Automatic classification of Human Epithelial type 2 cell Indirect Immunofluorescence images using Cell Pyramid Matching

Arnold Wiliema,*, Conrad Sandersonb,c, Yongkang Wondo, Peter Hobsone, Rodney F. Minchinb, Brian C. Lovella

a The University of Queensland, QLD 4072, Australia
b NICTA, GPO Box 2434, Brisbane, QLD 4001, Australia
c Queensland University of Technology, QLD 4000, Australia
d National University of Singapore, Singapore
e Sullivan Nicolaides Pathology, Australia

1. Introduction

The Anti-Nuclear Antibody (ANA) test is commonly used by clinicians to identify the existence of Connective Tissue Diseases such as Systemic Lupus Erythematosus, Sjögren’s syndrome, and Rheumatoid Arthritis [1]. The hallmark protocol for doing this is through Indirect Immunofluorescence (IIF) protocol. The IIF protocol on HEp-2 cells has been the hallmark method to identify the presence of ANAs, due to its high sensitivity and the large range of antigens that can be detected. However, it suffers from numerous shortcomings, such as being subjective as well as time and labour intensive. Computer Aided Diagnostic (CAD) systems have been developed to address these problems, which automatically classify a HEp-2 cell image into one of its known patterns (e.g., speckled, homogeneous). Most of the existing CAD systems use handpicked features to represent a HEp-2 cell image, which may only work in limited scenarios. We propose a novel automatic cell image classification method termed Cell Pyramid Matching (CPM), which is composed of regional histograms of visual words coupled with the Multiple Kernel Learning framework. We present a study of several variations of generating histograms and show the efficacy of the system on two publicly available datasets: the ICPR HEp-2 cell classification contest dataset and the SNPHEp-2 dataset.

This paper describes a novel system for automatic classification of images obtained from Anti-Nuclear Antibody (ANA) pathology tests on Human Epithelial type 2 (HEp-2) cells using the Indirect Immunofluorescence (IIF) protocol. The IIF protocol on HEp-2 cells has been the hallmark method to identify the presence of ANAs, due to its high sensitivity and the large range of antigens that can be detected. However, it suffers from numerous shortcomings, such as being subjective as well as time and labour intensive. Computer Aided Diagnostic (CAD) systems have been developed to address these problems, which automatically classify a HEp-2 cell image into one of its known patterns (e.g., speckled, homogeneous). Most of the existing CAD systems use handpicked features to represent a HEp-2 cell image, which may only work in limited scenarios. We propose a novel automatic cell image classification method termed Cell Pyramid Matching (CPM), which is composed of regional histograms of visual words coupled with the Multiple Kernel Learning framework. We present a study of several variations of generating histograms and show the efficacy of the system on two publicly available datasets: the ICPR HEp-2 cell classification contest dataset and the SNPHEp-2 dataset.

Keywords: Indirect Immunofluorescence tests Bag of visual words HEp-2 cell classification

* Corresponding author. Tel.: +61 7 3365 1643
E-mail addresses: a.wiliem@uq.edu.au, arnold.wiliem@ieee.org (A. Wiliem).

Please cite this article as: A. Wiliem, et al., Automatic classification of Human Epithelial type 2 cell Indirect Immunofluorescence images using Cell Pyramid Matching, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.10.014
Support Vector Machine (SVM) classifier for each region. The histograms from all regions are then fed into a spatial pyramid of levels, with each level containing non-overlapping patches, and the average histogram of visual words is computed.

Spatial Pyramid Matching (SPM) was proposed to exploit this kind of vehicle (i.e. whether it is a motorcycle, car, or truck). For example, one needs to determine which patch-level features and encoding technique is most suitable for the task at hand. Our previous work [15] proposed a Dual-Region (DR) structure specifically designed for cell images.

A single histogram of visual words of an image only describes information in that image. However, a mixing coefficient which indicates relative region importance needs to be empirically determined.

The work presented in this paper extends our previous study by proposing a novel approach termed Cell Pyramid Matching (CPM), which incorporates the positive aspects of the SPM and DR approaches, while omitting their negative aspects. Furthermore, we show that combining the CPM approach with a learning framework known as Multiple Kernel Learning [30] (where several variants of CPM are employed concurrently) leads to state-of-the-art performance on the SNPHep-2 dataset [15], and is comparable to the state-of-the-art on the ICPRcontest dataset [7]. We continue this paper as follows. We first delineate the HEp-2 cell classification task in Section 2. In Section 3 we discuss various forms of BoW descriptors and the proposed CPM approach. Section 4 is devoted to experiments and discussions, followed by the main findings in Section 5.

