

# HEp-2 Cell Classification using Multi-resolution Local Patterns and Ensemble SVMs

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

**Abstract**—We describe a pattern recognition system for classifying immunofluorescence images of HEp-2 cells into six classes: homogeneous, speckled, nucleolar, centromere, golgi, and nuclear membrane. We use locality-constrained linear coding to encode multiple local features and two-level cell pyramids to capture spatial structure of cells. An ensemble of linear support vector machines is used to classify each cell image. Leave-one-specimen-out experiments on the I3A Contest *Task 1* training data set predicted a mean class accuracy of 80.25%.

## I. INTRODUCTION AND RELATED WORK

The Anti-Nuclear Antibody (ANA) clinical pathology test is commonly used to identify various autoimmune diseases. A common method for identifying the presence of ANAs is *Indirect Immunofluorescence (IIF)* on Human Epithelial (HEp-2) cells due to its high sensitivity and the large range of antigens that can be detected [1]. The IIF HEp-2 cell patterns which are imaged need to be recognized for a more detailed diagnosis. This is usually done by human observers and is therefore subjective, relying heavily on the experience of the observer. Partly due to large inter- and intra-observer variability there is a strong demand for developing automated procedures to identify IIF patterns that would improve test repeatability and reliability.

In this work we focus on building an automated pattern recognition system to classify IIF HEp-2 cell images into predefined classes. Recent research has focussed on extracting suitable image features from IIF images for classification. Various morphological features and texture features such as Local Binary Patterns (LBP) have been widely applied for classification of segmented images of HEp-2 cells [2], [3], [4], [5], [6]. Due to poor quality images, utilising low-level features directly may be inefficient. Ersoy et al. [7] and Li et al. [8] handle noisy features during classification via boosting. In addition cell patterns such as Golgi (Fig. 1(f)) have some spatial information which is not captured well by many traditional feature representations; to partly encode this spatial information Wiliem et al. [1] used features computed from inner and outer cell regions.

In this paper we develop a system for cell image classification which is based on multiple types of local feature and which uses a two-level pyramid to retain some spatial information.

## II. METHOD

We describe a system to classify pre-segmented immunofluorescence images of HEp-2 cells into six classes: homoge-

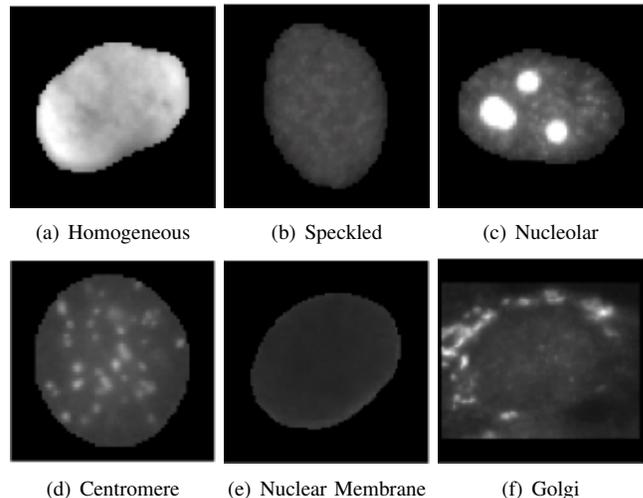


Fig. 1: Sample images from I3A 2014 dataset

neous, speckled, nucleolar, centromere, golgi, and nuclear membrane. Fig. 1 shows an example cell from each class. Firstly, cell images were intensity-normalised. A set of local features was then extracted and a feature encoding method based on sparse coding was employed to aggregate the local features into a cell image representation. A two-level cell pyramid was used to capture spatial structure. Finally an ensemble of one-vs-rest linear support vector machines (SVMs) was trained to classify cell images into one of the six classes. The following sections describe the proposed method in detail.

### A. Preprocessing and Feature Extraction

Prior to feature extraction, each cell's image was intensity-normalised; specifically, a cell's segmentation mask was dilated (using a  $5 \times 5$  structuring element) and image intensity values within the dilated mask region were then linearly rescaled so that 2% of pixels became saturated at low and high intensities.

