{LBP}_{P,R}^u(B_i, i(A))$ has N patterns, then their co-occurrence has $M \times N$ patterns.

When $P = 8$ and $R = 1$, the $\text{LBP}_{8,1}^{ru}$ has ten patterns, the $\text{LBP}_{8,1}^u$ has 59 patterns, thus, the dimension of the PRICoLBP$_{8,1}$ is $10 \times 59 = 590$. If ten templates are used, the dimension for the PRICoLBP is $10 \times 590 = 5,900$.

3) Multi-Resolution Texture Information: The PRI-CoLBP is effective to capture the structures in the small scales (such as co-occurrence of $\text{LBP}_{8,1}$ and co-occurrence of $\text{LBP}_{8,2}$). However, multi-resolution texture information is ignored. According to prior knowledge on texture classification, the multi-resolution texture information is highly effective.

To capture the multi-resolution texture information, this paper proposes a novel pairwise rotation invariant co-occurrence of local Gabor binary pattern (PRICoLGBP) descriptor. The
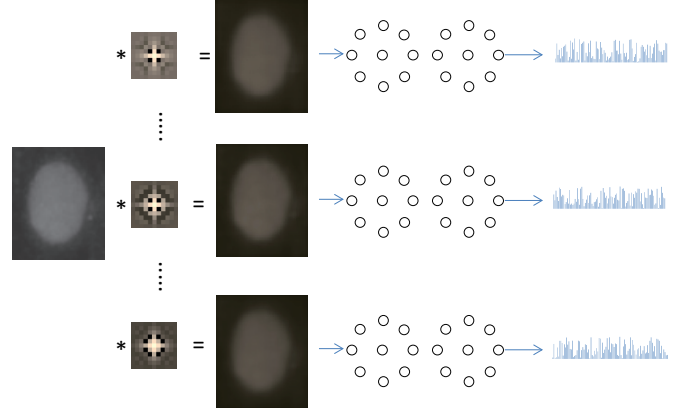

Fig. 3: The framework of PRICoLGBP. First, the image is filtered by Gabor filters with different scales. Then, the PRICoLBP features are extracted from each filtered images. Finally, the features extracted from each scale will be concatenated into the final feature.

original Gabor wavelet [29] [37] is proposed to capture multi-resolution and multi-orientation information. In the HEp-2 cell classification, the multi-resolution texture information is important. Meanwhile, we target an orientation-insensitive image representation. Therefore, we only focus on the multi-scale information and ignore the multi-orientation part of the Gabor filters. In this paper, we use a variant of Gabor filters, which can be denoted as:

$$G(x, y; s_x, s_y, f) = \frac{1}{2\pi s_x s_y} \exp\left(\frac{-1}{2}\left(\frac{x^2}{s_x{}^2} + \frac{y^2}{s_y{}^2}\right)\right) \cos(2\pi f(x^2 + y^2)^{\frac{1}{2}}) \tag{3}$$

where s_x and s_y are the variances of the x-axis and the y-axis respectively, f is the frequency of the sinusoidal function. We remove the orientation factor of the original Gabor filters. The filters of the used variants are radial symmetry instead of the orientation-selectivity in the original Gabor filters.

We illustrate the framework of our PRICoLGBP in Fig. 3. The original image is convolved with different filters, and then the PRICoLBP is extracted from each filtered image. Finally all PRICoLBPs are concatenate into the final feature. The used filters are radially symmetry, and the used PRICoLBP feature is rotation invariant. These two properties entitle the PRICoLGBP good robustness to rotation.

The PRICoLGBP shares some similar properties with the Local Gabor Binary Pattern (LGBP) [38] that is seen as a powerful LBP variants in face recognition. But, different from the LGBP, our PRICoLGBP is built on a more discriminative co-occurrence of LBPs. Thus, we can expect that PRICoLGBP can capture stronger multi-resolution texture information. Two strong properties of the PRICoLGBP makes it effective for the IIF HEp-2 cell classification. Firstly, the texture-based methods prove to be effective in the IIF HEp-2 cell classification. Secondly, the PRICoLGBP inherits the properties (illumination
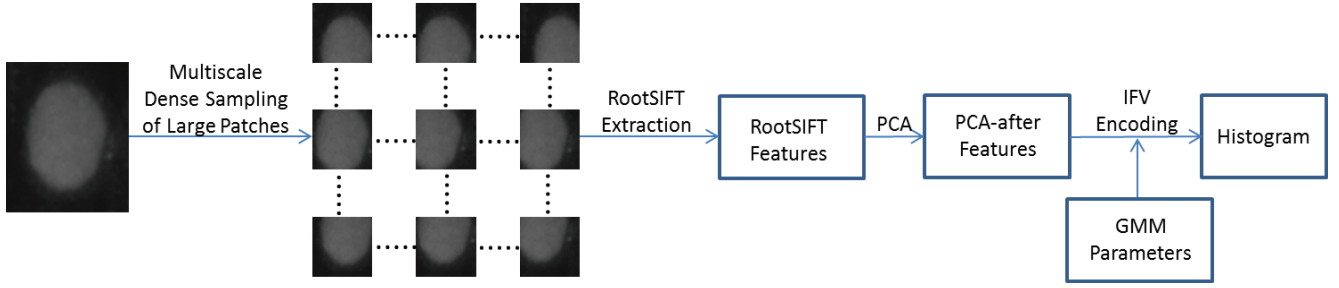
Fig. 4: The framework of discriminative shape information description. First, we densely sample thousands of large patches in multiple scales with a preset step (such as 2), then the RootSIFT feature is extracted for each patch. PCA is applied to the RootSIFT. Finally, IFV encoding is applied to the PCA-after features with pre-learned GMM parameters. The obtained aggregation from IFV encoding can be input into a linear SVM for the final classification.

and rotation robustness) from both Gabor and PRICoLBP. In the IIF HEp-2 cells, the "Positive" and the "Intermediate" cells from the same categories show extremely varying illumination.