2. HEp-2 cell classification task

Each positive HEp-2 cell image is represented as a three-tuple \((I, M, \delta)\) which consists of (i) the Fluorescein Isothiocyanate (FITC) image channel \(I\); (ii) a binary cell mask image \(M\) which can be manually defined, or extracted from the (DAPI) image channel [5]; and (iii) the fluorescence intensity \(\delta \in \{\text{strong}, \text{weak}\}\) which specifies whether the cell is a strong positive or weak positive. Strong content to be modelled and compared separately, leading to higher recognition accuracies than using only one average region (i.e. single histogram) for each cell image. An advantage of this approach is that it has lower dimensionality than SPM (i.e. approximately 90% less), leading to considerably lower storage requirements. However, a mixing coefficient which indicates relative region importance needs to be empirically determined.

Please cite this article as: A. Wiliem, et al., Automatic classification of Human Epithelial type 2 cell Indirect Immunofluorescence images using Cell Pyramid Matching, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.10.014

Table 1

<table>
<thead>
<tr>
<th>Approach</th>
<th>Descriptors</th>
<th>Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perner et al. [13]</td>
<td>Textural</td>
<td>Decision Tree</td>
</tr>
<tr>
<td>Hiemann et al. [5]</td>
<td>Structural; textural</td>
<td>Logistic Tree</td>
</tr>
<tr>
<td>Elbschiger et al. [11]</td>
<td>Image statistics; cell shape; textural</td>
<td>Nearest Neighbour (NN)</td>
</tr>
<tr>
<td>Hsieh et al. [12]</td>
<td>Image statistics; textural</td>
<td>Learning Vector Quantisation (LVQ)</td>
</tr>
<tr>
<td>Soda et al. [6]</td>
<td>Specific set of features (e.g., textural) for each class</td>
<td>Multi-Expert System</td>
</tr>
<tr>
<td>Cordelli et al. [10]</td>
<td>Image statistics; textural; morphological</td>
<td>AdaBoost</td>
</tr>
<tr>
<td>Strandmark et al. [14]</td>
<td>Morphological; image statistics; textural</td>
<td>Random Forest</td>
</tr>
<tr>
<td>Ali et al. [16]</td>
<td>Biological-inspired descriptor</td>
<td>Boosted k-NN Classifier</td>
</tr>
<tr>
<td>Theodorakopoulos et al. [17]</td>
<td>Morphological and texture features</td>
<td>Kernel SVM (RVM)</td>
</tr>
<tr>
<td>Thibault et al. [18]</td>
<td>Morphological and texture features</td>
<td>Linear Regression, Random Forest</td>
</tr>
<tr>
<td>Ghosh et al. [19]</td>
<td>Histograms of oriented gradients, image statistics and textural</td>
<td>SVM</td>
</tr>
<tr>
<td>Li et al. [20]</td>
<td>Textural and image statistics</td>
<td>SVM</td>
</tr>
<tr>
<td>Di Cataldo et al. [21]</td>
<td>GLCM and DCT features</td>
<td>SVM</td>
</tr>
<tr>
<td>Snell et al. [22]</td>
<td>Texture and shape</td>
<td>Multistage classifier</td>
</tr>
<tr>
<td>Ersoy et al. [23]</td>
<td>Local shape measures, gradient and textural</td>
<td>ShareBoost</td>
</tr>
<tr>
<td>Wiliem et al. [15]</td>
<td>Bag of visual words with dual-region structure</td>
<td>Nearest Convex Hull Classifier (NCH)</td>
</tr>
</tbody>
</table>

Fig. 1. Examples of strong positive ANA specimens. See Fig. 2 for images of individual cells.
positive images normally have more defined details, while weak positive images are duller.

Let $Y$ be a probe image $Y = (I, M, \delta)$, and $\epsilon$ be its class label. Given a gallery set $G = (\{I, M, \delta\}, \{I, M, \delta\}, \ldots, \{I, M, \delta\})$, the task of a classifier $\varphi : Y \times G \rightarrow \epsilon$ is to produce $\epsilon$, where ideally $\epsilon = \epsilon$.

