HEp-2 Specimen Classification using Multi-resolution Local Patterns and SVM

Siyamalan Manivannan, Wenqi Li, Shazia Akbar, Ruixuan Wang, Jianguo Zhang and Stephen J. McKenna
CVIP, School of Computing, University of Dundee, UK

***Abstract*—A pattern recognition system was developed to classify immunofluorescence images of HEp-2 specimens into seven classes: homogeneous, speckled, nucleolar, centromere, golgi, nuclear membrane, and mitotic spindle. Root-SIFT features together with multi-resolution local patterns were used to capture local shape and texture information. Sparse coding with max-pooling was applied to get an image representation from these local features. Specimens were classified using a linear support vector machine. Leave-one-specimen-out experiments on the I3A Contest *Task 2* data set predicted a mean class accuracy of 89.9%.**

I. Introduction

A recommended method for anti-nuclear antibody testing involves *Indirect Immunofluorescence (IIF)* imaging of Human Epithelial (HEp-2) cells. There is a lot of interest in recent years in reliably automating recognition of cell patterns in these images. In a companion paper we describe a pattern recognition system for classifying the individual cells [1]. This paper focuses on the related task of classifying specimens containing multiple cells.

Various morphological features and texture features such as Local Binary Patterns (LBP) have been widely applied for classification of individual HEp-2 cell images [2], [3], [4], [5], [6], [7]. Less attention has been paid to the classification of whole specimen images. Recently, Wiliem *et. al.* [8] proposed a specimen-level image classification system in which a specimen-level image descriptor is learned based on cell-level attributes. In another system, proposed by Soda *et. al.* [9], each specimen image is classified based on the strength of the recognition of the cells in that specimen image.

Here we focus on building a classification system to classify IIF HEp-2 specimen images into seven predefined classes based on the I3A-2014 Task-2 dataset. Some images from this dataset are shown in Fig. 1. Please refer to Section III for more details about the dataset. Our system for specimen image classification is based on multiple types of local feature and a linear SVM. Our proposed approach is explained in detail in Section II. Experimental results are reported in Section III and conclusions are made in Section IV.

II. Methodology

We describe a system to classify pre-segmented immunofluorescence images of HEp-2 specimens into seven classes: homogeneous, speckled, nucleolar, centromere, golgi, nuclear membrane and mitotic spindle. Prior to feature extraction, each image was intensity-normalised. Then a set of features was extracted from these images. A feature encoding method based on sparse coding was employed to aggregate the local features into an image representation. Finally a linear support vector machine (SVM) classifier was employed to classify each test image into one of the seven classes. The following sections describe the proposed method in some more detail.

A. Preprocessing and feature extraction

Prior to feature extraction, each image was intensity-normalised. Specifically, a specimen's segmentation mask (consisting of many disjoint masks of individual cells or cell clusters) was dilated using a 5×5 structuring element, and image intensity values within the dilated mask regions were then linearly rescaled so that 2% of pixels became saturated at low and high intensities respectively.

Two types of *local* features were then extracted:

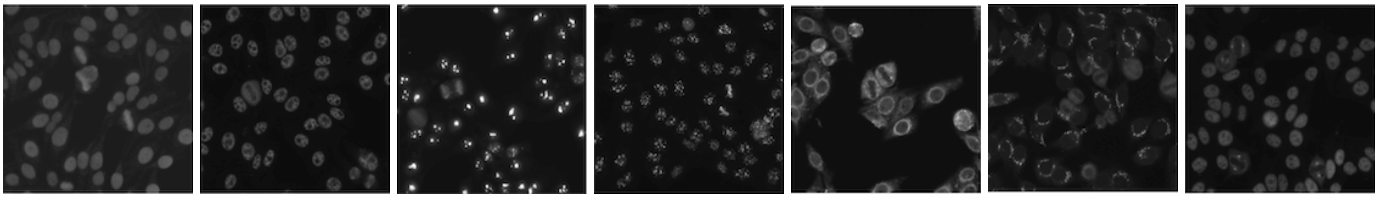
1) Multi-resolution local patterns (mLP): These are a multi-resolution version of local higher-order statistical (LHS) patterns. LHS is a non-binarized version of local binary patterns proposed by Sharma et al. for texture classification [10]. It operates on a small image neighbourhood of size 3×3 pixels. To capture information from a larger neighbourhood and reduce noise effects, we used the sampling patterns described by Maenpaa [11]. See [1] for further details.