B. Large Regional Shape Information

The proposed approach, to depict the richer shape information, consists of three steps: a) patch sampling, b) feature description with RootSIFT, and c) Improved Fisher Vector (IFV) encoding. The flowchart is illustrated in Fig. 4.

1) **Patch Sampling for Depicting Shape Information:** To increase the discriminativeness in shape information, we propose to sample large patches, since that the large patches preserve stronger shape structures. To be more specific, instead of sampling patches of small size, e.g., 16×16, or 19×19 as in object categorization tasks, we sample much larger patches, e.g., 41×41. We can observe in Fig. 4 that the sampled patches cover more than $\frac{1}{3}$ of the whole image.[1]

In Fig. 5, we show some samples of all six categories in the ICIP 2013 contest data set.[2]

- The shape structures from different categories vary a lot. Each category has its own characteristics. For instance, the category "*NuMum*" has bright and thick boundaries, the category "*Centermere*" has many bright spots, and the category "*Golgi*" does not have well-formed boundary. When considering the texture and the shape structures jointly, the difference between some categories is large. For instance, the categories "*Nucleolar*" and "*Centromere*" are easy to differentiate when jointly considering the shape and the texture.
- The "*Positive*" and the "*Intermediate*" HEp-2 cells from the same category have similar shape structures, although the structures of the "*Intermediate*" cells are hardly visible.

These observations are the rationales to explore the shape information for the HEp-2 cell image classification.

---

[1]In general, a HEp-2 cell image is of 65×70 and hence preserve stronger shape structure from the sampled patches.

[2]To visualize the shape structures, we enhance the images by using a logarithmic operator $loge(I)$ on the image and then normalize the image to the range of $[0, 255]$. This preprocessing method was proposed in [15].
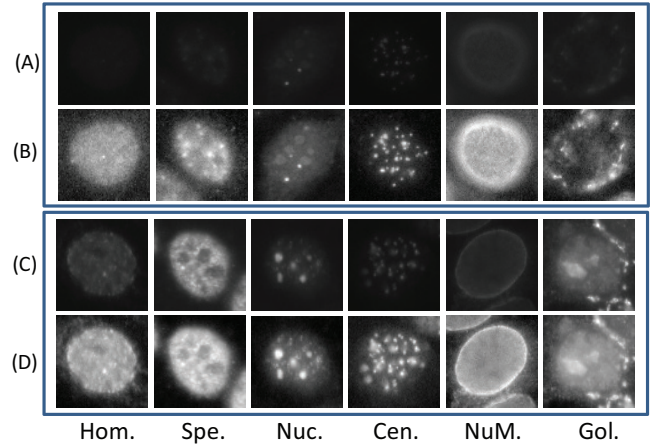

Fig. 5: Illustration of discriminative power of the shape information. Each column comes from one category. It is easy to find that the shapes vary a lot among different categories. (A) shows the "*Intermediate*" cell images from each category, and (B) shows (A)'s corresponding enhanced images by a logarithmic operator $loge(I)$. (C) shows the "*Positive*" cell images, and (D) is (C)'s corresponding enhanced images.

2) **RootSIFT Feature Extraction on Large Patches:** We extract 128-dimensional SIFT features [33] from the sampled large patches. For each feature **f**, we normalize it with L_1-norm and then take the componentwise square root operation, i.e.,

$$\bar{\mathbf{f}} = \frac{\mathbf{f}}{\|\mathbf{f}\|_{\mathbf{1}}},$$
$$\tilde{f}_j = \sqrt{\bar{f}_j}, \quad j = 1, \cdots, 128. \tag{4}$$

The obtained $\tilde{f}$ is termed as "RootSIFT" [39], which was proposed to enhance the discriminative power of the SIFT.

3) **Improved Fisher Vector (IFV) Encoding:** We encode the RootSIFT features by an Improved Fisher Vector (IFV)

approach [36] [40], which consists of three steps:

- Data decorrelation by Principal Component Analysis (PCA).
- Training a Gaussian Mixture Model (GMM).
- Calculating the IFV by using the first and the second order statistics with respect to the GMMs.

Denote the parameters in the GMMs as $\{\pi_k, \mu_k, \Sigma_k\}_{k=1}^K$ where π_k is the membership probability, μ_k is the mean of the k-th component Gaussian, and Σ_k is the covariance matrix which is enforced to be diagonal. Let $X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N] \in \mathcal{R}^{d \times N}$ be a set of feature vectors of an image after the decorrelation, where d is the reduced feature dimension of using the PCA and N is the number of the RootSIFT features in the image. The IFV captures the deviation of the features in an image from the first and the second statistics of the GMM. To be more specific, the IFV is defined as follows:

$$\mathbf{F} = [\mathbf{u_1}^\mathrm{T}, \mathbf{v_1}^\mathrm{T}, ..., \mathbf{u_K}^\mathrm{T}, \mathbf{v_K}^\mathrm{T}]^\mathrm{T}, \tag{5}$$

where

$$\mathbf{u}_k = \frac{1}{N\sqrt{\pi_k}} \sum_{i=1}^N s_{ki} \Sigma_k^{-\frac{1}{2}} (x_i - \mu_k), \tag{6}$$

$$\mathbf{v}_k = \frac{1}{N\sqrt{2\pi_k}} \sum_{i=1}^N s_{ki} [(x_i - \mu_k) \cdot (\Sigma_k^{-1}(x_i - \mu_k)) - \mathbf{1}], \tag{7}$$

in which the s_{ki} is defined as

$$s_{ki} = \frac{p(x_i|\mu_k, \Sigma_k)\pi_k}{\sum_{i=1}^K p(x_i|\mu_i, \Sigma_k))\pi_i}, \qquad k = 1, ..., K. $$

The parameter s_{ki} is the responsibility of feature $\mathbf{x}_i$ belonging to the k-th GMM component.

Note that the dimension of $\mathbf{F}$ is $2 \times d \times K$, where K is the number of the mixture components in GMM. In our experiments, we set d as 80, and K as 256. The final dimension of the IFV representation is $2 \times 80 \times 256 = 41,960$. Note that this is the first time that the IFV is used in the HEp-2 cell classification task.