We consider six HEp-2 cell patterns [2] listed below; example images are shown in Fig. 2:

1. **Homogeneous**: A uniform diffuse fluorescence covering the entire nucleoplasm sometimes accentuated in the nuclear periphery.
2. **Coarse speckled**: Densely distributed, variously sized speckles, generally associated with larger speckles, throughout nucleoplasm of interphase cells; nucleoli are negative.
3. **Fine speckled**: Fine speckled staining in a uniform distribution, sometimes very dense so that an almost homogeneous pattern is attained; nucleoli may be positive or negative.
4. **Nucleolar**: Brightly clustered larger granules corresponding to decoration of the fibrillar centers of the nucleoli as well as the coiled bodies.
5. **Centromere**: Rather uniform discrete speckles located throughout the entire nucleus.
6. **Cytoplasmic**: A very fine dense granular to homogeneous staining or cloudy pattern covering part or the whole cytoplasm.

## 3. Bag of words classification systems

A conceptual illustration of the general approach for obtaining histograms of visual words from HEp-2 cell images is shown in Fig. 3. Each cell image is first resized into a canonical size and then divided into small overlapping patches. The patches are in turn represented by patch-level features. The local histogram from each patch is then extracted by using the pre-trained visual word dictionary. The local histograms located inside a region are pooled to compute the overall histogram for the region. Finally, the cell image is represented by a set of regional histograms; examples of regional structures are shown in Fig. 4.

In the following subsections, we first describe low-level patch-level features, followed by presenting various methods for local histogram extraction. The regional structures (i.e. SPM, DR and the proposed CPM) are discussed afterwards. Finally, we overview a framework known as Multiple Kernel Learning (MKL), which combines information captured from several descriptors.

### 3.1. Patch-level feature extraction

Given a HEp–2 cell image $(I, M, \delta)$, both the FITC image $I$ and mask image $M$ are divided into small overlapping patches $P_I = \{p_{I_1}, p_{I_2}, \ldots, p_{I_n}\}$ and $P_M = \{p_{M_1}, p_{M_2}, \ldots, p_{M_m}\}$. The division is accomplished in the same manner for both images, resulting in each patch in the FITC image having a corresponding patch in the mask image. Let $f$ be a patch-level feature extraction function $f : p_{I} \rightarrow x$, where $x \in \mathbb{R}^d$, $P_I$ now can be represented as $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$.

For evaluation purposes, we selected two popular patch-level feature extraction techniques, based on the Scale Invariant Feature Transform (SIFT) and the Discrete Cosine Transform (DCT). The SIFT descriptor is invariant to uniform scaling, orientation and partially invariant to affine distortion and illumination changes [31]. These attributes are advantageous in this classification task as cell images are unaligned and have high within class variabilities. DCT based features proved to be effective for face recognition in video surveillance [25,27]. By using only the low frequency DCT coefficients (essentially a low-pass filter), each patch representation is relatively robust to small alterations [25]. We follow the extraction procedures for SIFT and DCT as per [32,25].

The dictionary of visual words, denoted as $\mathcal{D}$, is trained from patches extracted in sliding window manner from training cell images. Each histogram encoding method has specific dictionary training procedure.

### 3.2. Generation of local histograms

For each patch-level feature that belongs to region $r$, $x_r \in \mathcal{X}_r$, a local histogram $h_r$ is obtained. In this work we consider three prominent histogram encoding methods: (1) Vector Quantisation; (2) Soft Assignment; (3) Sparse Coding. The methods are elucidated below.

#### 3.2.1. Vector Quantisation (VQ)

Given a set $\mathcal{D}$, the dictionary of visual words, the $i$-th dimension of local histogram $h_r$ for patch $x_r$ is computed via

$$h_{ri} = \begin{cases} 1 & \text{if } i = \arg \min_{i \in \mathcal{D}} \text{dist}(x_r, d_i) \\ 0 & \text{otherwise} \end{cases}$$

---

Please cite this article as: A. Wiliem, et al., Automatic classification of Human Epithelial type 2 cell Indirect Immunofluorescence images using Cell Pyramid Matching, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.10.014
where $\text{dist}(\mathbf{x}_j, \mathbf{d}_k)$ is a distance function between $\mathbf{x}_j$ and $\mathbf{d}_k$, while $\mathbf{d}_k$ is the $k$-th entry in the dictionary $\mathcal{D}$ and $|\mathcal{D}|$ is the number of elements in $\mathcal{D}$. The dictionary is obtained via the $k$-means algorithm [33] on training patches, with the resulting cluster centers representing the entries in the dictionary.

The VQ approach is considered as a hard assignment approach since each image patch is only assigned to one of the visual words. Such hard assignment can be sensitive to noise [26].