2) Root-SIFT (rSIFT): Root-SIFT is a variant of the widely used SIFT descriptor that produces better performance than SIFT on some image matching and retrieval tasks [12]. The standard SIFT descriptor is a histogram representation of local image derivatives. Using Euclidean distance between histograms often yields inferior performance compared to other measures such as χ^2 or Hellinger for texture classification and image categorization [12]. Therefore in [12] the standard SIFT descriptor was been modified to create Root-SIFT so that comparing RootSIFT descriptors using Euclidean distance is equivalent to using the Hellinger kernel to compare the original SIFT vectors.

Multi-scale feature extraction: From each image, all the features were densely extracted from image patches using a patch step-size of 4 pixels. Both small patches (12×12 pixels and 16×16 pixels) and large patches (48×48 pixels and 64×64 pixels) were used. Intuitively, small patches can capture local properties at cellular level while large patches can capture information about groups of cells. Features from outside the dilated cell masks were discarded.

B. Feature encoding and image representation

Bag-of-words based feature encoding methods are widely applied for image classification [13], [14]. However, sparse



(a) Homogeneous (b) Speckled (c) Nucleolar (d) Centromere (e) Nuclear Memb. (f) Golgi (g) Mitotic Spindle

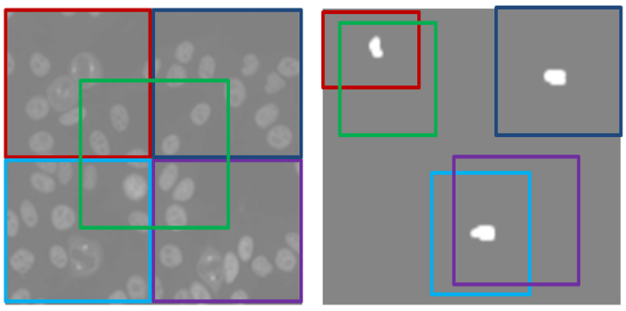
Fig. 1: Sample images from I3A-2014 Task-2 dataset.

coding often shows improved performance [15], [16]. We use locality-constrained linear coding [16], an efficient variant of sparse coding, that utilizes the local linear property of manifolds to project each descriptor into its local-coordinate system.

In our classification system, a separate dictionary of size M was learned for each feature type with each group of patches (i.e., one for *small* and one for *large* patches). Max-pooling was used to aggregate the locality-constrained linear codes. The image representations using the small-patch dictionary and the large-patch dictionary for each feature type were concatenated to give a $4M$-vector on which classification was based.

C. Classifier training

The I3A Contest *Task 2* provides a training set of 1008 images from 252 specimens. This dataset was augmented by including a 90°-rotated version of each original image, resulting in a set of 2016 images. Five sub-images were extracted from each image based on the layout shown in Fig 2(a) with the exception of images in the mitotic spindle class. In mitotic spindle images, metaphase cells in which stained mitotic spindle was apparent were manually identified. Five sub-images were then extracted around those cells, with some random variation, as shown in Fig 2(b). Finally, the $10,080$ extracted sub-images were added to the $2,016$ images, resulting in an augmented dataset of $12,096$ images.



(a) (b)

Fig. 2: Sub-images extracted from specimen images for (a) the homogeneous, speckled, nucleolar, centromere, golgi, and nuclear membrane classes, and (b) the mitotic spindle class. White blobs in each image indicate the segmented regions of individual cells.

A one-vs-rest, multi-class SVM classifier was then trained on the augmented dataset. Since each specimen was imaged at four different locations in the test phase, generating four images for each specimen, we ran the trained SVM on each of these four images, resulting in four sets of classification scores per specimen. Scores were treated as probabilities using Platt rescaling [17]. The final classification decision was made by averaging these probabilistic scores and selecting the class with the highest average score.

III. Experiments

A. Dataset

The I3A-2014 Task-2 dataset was used to evaluate our classification system. This dataset was acquired in 2013 at the Sullivan Nicolaides Pathology Laboratory, Australia. It was collected from 1001 patients' sera with positive ANA test results. Each patient sera was diluted to 1:80 and the specimen was photographed using a monochrome camera fitted on a microscope. Each specimen was photographed at four different locations, resulting in four images for each specimen. The dataset includes seven classes of cell patterns: homogeneous, speckled, nucleolar, centromere, nuclear membrane, golgi, and mitotic spindle. The first four classes represent common ANA patterns whilst the other three classes are less common. The dataset is divided into training and testing set, with approximately 25% for training and 75% for testing. All images are in monochromatic uncompressed format with resolution 1388×1040 pixels and cell masks obtained based on automatic segmentation for each image.