C. Aggregation Normalization and Classification

Aggregation normalization is a key step before training a SVM model. We normalize the aggregation componentwisely as follows:

$$\hat{H}_i = sign(H_i)\sqrt{\mid H_i \mid}, \forall i = 1, \cdots, D, \tag{8}$$

where D is the dimension of H, $sign(\cdot)$ is a sign function. And then we further normalize the aggregation with L_2 norm.

For classification we use a linear SVM since it is widely used in large scale problems. For the linear SVM, the training is fast and the speed of classification in test phase is also fast. We use the *one-vs-the-rest* strategy to handle the multi-class classification problem.
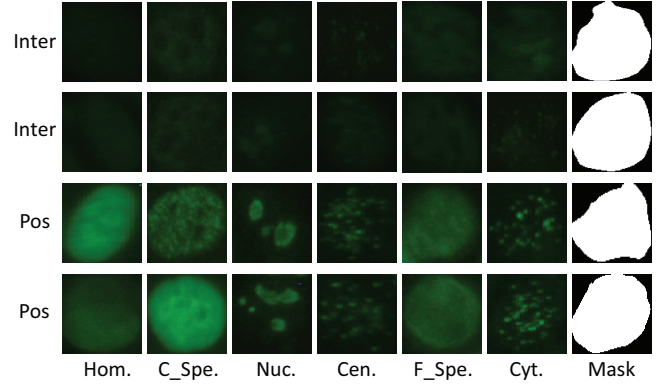

Fig. 6: Sample images from the ICPR 2012 contest dataset with each column from one category. The first two rows show the "*Intermediate*" samples, and the last two rows show the "*Positive*" samples.

TABLE I: Details of the ICPR 2012 contest data set.

	Homo	Coar	Fine	Nucl	Cent	Cyto	Total
Instances/train	3	2	2	2	3	2	14
Cells/train	150	109	94	102	208	60	723
Instances/test	2	3	2	2	3	2	14
Cells/test	180	101	114	139	149	51	734

IV. DATASETS AND EVALUATION METRICS

A. ICPR 2012 Contest Dataset

ICPR 2012 cell images were acquired by means of a fluorescence microscope (40-fold magnification) coupled with a 50W mercury vapor lamp and with a digital camera. The images have a resolution of $1,388 \times 1,038$ pixels, a color depth of 24 bits and they are stored in an uncompressed format. Specialists manually segmented and annotated each cell. In particular, a biomedical engineer manually segmented the cells by the use of a tablet PC. Subsequently, each image was verified and annotated by a medical doctor specialized in immunology. The dataset contains 28 images almost equally distributed with respect to the different patterns. In the contest, the 28 images are divided into training and testing sets. The information for the training and the testing sets is shown in Tab. I. More detailed information can be found in [41]. Some samples are shown in Fig. 6.

TABLE II: Details of the ICIP 2013 contest data set.

	Ho	Sp	Nu	Ce	NM	Go	Total
Specimens	16	16	16	16	15	4	83
Cells	2,494	2,831	2,598	2,741	2,208	724	13,596

B. ICIP 2013 Contest Dataset

The ICIP 2013 data set used 419 patients positive sera with screening dilution 1:80. The specimens were automatically

captured using a monochrome high dynamic range cooled microscopy camera. For each patient serum, 100-200 cell images were extracted. In total, there were 68,429 cell images extracted. The whole 68,429 cell images were divided into 13,596 training samples and 54,833 testing samples.

The labeling process involved at least two scientists who read each patient's specimen under a microscope. A third expert's opinion was sought to adjudicate any discrepancy between the two opinions. In this way, a ground-truth mask can be extracted from each cell image.

The testing images are not released. The training set is big; it contains 83 specimens which contains 13,596 cells in total. Some basic information for the training data in the ICIP 2013 contest are shown in Tab. II. More detailed information can be found in [31]. Some sample images are shown in Fig. 1.

It should be noted that in the ICPR 2014 contest, the Task-1 used the exactly same dataset as the ICIP 2013 contest.

C. Evaluation Metrics

The cells from the same specimen can only be used for training or testing. As pointed out before, each specimen always has 100-200 cells. The cells in the same specimen always have higher similarity than the cells from different specimens. Thus, to evaluate the generalization ability of methods, the cells in one specimen should not be splited into both training and testing. This is because, if part of the cells in one specimen are used for training, and the rest cells in the same specimen, that are used for testing, are too easily correctly classified. In this way, whereas this strategy usually obtains a high validation performance, it can not be generalized to other unknown specimens.

In the previous ICPR 2012 and ICIP 2013 contests, accuracy of maximum classification number was used as a performance metric. For specimen, in the ICPR 2012 data set, the testing number of images is 734, if the 500 images are classified correctly, then the accuracy is $\frac{500}{734}$. In this paper, we followed the metric of the previous ICPR 2012 and ICIP 2013 contest, and used the maximum classification number as the metric.

When comparing our method to the ICPR 2014 winner [16], we strictly followed the winner's protocol, and used the leave-one-specimen-out protocol. The averaged Mean Class Accuracy (MCA) was reported.

V. EXPERIMENTS

A. Implementation Details

PRICoLGBP. For multi-resolution PRICoLGBP feature, we use the original image and seven Gabor-filtered images under seven different scales $\{1, 2, 3, 4, 5, 6, 7\}$. For each filtered image, we can extract one PRICoLBP feature. In each PRICoLBP feature, we use ten templates. As we described before, the dimension of PRICoLBP using one template is 590. Thus, the final dimension for PRICoLGBP is $8 \times 10 \times 590 = 47,200$.