Four types of *local* feature were then extracted:

1) *Multi-resolution local patterns (mLP)*: These are a multi-resolution version of the local higher-order statistical (LHS) patterns proposed by Sharma et al. [9] for texture classification. LHS is a non-binarized version of the well-known Local Binary Patterns method. It operates on a small image neighbourhood of size  $3 \times 3$ . To capture information from a larger neighbourhood and reduce noise effects, we used the sampling patterns described by Maenpaa [10]. This

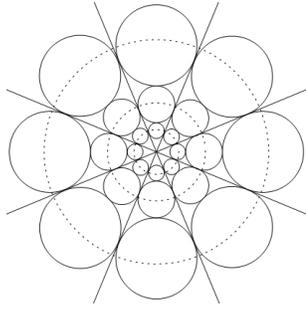


Fig. 2: The Gaussian filtered sampling points

is inspired by the spatial structure of receptive fields in the human retina and has been widely adopted in recently developed visual features in computer vision such as FREAK [11], BRISK [12], and DAISY [13]. Fig. 2 shows an example sampling pattern where the local neighborhood is quantized radially into three resolutions (radii), and at each resolution a set of ( $N = 8$ ) sampling regions (indicated as circles) are considered. At each sampling point a Gaussian filter was applied, integrating information from the filter’s region of support. We call the combination of LHS and these sampling patterns *multi-resolution local patterns*.

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 [14]. The standard SIFT descriptor is a histogram representation of local image derivatives and was originally designed to be used with Euclidean distance. Using Euclidean distance to compare histograms often yields inferior performance compared to other measures such as  $\chi^2$  or Hellinger for texture classification and image categorization [14]. Therefore, standard SIFT was modified in [14] to create Root-SIFT such that comparing RootSIFT descriptors using Euclidean distance is equivalent to using the Hellinger kernel to compare SIFT vectors.

3) *Random projections (RP)*: Random projection, a simple yet powerful method for dimensionality reduction [15], projects patch intensity vectors from the original patch-vector space  $\mathbb{R}^{D'}$  to a compressed space  $\mathbb{R}^D$  using randomly chosen projection vectors. Such a scheme has been successfully applied to texture image classification [16]. Let  $\mathbf{x}$  be a  $D'$ -dimensional patch vector and  $\hat{\mathbf{x}}$  be its  $D$ -dimensional representation in the compressed space. The RP method simply maps these vectors using a  $D \times D'$  random projection matrix  $R$ , such that:

$$\hat{\mathbf{x}}_{D \times 1} = R_{D \times D'} \mathbf{x}_{D' \times 1} \quad (1)$$

Each element in matrix  $R$  is sampled from a Gaussian distribution with zero mean and unit variance. The key point of RP is that when projecting the patch-vectors from the original space to the compressed space their relative distances are approximately preserved.

4) *Intensity histograms (IH)*: We compute intensity histograms from small image patches to capture the local intensity information.

## B. Feature encoding and image representation

Bag-of-words based feature encoding methods are widely applied for image classification [17], [18]. However, sparse coding which assigns each local feature to a set of bases (dictionary elements) often shows improved performance [19], [20]. Locality-constrained linear coding (LLC), an efficient variant of sparse coding, utilizes the local linear property of manifolds to project each descriptor into its local-coordinate system [20].

Let  $X$  be a set of  $D$ -dimensional local descriptors extracted from an image, *i.e.*  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$ . Given a codebook (dictionary) with  $M$  entries,  $B = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M] \in \mathbb{R}^{D \times M}$ , locality-constrained linear coding uses the following criterion to compute the codes  $C = [\mathbf{c}_1, \mathbf{c}_2 \dots \mathbf{c}_N]$ :

$$\begin{aligned} \underset{C}{\operatorname{argmin}} \sum_{i=1}^N \|\mathbf{x}_i - B\mathbf{c}_i\|^2 + \lambda \|\mathbf{d}_i \odot \mathbf{c}_i\|^2 \\ \text{s.t. } \mathbf{1}^T \mathbf{c}_i = 1, \quad \forall i \end{aligned} \quad (2)$$

where  $\odot$  denotes the element-wise multiplication and

$$\mathbf{d}_i = \exp\left(\frac{\operatorname{dist}(\mathbf{x}_i, B)}{\sigma}\right) \quad (3)$$

where  $\operatorname{dist}(\mathbf{x}_i, B) = [|\mathbf{x}_i - \mathbf{b}_1|, \dots, |\mathbf{x}_i - \mathbf{b}_M|]^T$  and  $\sigma$  is a decay parameter. A fast approximation to LLC described in [20] was used to speed up the encoding process. Specifically, instead of solving Eq. (2), the  $K (< D < M)$  nearest neighbours of  $\mathbf{x}_i$  in  $B$  were considered as the local bases  $B_i$  and a much smaller linear system (Eq. (4)) was solved to get the local linear codes.

$$\begin{aligned} \underset{C}{\operatorname{argmin}} \sum_{i=1}^N \|\mathbf{x}_i - B_i \mathbf{c}_i\|^2 \\ \text{s.t. } \mathbf{1}^T \mathbf{c}_i = 1, \quad \forall i \end{aligned} \quad (4)$$

We learned separate dictionaries of size  $M$  for each feature type. Max-pooling was used to aggregate the locality-constrained linear codes. For each feature type, a 2-level cell pyramid was used to capture spatial structure. At the first level, sparse codes from the whole cell were pooled to get a feature vector of size  $M$ . At the second level, feature vectors were computed from the inner region and from the border region of each cell respectively. These three feature vectors were concatenated to give a  $3M$ -vector. Finally, the four feature types were concatenated to give a  $12M$ -vector on which classification was based.