B. Experimental setting

A 3-resolution mLP was used, with 8 sampling points at each resolution. The parameters of the Gaussian filters at each sampling point were selected as in [11]. K-means was used for dictionary learning with $300,000$ randomly selected instances of each type of local feature. The size of the dictionary, M, was empirically set to 5000. We used the implementation of LLC from [16] with 10 nearest neighbours ($K = 10$). The Liblinear package [18] was used to build the classifier. The Mean Class Accuracy (MCA) was used as the evaluation metric, defined as

$$\text{MCA} = \frac{1}{K}\sum_{k=1}^{K}\text{CCR}_k \tag{1}$$

where CCR_k is the correct classification rate for class k and K is the number of classes. In testing, the rotated images (Section II-C) were not used.

C. Experiment 1: Comparison of different features

We compared the performance obtained when using different features. *5 fold-cross-validation* experiments were carried out. Table I reports the MCA for each feature type as well as their combination. mLP with larger patch sizes outperformed other features. rSIFT with larger patch sizes gave the worst result. Combining all the features together resulted in a performance improvement with an MCA of 89.93% obtained; Table II reports the confusion matrix for this case.

Feature type.	MCA(%)
mLP $(12 \times 12, 16 \times 16)$	85.63
rSIFT $(12 \times 12, 16 \times 16)$	85.62
mLP $(48 \times 48, 64 \times 64)$	87.34
rSIFT $(48 \times 48, 64 \times 64)$	82.20
All	89.93

TABLE I: The performance of different features based on 5 fold-cross-validation. (All = histogram of rSIFT and histogram of mLP, with *smaller* and *larger* patch sizes).

D. Experiment 2: Leave-one-specimen-out experiments

Leave-one-specimen-out experiments were carried out using the specimen IDs provided to split the data into training and validation sets. Since 252 different specimens were available, we used images and their rotated versions from 251 specimens for training in each fold. Table III reports the confusion matrix of different classes. The trained classifier obtained 100% accuracy for the centromere and golgi classes, but performed worst for the mitotic spindle class by confusing this class mostly with the homogeneous class. An MCA of 89.93% was obtained. (Although the MCA is the same as that of Experiment 1, the confusion matrix is not exactly the same.)

E. Experiment 3: Classification of individual images taken at different locations

In the above two experiments (Exp. III-C and Exp. III-D) the final decision of each specimen is made based on averaging the classification scores of its four images taken from different locations as explained in Section II-C. In this experiment we consider each of the four images separately and classify them individually. A leave-one-specimen-out experiment was performed, where at each iteration a classifier is trained on the images (and their rotated versions) obtained from 251 specimens, and tested on each of the four images of the test specimen. An MCA of 87.93% was obtained which is 2% less then the best accuracy (89.93%) obtained in Exp. III-D. Figs. 3 and 4 show some examples of images that were erroneously classified.

IV. Conclusions

We developed a pattern recognition system to classify immunofluorescence images of HEp-2 specimen into seven classes. Leave-one-specimen out experiments on the I3A-2014 Task-2 training images predicted a mean class accuracy of 89.93%.

References

[1] 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 *1st Workshop on Pattern Recognition Techniques for Indirect Immunofluorescence Images (I3A)*, 2014.

[2] A. B. L. Larsen, J. S. Vestergaard, and R. Larsen, "HEp-2 cell classification using shape index histograms with donut-shaped spatial pooling," *IEEE Transactions on Medical Imaging*, vol. To appear, 2014.

[3] 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*, 2012.

[4] J. L. Linlin Shen and S. Yu, "Hep-2 cell classification using rotationally invariant features," School of Computer Science and Software Engineering, Shenzhen University, China, Tech. Rep., 2013.

[5] P. Foggia, G. Percannella, P. Soda, and M. Vento, "Benchmarking hep-2 cells classification methods." *IEEE Transactions on Medical Imaging*, 2013.

[6] I. Ersoy, F. Bunyak, J. Peng, and K. Palaniappan, "Hep-2 cell classification in iif images using shareboost," in *21st International Conference on Pattern Recognition*, 2012.

[7] 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 *21st International Conference on Pattern Recognition*, Nov 2012.

[8] B. C. L. A. Wiliem, P. Hobson, "Discovering discriminative cell attributes for hep-2 specimen image classification." in *IEEE Winter Conference on Applications of Computer Vision*, 2014.

[9] P. Soda and G. Iannello, "Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis." *IEEE Transactions on Information Technology in Biomedicine*, 2009.

[10] G. Sharma, S. Hussain, and F. Jurie, "Local higher-order statistics for texture categorization and facial analysis," in *IEEE European Conference on Computer Vision*, 2012.