RootSIFT(IFV). We densely sample the RootSIFT feature at six scales with grid step 2. The sampled patch size is 41×41. If the image size (height or width) is less 64, we will resize it to the image with minimum size 64 and keep the height/width

ratio. Six scales are achieved by filtering the images with Gaussians with different scales of different standard deviates $\{1.5, 1.5^2, 1.5^3, 1.5^4, 1.5^5, 1.5^6\}$. For specimen, for an image with image size 70×70, we can sample 225 points for each scale. Thus, for six scales, we can get 1,350 sampled patches. For a larger image, such as 70×80, we will sample more points. In the IFV, we firstly sample 100,000 RootSIFT features from the training samples, then the 100,000 RootSIFT features are used to learn the PCA components, and 80 principal components are preserved as the basis for dimension reduction. As pointed out by [40], the PCA is a key step in the IFV framework. After the PCA step, we learn a GMM with 256 components. For the GMM, we use the Vlfeat to learn the parameters $\theta = \{\pi_k, \mu_k, \Sigma_k, k = 1, \cdots, K\}$. The final dimension using the IFV encoding is $2 \times 80 \times 256 = 41,960$.

Experimental Setups. The Vlfeat toolbox [42] is used for fast RootSIFT extraction and IFV encoding, and the Liblinear [43] is used for the linear SVM training and classification. For the parameter C, we cross-validated it in $\{0.001, 0.01, 0.1, 1, 100, 1,000\}$. As for the method of the ICIP winner (Shen et al., the ICIP 2013 winner), we use the same code that had been submitted into the ICIP 2013 contest and achieved the 1st place. All experimental comparisons are conducted in the same framework. Take the ICIP 2013 contest data set as an example, first, we create ten splits for ten repeated experiments. For each split, the whole ICIP contest 2013 data set are randomly divided into the training and testing sets. Meanwhile, to truly show the generalization performance of approaches, the images from the same cell are only divided into training or testing set. Thus, all comparisons are fair in this paper. We have provided the matlab code[3] to reproduce the experimental results.

TABLE III: Four experimental setups for the ICIP 2013 data set. For example, "1" means all images from one specimen are used for training, and the remaining specimens for testing.

	Ho	Sp	Nu	Ce	NM	Go	total
Setup A	1	1	1	1	1	1	6
Setup B	2	2	2	2	2	2	12
Setup C	4	4	4	4	4	2	22
Setup D	8	8	8	8	8	2	42

TABLE IV: Comparison between the single-resolution and the multi-resolution texture feature on the ICIP 2013 data set.

	Setup A	Setup B	Setup C	Setup D
PRICoLBP	43.1 ± 6.8	55.2 ± 5.4	67.6 ± 4.0	73.9 ± 2.9
PRICoLGBP	50.4 ± 6.2	61.4 ± 5.1	72.2 ± 3.8	78.4 ± 2.7

B. Evaluation of Features

In this subsection, we will mainly evaluate some aspects of the proposed texture and shape features. The ICPR 2012 data

---

[3]https://www.dropbox.com/s/eoifdhqjs1o7vky/HEp2Cell.zip?dl=0

set is too small to fully evaluate the properties of the proposed methods. Thus, we will use the ICIP 2013 data set in this subsection. To fully evaluate the properties, we use four sets of different experimental setups, as shown in Tab. III. Take the setup "D" as an example, in the experimental setup "D", 42 specimens (including eight specimens from "*Homogeneous*", eight specimens from "*Speckled*", eight specimens from "*Nucleolar*", eight specimens from "*Centromere*", eight specimens from "*NuMem*" and two specimens from "*Golgi*") are used for training, and the rest 41 specimens are used for testing. We pre-create ten training and testing splits randomly, and report the averaged results.

Evaluation of Multi-Resolution Texture Extraction Strategy. Here, we conduct experiments to compare the PRICoLBP and PRICoLGBP using the above-mentioned four experimental setups. The results are shown in Tab. IV.

We can observe that from Tab. IV, multi-resolution texture feature significantly improves the single-resolution texture feature. For specimen, the multi-resolution PRICoLGBP improves the PRICoLBP by 7.3% and 4.5% for the experimental setup "A" and "D". We believe the performance gain monotonically decreasing from 7.3% to 4.5% is reasonable. From the setup A to the setup D, more and more samples are used for the training, the performance for both the proposed and the compared approaches will increase. When the training samples are limited, the PRICoLGBP shows better discriminative power, the performance gain thus is high. With the increase of the training samples, the performance gain decreases, whereas the PRICoLGBP will consistently outperform the PRICoLBP.

Evaluation of Improved Fisher Vector Encoding. To evaluate the effectiveness of the Improved Fisher Vector, we compare it with the traditional Vector Quantization (VQ) and Vector of Locally Aggregated Descriptors (VLAD) encodings. The VQ statistics the occurrences of visual words, but the VLAD and the IFV calculate the differences between the pooled features and the words. For the VLAD and the IFV, the former encodes the first order moment of the descriptors assigned to a word, and the latter depicts both the first order and the second order moments. The K-means is usually used in the VLAD, but the GMM is typically used in the IFV. The dictionary size of the VQ used in this paper is 4,096. In the VLAD, we use the K-means to cluster 256 words. For the IFV, we used 256 GMMs. For all VQ, VLAD and IFV, the feature is normalized according to Eq. 8. A linear SVM is used for training and classification. The results are shown in Tab. V.

TABLE V: Comparison between the Vector Quantization and the Improved Fisher Vector on the ICIP 2013 data set.

	Setup A	Setup B	Setup C	Setup D
RootSIFT(VQ)	54.9 ± 6.8	62.2 ± 4.5	68.3 ± 2.4	71.2 ± 2.3
RootSIFT(VLAD)	57.9 ± 6.8	65.6 ± 4.6	73.5 ± 2.5	77.2 ± 2.5
RootSIFT(IFV)	58.7 ± 6.7	66.7 ± 5.1	74.7 ± 2.7	78.4 ± 2.7

From Tab. V, we can find that the IFV encoding method sharply improves the performance of the VQ encoding method,

and slightly outperforms the VLAD as shown in [42][4]. For instance, under the experimental configuration "D", the IFV improves the VQ from 71.2% to 78.4%, and improves the performance of the VLAD by 1.2%. This observation validates that, to encode the differences between the pooled features and the words is more effective than to simply encode the occurrences of visual words. Meanwhile, it also demonstrates that to depict both the first order and the second order moments at that same time performs better than to only capture the first order information although the VLAD encoding is faster than the IFV encoding. However, in the current system, the speed of the IFV is acceptable; it only takes around 0.1s to encode the features of an image.