## C. Classification

Augmenting a classifier’s training set with rotated versions of the images it contains may improve classification performance but it also increases memory requirements in the implementation we used. Instead we used an ensemble of one-vs-rest, multi-class, linear SVMs; the ensemble consisted of four SVMs, one trained on the original training set images, and others trained on images after they were rotated through  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$  respectively. At test time, each test image

	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
Golgi.	5.66	3.73	20.72	2.35	9.53	58.01

TABLE I: Confusion matrix for leave-one-specimen-out experiments on I3A Task-1 dataset.

was rotated by  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ , and each rotated image was then given to the ensemble. This resulted in a set of 16 classification scores for each class. Scores were treated as probabilities using Platt rescaling [21]. The final classification decision was made by averaging these probabilistic scores and selecting the highest scoring class.

### III. EXPERIMENTS

#### A. Dataset

The I3A-2014 Task-1 dataset was collected between 2011 and 2013 at the Sullivan Nicolaidis Pathology laboratory, Australia. The training dataset contains 13,596 cell images collected from 83 different specimens and assigned to the following classes: homogeneous, speckled, nucleolar, centromere, golgi, and nuclear membrane (see Fig. 1). All the cell images are monochrome and approximately  $70 \times 70$  pixels in size. We also made use of cell images segmented from the I3A-2014 Task-2 dataset which contains images of specimens from 7 different categories: homogeneous, speckled, nucleolar, centromere, golgi, nuclear membrane and mitotic spindle; see Fig. 3(a) and (c) for examples. (We did not use the *mitotic spindle* images in the experiments reported in this paper). A companion paper describes experiments classifying specimens into seven classes [22].

#### B. Experimental setting

All features were densely extracted from patches of size  $12 \times 12$ ,  $16 \times 16$ , and  $20 \times 20$  pixels with a step-size of 2 pixels. The following parameter settings were used:

- mLP: a 3-resolution version with 8 sampling points at each resolution was used as shown in Fig. 2. The parameters of the Gaussian filters at each sampling point were selected as in [10].
- RP: The dimension  $D'$  of each linearised patch was reduced to  $D = 300$  whenever  $D' > 300$ .
- IH: Local intensity histograms of 256 bins were used.

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 1500. We use the implementation of LLC from [20] with 10 nearest neighbours ( $K = 10$ ). The Liblinear package [23] was used to build the ensemble classifiers. The Mean Class Accuracy (MCA) was used as the evaluation metric. It is defined as

$$\text{MCA} = \frac{1}{K} \sum_{k=1}^K \text{CCR}_k \quad (5)$$

where  $\text{CCR}_k$  is the correct classification rate for class  $k$  and  $K$  is the number of classes.

#### C. Experiment 1: Comparison of different features

We compared the performance obtained when using different features. *Leave-one-specimen-out* experiments were carried out using the specimen IDs provided to split the data into training and validation sets. Since 83 different specimens were available, we used images from 82 specimens for training in each fold. Table II reports the MCA for each feature type as well as for their combination. RP gave a slightly better MCA than the other features. IH gave the worst result. Combining all the features together resulted in a small increase to an MCA of 80.25%. Table I reports the confusion matrix in this case.

Feature type.	MCA(%)
rSIFT	78.00
mLP	78.63
IH	61.26
RP	79.60
All	80.25

TABLE II: MCA obtained when using different features. (All = rSIFT+mLP+IH+RP).

#### D. Experiment 2: Evaluation on cell images extracted from the Task-2 dataset

An automatic procedure was used to select cells from the Task-2 dataset given the segmentation masks provided with that dataset. Firstly, all disjoint regions were identified in the segmentation mask images using connected component analysis. Secondly, eccentricity values were calculated for each connected component. Finally, low-eccentricity components that could be bounded by an  $80 \times 80$  square with which no other component overlapped were selected. Approximately 5000 isolated cells were selected in this way. This is illustrated in Fig. 3 where red bounding boxes denote cell images that were extracted.

We trained an ensemble classifier using all the images from the Task-1 training dataset and then tested it on the cell images extracted from the Task-2 dataset. The results are reported in Table III; an MCA of 85.5% was obtained.