[11] T. Maenpaa, "The local binary pattern approach to texture analysis - extensions and applications," Ph.D. dissertation, University of Oulu, 2003.

[12] R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval," in *IEEE Computer Vision and Pattern Recognition*, 2012.

[13] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *IEEE International Conference on Computer Vision*, 2003.

[14] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *IEEE Computer Vision and Pattern Recognition*, 2006.

[15] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *IEEE Computer Vision and Pattern Recognition*, June 2009.

[16] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *IEEE Computer Vision and Pattern Recognition*, 2010.

[17] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in large margin classifiers*, 1999.

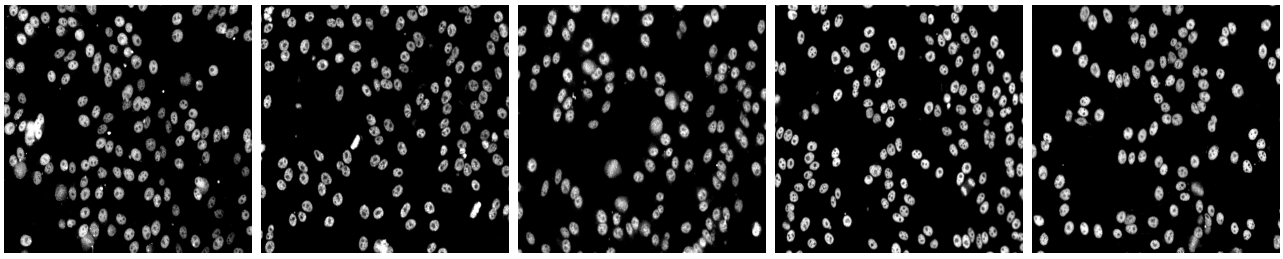
[18] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *Journal of Machine Learning Research*, 2008.

	Homog.	Speck.	Nucle.	Centr.	golgi.	numem.	mitsp.
Homog.	86.79	9.43	1.89	0.00	0.00	1.89	0.00
Speck.	1.92	96.15	0.00	0.00	0.00	1.92	0.00
Nucle.	0.00	0.00	98.00	2.00	0.00	0.00	0.00
Centr.	0.00	0.00	0.00	100.00	0.00	0.00	0.00
golgi.	0.00	0.00	0.00	0.00	100.00	0.00	0.00
numem.	0.00	0.00	0.00	0.00	0.00	95.24	4.76
mitsp.	26.67	6.67	0.00	0.00	0.00	13.33	53.33

TABLE II: Confusion matrix based on 5 fold-cross-validation.

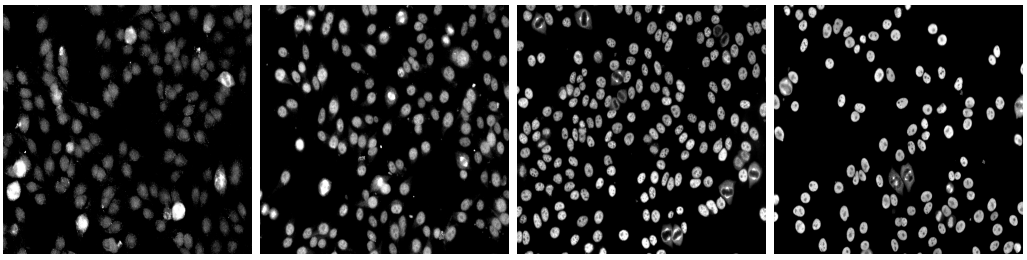
	Homog.	Speck.	Nucle.	Centr.	golgi.	numem.	mitsp.
Homog.	88.68	9.43	0.00	0.00	0.00	1.89	0.00
Speck.	3.85	94.23	0.00	0.00	0.00	1.92	0.00
Nucle.	0.00	0.00	98.00	2.00	0.00	0.00	0.00
Centr.	0.00	0.00	0.00	100.00	0.00	0.00	0.00
golgi.	0.00	0.00	0.00	0.00	100.00	0.00	0.00
numem.	0.00	0.00	0.00	0.00	0.00	95.24	4.76
mitsp.	26.67	6.67	0.00	0.00	0.00	13.33	53.33

TABLE III: Confusion matrix for leave-one-specimen-out experiment.



(a) (b) (c) (d) (e)

Fig. 3: The images from the Homogeneous class (specimen numbers $[14, 22, 26, 37, 47]$) that were wrongly classified as Speckled.



(a) (b) (c) (d)

Fig. 4: Some images from the Mitotic Spindle class (specimen numbers $[240, 241, 248, 249]$) that were wrongly classified as Homogeneous.