Evaluation of Large-Patch Sampling Strategy. In order to evaluate the RootSIFT(IFV) under different patch sampling strategy, we conduct experiments under two configurations: the patch size 41×41 and the patch size 17×17. For fair comparison, we keep the numbers of samples points for both settings are approximately same. The results are shown in Tab. VI. According to the Tab. VI, the large patch sampling strategy

TABLE VI: Comparison between two different patch sampling strategy on the ICIP 2013 data set.

	Setup A	Setup B	Setup C	Setup D
RootSIFT(small)	56.9 ± 7.4	65.1 ± 5.6	73.2 ± 3.0	77.2 ± 2.8
RootSIFT(large)	58.7 ± 6.7	66.7 ± 5.1	74.7 ± 2.7	78.4 ± 2.7

is more effective than the small patch sampling strategy. For instance, the former outperforms the latter 1.2% under the experimental setup "D".

Evaluation of Normalization Method. Here, we evaluate the importance of the normalization method. For both PRICoLGBP feature and RootSIFT(IFV), we normalized the aggregations according to Eq. 8. We compare them with the direct L_2 normalized aggregations under the linear SVM framework. The results are shown in Tab. VII.

TABLE VII: Evaluation of the Normalization Methods on RootSIFT(IFV) and PRICoLGBP on the ICIP 2013 data set. In the table, "PRICoLGBP" and "RootSIFT(IFV)" denote the feature under the direct L_2 normalization, "PRICoLGBP*" and "RootSIFT(IFV*)" denote the features normalized by Eq. 8 and then by L_2 normalization.

	Setup A	Setup B	Setup C	Setup D
PRICoLGBP	48.4 ± 6.2	59.4 ± 4.6	70.3 ± 3.2	77.5 ± 3.2
PRICoLGBP*	50.4 ± 6.2	61.4 ± 5.1	72.2 ± 3.8	78.4 ± 2.7
RootSIFT(IFV)	57.8 ± 7.0	65.7 ± 5.4	73.6 ± 2.9	77.4 ± 2.7
RootSIFT(IFV*)	58.7 ± 6.7	66.7 ± 5.1	74.7 ± 2.7	78.4 ± 2.7

From Tab. VII, it is easy to find that the PRICoLGBP with normalization according to Eq. 8 consistently outperforms the

---

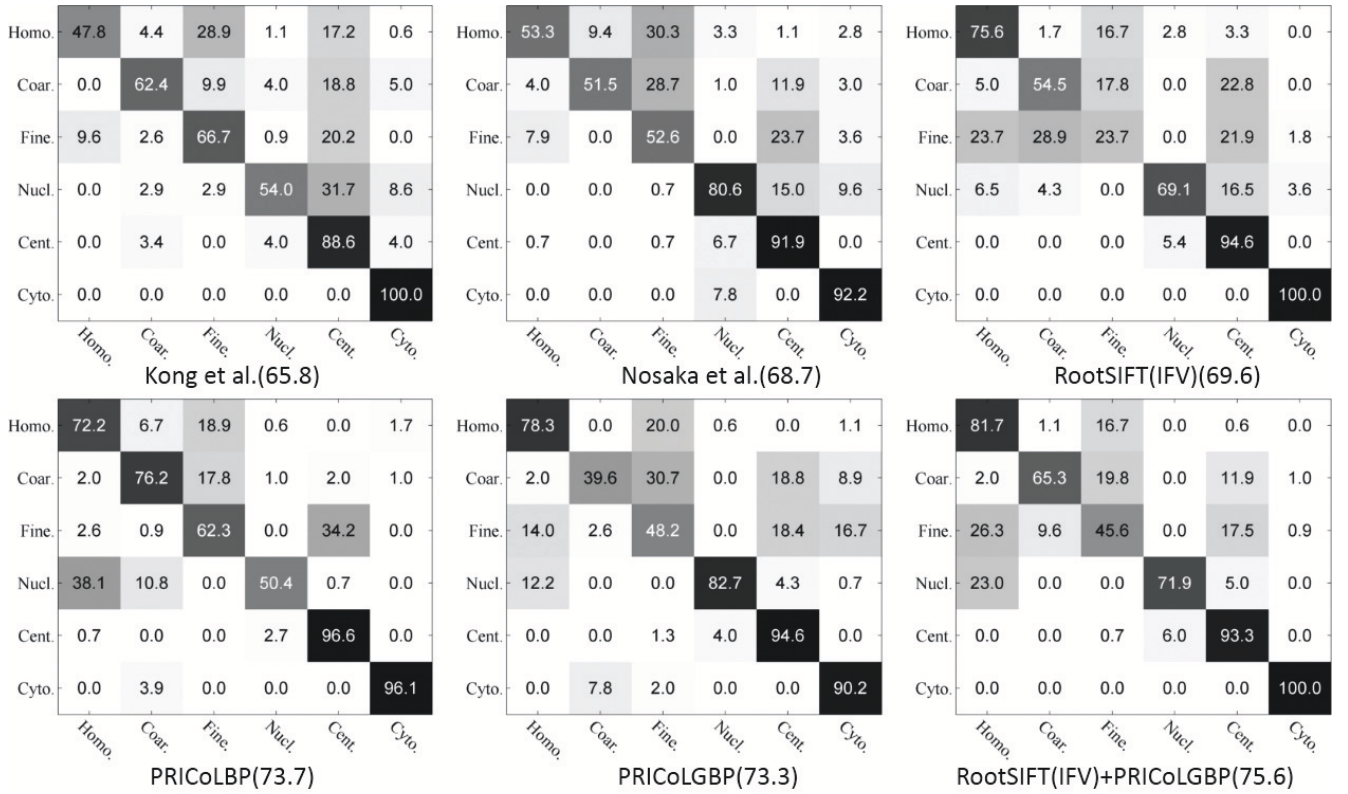[4]http://www.vlfeat.org/applications/apps.html

Fig. 7: Classification confusion matrices and averaged accuracies on the ICPR 2012 contest data set.