### IV. CONCLUSIONS

We developed a pattern recognition system to classify IIF images of HEp-2 cells into six classes. Two training regimes

	Homo.	Spec.	Nucl.	Cent.	NuMe.	Golgi
Homo.	64.36	29.68	1.70	0.10	3.51	0.63
Spec.	4.91	90.56	0.50	2.00	1.70	0.30
Nucl.	1.44	1.88	95.89	0.11	0.22	0.44
Cent.	0.32	9.78	6.15	83.62	0.00	0.10
NuMe.	3.70	2.16	0.30	0.00	92.28	1.54
Golgi.	0.10	1.66	3.64	0.72	7.38	86.47

TABLE III: Confusion matrix obtained when testing on images extracted from the I3A Task-2 dataset.

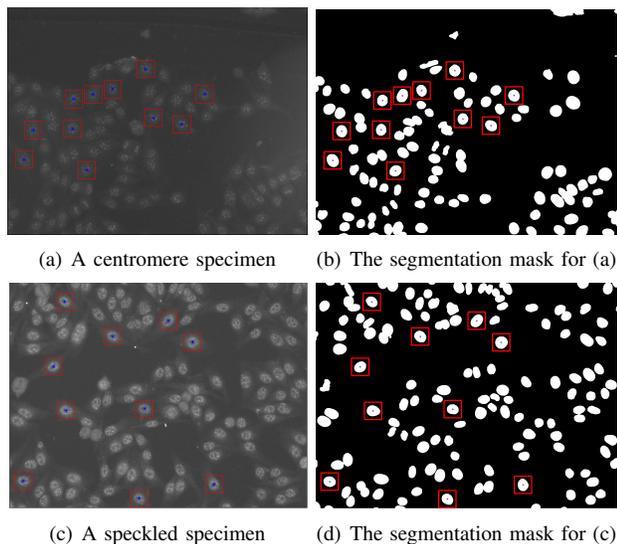


Fig. 3: Sample specimen images from I3A Task-2 dataset. The red bounding boxes indicate the cell images which are automatically extracted from these specimen images.

were used to generate submissions to the I3A Contest associated with the ICPR 2014 Workshop. The first regime used only data made available in the Task 1 training set; leave-one-specimen-out experiments with this regime predicted a MCA of 80.25%. This regime was also tested on about 5000 cell images automatically extracted from the Task 2 training dataset, giving a MCA of 85.5%. The second regime trained classifiers on a data set consisting of the Task 1 training set together with these additional 5000 cell images.

#### REFERENCES

- [1] A. Wiliem, Y. Wong, C. Sanderson, P. Hobson, S. Chen, and B. Lovell, "Classification of human epithelial type 2 cell indirect immunofluorescence images via codebook based descriptors," in *IEEE Workshop on Applications of Computer Vision*, 2013.
- [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 Trans. 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 Int. Conf. on Bioinformatics Bioengineering*, 2012.
- [4] L. Shen, J. Lin, 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 Trans. on Medical Imaging*, 2013.
- [6] P. Foggia, G. Percannella, A. Saggese, and M. Vento, "Pattern recognition in stained hep-2 cells: Where are we now?" *Pattern Recognition*, 2014.
- [7] I. Ersoy, F. Bunyak, J. Peng, and K. Palaniappan, "Hep-2 cell classification in IIF images using shareboost," in *21st Int. Conf. on Pattern Recognition*, 2012.
- [8] 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 Int. Conf. on Pattern Recognition*, Nov 2012.
- [9] G. Sharma, S. Hussain, and F. Jurie, "Local higher-order statistics for texture categorization and facial analysis," in *IEEE European Conf. on Computer Vision*, 2012.
- [10] T. Maenpaa, "The local binary pattern approach to texture analysis - extensions and applications," Ph.D. dissertation, University of Oulu, 2003.
- [11] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast Retina Keypoint," in *IEEE Computer Vision and Pattern Recognition*, 2012.
- [12] S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *IEEE Int. Conf. on Computer Vision*, 2011.
- [13] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2010.
- [14] R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval," in *IEEE Computer Vision and Pattern Recognition*, 2012.
- [15] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," in *Knowledge Discovery and Data Mining*, 2001.
- [16] L. Liu and P. Fieguth, "Texture classification from random features," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2012.
- [17] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *IEEE Int. Conf. on Computer Vision*, 2003.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in large margin classifiers*, 1999.
- [22] S. Manivannan, W. Li, S. Akbar, R. Wang, J. Zhang, and S. J. McKenna, "Hep-2 specimen classification using multi-resolution local patterns and SVM," in *ICPR I3A workshop on pattern recognition techniques for IIF images*, 2014.
- [23] 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.