### 3.2.2. Soft Assignment (SA)

In comparison to the VQ approach above, a more robust approach is to apply a probabilistic method [25]. Here the visual dictionary $\mathcal{D}$ is a convex mixture of Gaussians. The $i$-th dimension of the local histogram for $\mathbf{x}_j$ is calculated by

$$h_{i,j} = \frac{w_i p_i(\mathbf{x}_j)}{\sum_{k=1}^{|\mathcal{D}|} w_k p_k(\mathbf{x}_j)}$$

(2)

where $p_i(\mathbf{x})$ is the likelihood of $\mathbf{x}$ according to the $i$-th component of the visual dictionary $\mathcal{D}$:

$$p_i(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \left| \mathbf{C}_i \right|^{1/2} } \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{\mu}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{\mu}_i) \right]$$

(3)

with $w_i$, $\mathbf{\mu}_i$, and $\mathbf{C}_i$ representing the weight, mean and diagonal covariance matrix of Gaussian $i$, respectively. The scalar $d$ represents the dimensionality of $\mathbf{x}$. The dictionary $\mathcal{D}$ is obtained using the Expectation Maximisation algorithm [33] on training patches.

### 3.2.3. Sparse Coding (SC)

It has been observed that each local histogram produced via Eq. (2) is sparse in nature (i.e. most elements are close to zero) [27]. In other words, the SA approach described in Section 3.2.2 is an indirect Sparse Coding approach. Hence, it is possible to adapt direct Sparse Coding algorithms in order to represent each patch as a combination of dictionary atoms [34,28], which theoretically can lead to better recognition results [27]. A vector of weights $\mathbf{\theta} = [\theta_1, \theta_2, ..., \theta_D]^T$ is computed for each $\mathbf{x}_j$ by solving a minimisation problem that selects a sparse set of dictionary atoms. As the theoretical optimality of the $\ell_1$-norm minimisation solution is guaranteed [35], in this work we used

$$\min_{\mathbf{\theta}} \frac{1}{2} || \mathbf{D} \mathbf{\theta} - \mathbf{x}_j ||^2_2 + \lambda \sum_{l} || \theta_l ||_1$$

(4)

where $\| \cdot \|_p$ denotes the $\ell_p$-norm and $\mathbf{D} \in \mathbb{R}^{d \times |\mathcal{D}|}$ is a matrix of dictionary atoms. The dictionary $\mathbf{D}$ is trained by using the K-SVD algorithm [36], which is known to be suitable for obtaining...
reasonably dictionaries in similar cases, i.e. using a large number of small image patches [37].

As $\theta$ can have negative values due to the objective function in Eq. (4), we construct each local histogram using the absolute value of each element in $\theta$ [27]:

$$h_j = |\theta_1| + |\theta_2| + ... + |\theta_M|$$

(5)

Compared to Eqs. (1) and (2), obtaining the histogram using Sparse Coding is considerably more computationally intensive, due to the need to solve a minimisation problem for each patch.

3.3. Histogram pooling

Let $X_r$ be the set of patch-level features belonging to region $r$. The overall histogram representation for region $r$ is then obtained via averaging local histograms [25,27]:

$$H_d = \frac{1}{|X_r|} \sum_{k=1}^{|X_r|} h_j$$

(6)

where $|X_r|$ is the number of elements in set $X_r$. In the following subsections, we describe several possible spatial layouts for the regions and the corresponding similarity measures.

3.4. Spatial structures for multiple region descriptors

In this section we describe two existing spatial structures for using multiple regional descriptors (i.e. SPM and DR), followed by the proposed CPM approach. The conceptual diagram for each approach is shown in Fig. 4.

3.4.1. Spatial Pyramid Matching (SPM)

The regions are organised similar to an image pyramid with several levels [24]. At each level $l$, the image is divided into $(2^l)$ non-overlapping regions. For instance, at level 0 (i.e. the top level), the image is divided into $1 \times 1$ region; at level 1, the image is divided into $2 \times 2$ regions. In this work, we follow Lazebnik et al. [24] by using a three-level pyramid (i.e. levels 0, 1 and 2): $1 \times 1$, $2 \times 2$ and $4 \times 4$. In total, there are $1 + 4 + 16 = 21$ regions. The pyramid match kernel is used to measure the similarities between two images [24]:

$$K(H_1, H_2) = \frac{1}{2^l} \sum_{i=1}^{h(i)} G(H_1^{(i)}, H_2^{(i)}) + \sum_{l=1}^{L} \frac{1}{2^l} \sum_{i=1}^{h(i)} G(H_1^{(i)}, H_2^{(i)})$$

(7)

where $H_d^{(i)}$ is the $i$-th regional histogram of level $l$ of the $k$-th image, while $L$ is the maximum number of levels (i.e. $L=2$). $G(\cdot, \cdot)$ is a histogram intersection kernel, defined as [24]

$$G(H_1^{(i)}, H_2^{(i)}) = \sum_j H_{1j}^{(i)} H_{2j}^{(i)}$$

(8)

where $H_{kj}^{(i)}$ is the $j$-th dimension of a regional histogram for level $l$ and region $r$ of image $k$.