PRICoLGBP without normalization, and the RootSIFT(IFV) using normalization also consistently outperforms the non-normalized feature. All these two observations validate the effectiveness of the normalization.

TABLE VIII: Comparison with the state-of-the-art methods on the ICPR 2012 contest dataset.

Methods	Accuracy (%)
Li et al. [44]	64.2
Kong et al. [8]	65.8
Nokasa et al. [6]	68.7
GoC-LBPs [7]	69.2
Distribution of SIFT [7]	71.3
Combined Features (SRC classifier) [7]	75.1
SIFT (VLAD) [13]	70.57
Shape Index Histograms [15]	71.5
Human Expert	73.33
Gabor-PRICoLBP + RootSIFT(IFV)	**75.6**

C. Comparison with the State-of-the-art Methods

Experiments on the ICPR 2012 contest. To extract the PRICoLBP and the PRICoLGBP features, we only use the green channel. For the RootSIFT(IFV), we use the gray image. In the IFV, since the sampled patch is 41×41, when the minimal size of the image is less than 50, we will resize the image to the minimal size 64 while keeping the ratio between the height and the width. Since the dataset is very small, we directly use SVM with χ^2 kernel for the PRICoLBP and the PRICoLGBP. For the RootSIFT(IFV), and the combination of the PRICoLGBP and the RootSIFT(IFV), we used linear SVM. We used the official training/testing evaluation in the ICPR 2012 contest as shown in Table I. The classification confusion matrix and averaged accuracies using the provided experimental setup by the ICPR 2012 contest organizers are shown in Fig. 7. In Table VIII, we compare our method with some state-of-the-art approaches.

We observe that from Fig. 7 and Table VIII:
- The PRICoLGBP and the RootSIFT(IFV) work well on this dataset, and both outperform the winner of the ICPR 2012 contest. The PRICoLGBP also outperforms the RootSIFT(IFV) significantly. The combination of the PRICoLGBP and the RootSIFT(IFV) further improves each of them. Specifically, we want to emphasize that the combination improves the PRICoLGBP by 25.7% on the category "*Coarse Speckled*", and improves the Root-SIFT(IFV) by 21.9% on the category "*Fine Speckled*";
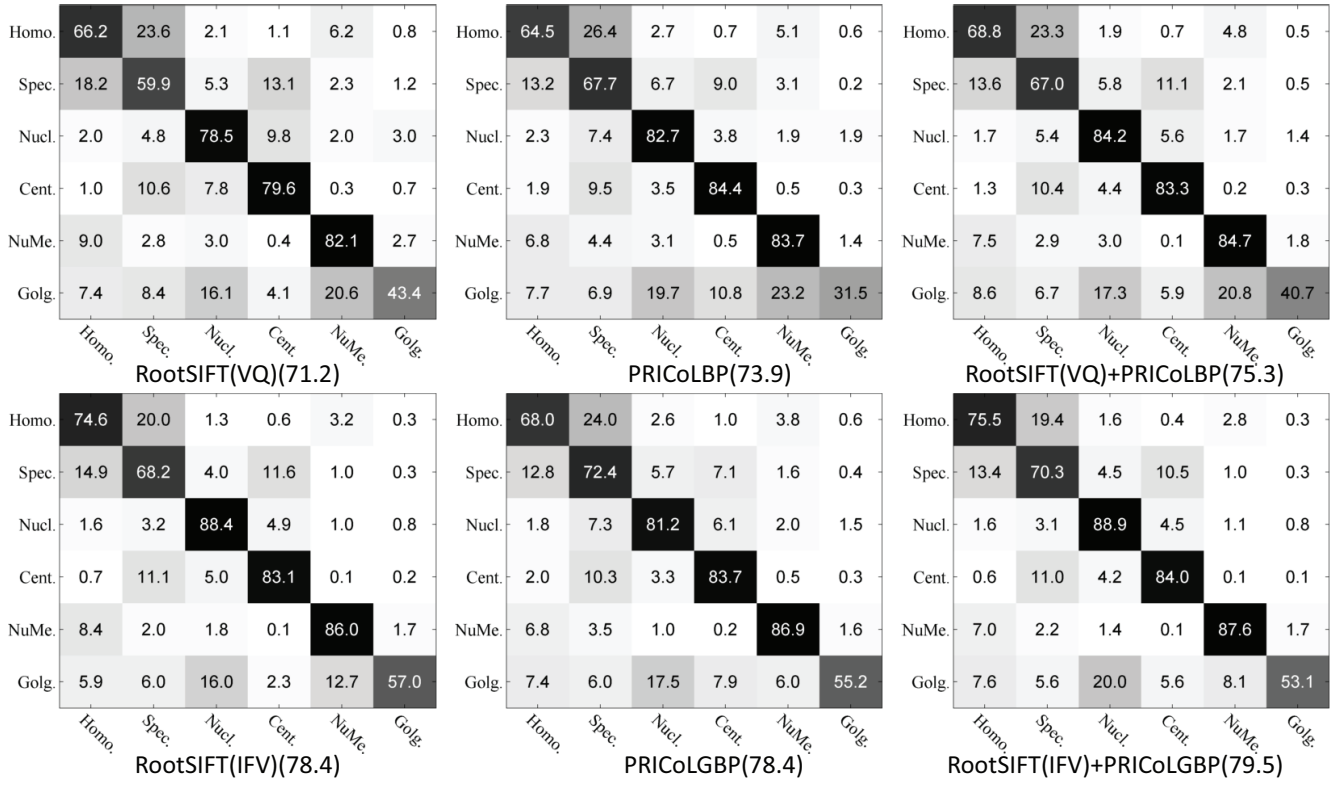- Our approach significantly improves the top three methods in the ICPR 2012 contest, and also outperforms two

Fig. 8: Classification confusion matrices and averaged accuracies on the ICIP 2013 contest data set.

other well-performing works [5], [13]. It should be noted that the task is even challenging for the human expert. As you can see from Table VIII, human expert only achieves 73.33% on this task. From this perspective, an accurate HEp-2 cell classification system is very valuable;

- The most confusing pairs are "*Fine Speckled*" and "*Centromere*", "*Nucleolar*" and "*Centromere*", and "*Homogeneous*" and "*Fine Speckled*". For instance, in all six algorithms, the category "*Homogeneous*" is easily misclassified into the category "*Fine Speckled*".

In the ICPR 2012 contest data set, the texture based methods work better than the shape based methods. This observation is consistent with the analysis in [2]. This may account for the small scale of this data set. Texture property is more stable than the shape property when the number of the data set is small. Meanwhile, along with increase of the data set, the statistics of shape property will become more stable. This argument is partly validated by the following experiment on the ICIP 2013 contest data set. We also want to point out another issue that when the scale of the data set is small, the classifier will be more sensitive to the parameter C. We used the training set to conduct cross-validation to determine the parameter C.

Experiments on the ICIP 2013 contest. We evaluate and compare six methods including the RootSIFT(VQ), the PRICoLBP, a combination of RootSIFT with VQ and the PRICoLBP, the PRICoLGBP, the RootSIFT(IFV), and a com-

bination of the PRICoLGBP and the RootSIFT(IFV). Here, we use the experimental setup "D". The classification confusion matrix and averaged accuracies over ten trials are shown in Fig. 8. We observe that from Fig. 8:

1). The PRICoLGBP greatly outperforms the PRICoLBP, and the RootSIFT(IFV) significantly outperforms the RootSIFT(VQ). This fully demonstrates the effectiveness of the proposed approaches (PRICoLGBP and RootSIFT(IFV)). The effectiveness of the PRICoLGBP verifies that the multi-resolution information is important in the HEp-2 cell classification. The IFV has proved to be more effective than the VQ on many applications, including object recognition, texture classification, and face verification. This point also explains why the RootSIFT(IFV) significantly outperforms the RootSIFT(VQ) on this task.

2). Compared to the RootSIFT(VQ), the RootSIFT(IFV) shows obvious improvement on all categories. This fully illustrates the high effectiveness of the IFV encoding. Compared to the PRICoLBP, the PRICoLGBP achieves improvement on four categories ("*Homogeneous*", "*Speckled*", "*NuMem*", and "*Golgi*"). This may reflect that the multi-resolution texture information on these four categories is discriminative, especially on the "*Golgi*". On the other two categories, the performances of the PRICoLBP and PRICoLGBP are comparable.

3). The combination between the texture and the shape features outperforms each of them, and improves the performance

78.4% for the single feature to 79.5%. We repeat experiments ten trials and perform a one-tailed statistical significance test. The statistical significance test shows that the feature combination is significantly better than the single feature. Meanwhile, we would like to point out that, in [16], the combination of four features with performance 80.25% only improves the best descriptor (the performances of four features are 78.00%, 78.63%, 61.26% and 79.60% individually) for 0.65%. In this perspective, the 1.1% improvement is significant.

4). The category "*Golgi*" obtains the lowest performance among all six categories, this is due to the less training samples in this category. The most confusing pairs are "*Golgi*" and "*Nucleolar*", and "*Speckled*" and "*Homogeneous*". This is mainly due to that the shape and the texture structures between the pairs are similar.

Comparison with the winner of the ICPR 2014 contest. In this part, we compare our method with the rank first approach [16] of the ICPR 2014 contest. We strictly follow their experimental protocol, and use the leave-one-specimen-out strategy. According to the specimen IDs, we can split the data into training and validation sets. Since we have 83 different specimens, in each test, we use 82 specimens for training and the rest one for testing. The result of [16] is shown in Table IX(a), and our result based on the combination of the PRIGCoLBP and the RootSIFT(IFV) is shown in Table IX(b).

According to Table IX, the Mean Class Accuracy (MCA) for [16] is 80.25%, and our method achieves a comparable performance 80.04%. Compared to their method, our method performs better on "*Nucleolar*" and "*Golgi*", and performs worse on "*Homogeneous*", "*Speckled*" and "*Centromere*".

There are two key differences between our method and [16]. 1). Our method is built on the proposed two features (PRICoLGBP and RootSIFT (IFV)), but their method [16] is based on the combination of four features. 2). Our method only uses a linear SVM classifier, but their method combines sixteen classifiers. In [16], Manivannan et al. combined four types of features. They rotated the images to 0°, 90°, 180°and 270°, and for each direction, they trained one model. Thus, for each feature, they trained four models. The final classification result is based on the summation of the probability of sixteen classifiers. However, we only trained one model using the combination of the RootSIFT (IFV) and the PRICoLGBP features. Usually, a model combination will work better than a single model. In this manner, the proposed approaches can be further improved if we use multiple models for the proposed approaches. Meanwhile, we also expect that the combination of the proposed two features and [16] will lead to an even better performance.

VI. CONCLUSION

This paper proposed an effective method for automatic classification of the HEp-2 cell via using multi-resolution co-occurrence texture and richer shape information. Specifically, we proposed to capture the multi-resolution co-occurrence texture information by a novel Pairwise Rotation Invariant Co-occurrence of Local Gabor Binary Pattern (PRICoLGBP) descriptor, and depict the richer shape information by using an

TABLE IX: Classification confusion matrices of (a). the ICPR 2014 rank first approach, and (b). Our method under the leave-one-specimen-out experiments on the I3A Task-1 dataset.

(a). The ICPR 2014 rank first approach [16].						
	Homo.	Spec.	Nucl.	Cent.	NuMe.	Golgi
Homo.	**81.8**	15.00	0.76	0.20	2.04	0.20
Spec.	8.87	**77.36**	3.67	9.18	0.74	0.18
Nucl.	1.12	3.89	**90.65**	2.08	1.27	1.00
Cent.	0.47	10.87	2.85	**85.66**	0.04	0.11
NuMe.	6.30	2.04	1.40	0.27	**88.04**	1.95
Colgi.	5.66	3.73	20.72	2.35	9.53	**58.01**
(b). Our method.						
	Homo.	Spec.	Nucl.	Cent.	NuMe.	Golgi
Homo.	**81.32**	15.56	0.80	0.12	1.96	0.24
Spec.	12.19	**73.86**	3.74	9.29	0.78	0.14
Nucl.	1.58	2.23	**92.49**	2.19	0.92	0.58
Cent.	0.15	10.98	3.43	**85.41**	0	0.04
NuMe.	6.52	1.54	1.27	0.05	**88.04**	2.58
Colgi.	8.56	4.70	19.34	3.31	4.97	**59.12**

Improved Fisher Vector (IFV) model with RootSIFT features which are sampled from large image patches in multiple scales. We systematically evaluated the proposed approach on the ICPR 2012, the ICIP 2013, and the ICPR 2014 HEp-2 cell classification contest data sets. The proposed approach demonstrated superior performance on all three data sets.

VII. ACKNOWLEDGMENT

REFERENCES

[1] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking hep-2 cells classification methods." *IEEE Transactions on Medical Imaging*, vol. 32, no. 10, p. 1878, 2013.

[2] P. Foggia, G. Percannella, A. Saggese, and M. Vento, "Pattern recognition in stained hep-2 cells: Where are we now?" *Pattern Recognition*, vol. 47, no. 7, pp. 2305–2314, 2014.

[3] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Early experiences in mitotic cells recognition on hep-2 slides," in *IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS)*, 2010, pp. 38–43.

[4] G. Percannella, P. Soda, and M. Vento, "A classification-based approach to segment hep-2 cells," in *IEEE 25th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2012, pp. 1–5.

[5] I. Theodorakopoulos, D. Kastaniotis, G. Economou, and S. Fotopoulos, "Hep-2 cells classification via fusion of morphological and textural features," in *IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE)*. IEEE, 2012, pp. 689–694.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] X.-H. Han, J. Wang, G. Xu, and Y.-W. Chen, "High-order statistics of micro-texton for hep-2 staining pattern classification," *IEEE Transactions on Biomedical Engineering*, 2014.

[12] M. Faraki, M. T. Harandi, A. Wiliem, and B. C. Lovell, "Fisher tensors for classifying human epithelial cells," *Pattern Recognition*, vol. 47, no. 7, pp. 2348–2359, 2014.

[13] D. Kastaniotis, I. Theodorakopoulos, G. Economou, and S. Fotopoulos, "Hep-2 cells classification using locally aggregated features mapped in the dissimilarity space," in *IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 2013, pp. 1–4.

[14] I. Theodorakopoulos, D. Kastaniotis, G. Economou, and S. Fotopoulos, "Hep-2 cells classification using morphological features and a bundle of local gradient descriptors," in *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*. IEEE, 2014, pp. 33–36.

[15] B. L. L. Anders, S. V. Jacob, and L. Rasmus, "Hep-2 cell classification using shape index histograms with donut-shaped spatial pooling," *IEEE Transactions on Medical Imaging*, 2014.

[16] S. Manivannan, W. Li, S. Akbar, R. Wang, J. Zhang, and S. J. McKenna, "Hep-2 cell classification using multi-resolution local patterns and ensemble svms," in *Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A), 2014 1st Workshop on*. IEEE, 2014, pp. 37–40.

[17] 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.

[18] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[19] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *In Workshop on statistical learning in computer vision, ECCV*, 2004.

[20] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," *Image and vision computing*, vol. 10, no. 8, pp. 557–564, 1992.

[21] 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.

[22] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[23] X. Qi, R. Xiao, C.-G. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise rotation invariant co-occurrence local binary pattern," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 11, pp. 2199–2213, 2014.

[24] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer vision using local binary patterns*. Springer, 2011, vol. 40.

[25] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.

[26] 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.

[27] 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.

[28] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, no. 6, pp. 610–621, 1973.

[29] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, 1946.

[30] G. V. Ponomarev, V. L. Arlazarov, M. S. Gelfand, and M. D. Kazanov, "Ana hep-2 cells image classification using number, size, shape and localization of targeted cell regions," *Pattern Recognition*, vol. 47, no. 7, pp. 2360–2366, 2014.

[31] P. Hobson, G. Percannella, M. Vento, and A. Wiliem, "http://nerone.diiie.unisa.it/contest-icip-2013-test/index.shtml."

[32] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *ACM International Conference on Image and Video Retrieval*. ACM, 2007, pp. 401–408.

[33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[34] A. Vedaldi and A. Zisserman, "Efficient additive kernels via explicit feature maps," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 3, pp. 480–492, 2012.

[35] B. Lovell, G. Percannella, M. Vento, and A. Wiliem, "Icpr 2014 hep-2 cells classification workshop and contest," in *Available: http://i3a2014.unisa.it/*.

[36] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," in *European Conference on Computer Vision 2010*. Springer, 2010, pp. 143–156.

[37] J. R. Movellan, "Tutorial on gabor filters," *Technical Report, MPLab Tutorials*, 2005.

[38] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," in *IEEE International Conference on Computer Vision (ICCV)*, vol. 1. IEEE, 2005, pp. 786–791.

[39] R. Arandjelovic and A. Zisserman, "Three things everyone should know to improve object retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2911–2918.

[40] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *International Journal of Computer Vision*, pp. 1–24, 2013.

[41] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Hep-2 cells classification contest website [online]," in *Available: http://mivia.unisa.it/hep2contest/*.

[42] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *ACM Multimedia*, 2010, pp. 1469–1472.

[43] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

[44] 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.