3.4.2. Dual Region (DR)

Each cell is divided into an inner region, which covers the cell content, and an outer region, which contains information related to cell edges and shape [15]. To this end, each patch is first classified as either belonging to the inner or outer region by inspecting its corresponding mask patch. More specifically, let $X = X^{(i)} \cup X^{(o)}$, with $X^{(i)}$ representing the set of outer patches, and $X^{(o)}$ the set of inner patches. The classification of patch $p$ into a region is done via

$$p_i \in \begin{cases} X^{(i)} & \text{if } fg(p_m) \in \{\tau_1, \tau_2\} \\ X^{(o)} & \text{if } fg(p_m) \in \{0, 1\} \end{cases}$$

(9)

where $p_i$ is the corresponding mask patch; $fg(p_m) \in \{0, 1\}$ computes the normalised occupation count of foreground pixels from mask patch $p_m$; $\tau_1$ is the minimum foreground pixel occupation of a patch belonging to the outer region; $\tau_2$ is the minimum pixel occupation of a patch belonging to the inner region. Note that the size of the inner and outer regions is indirectly determined via Eq. (9). Based on preliminary experiments, we have found that $\tau_1 = 0.3$ and $\tau_2 = 0.8$ provide good results. Unlike SPM, there are only two regional histograms required to represent a cell image. As such, the DR descriptor is $(21-2)/21 = 90\%$ smaller than SPM.

The similarity between two images is defined via

$$\mathcal{K}(H_1, H_2) = \exp[-\text{dist}(H_1, H_2)]$$

(10)

Adapting [25], dist$(H_1, H_2)$ is defined by

$$\text{dist}(H_1, H_2) = \alpha_1 \|H_1^{(i)} - H_2^{(i)}\|_1 + \alpha_2 \|H_1^{(o)} - H_2^{(o)}\|_1$$

(11)

where $H_1^{(i)}$ and $H_2^{(i)}$ are the inner and outer region histograms of image $k$, respectively; $\alpha_1$ and $\alpha_2$ are positive mixing parameters which weight the importance of information contained for each region, under the constraint of $\alpha_1 + \alpha_2 = 1$. A possible drawback of the DR approach is that determining good settings for the $\tau_1$, $\tau_2$ and $\alpha_1$ parameters is currently a time consuming procedure, where a heuristic or grid-based search is used. Furthermore, not all valid settings in such a search might be evaluated, which can lead to sub-optimal discrimination performance.

3.4.3. Cell Pyramid Matching (CPM)

The proposed CPM approach combines the advantages of both SPM and DR structures. It adapts the idea of using a pyramid structure from SPM as well as the inner and outer regions from DR. Unlike SPM, CPM has only two levels: level 0 which comprises the whole cell region, and level 1 which comprises the inner and outer regions. The advantages of this combination are two fold: (1) the CPM descriptor only requires 3 histograms to represent a cell image, and is hence $(21-3)/21 = 95\%$ smaller than SPM; (2) as the CPM follows the SPM construct, it employs the pyramid match kernel, which eliminates the mixing parameters in DR.

3.5. Multiple Kernel Learning

Fusing information provided by various image descriptors and spatial structures (each with a dedicated kernel, as shown above) may improve discrimination ability. If the given descriptors are at least partially capturing differing information. To that end we have elected to use the Multiple Kernel Learning (MKL) framework, which aims to learn the optimum mixing of various kernels [30]. Let $(x_i, y_i) \in \mathcal{G}$ be the training set, where $x_i$ is a feature vector and $y_i \in \{-1, 1\}$ is the corresponding groundtruth label. The MKL classifier is an extended form of the SVM classifier, defined as

$$\hat{\omega}(q) = \sum_{k=1}^{M} \beta_k \mathcal{K}(\mathbf{q}, \mathbf{x}_k) + \mathbf{b}$$

(12)

where $q$ is a query point, $\mathbf{x}_k \in \mathcal{G}$ is the $k$-th training point, $\beta_k$ is the “importance” weight of the $k$-th training point, $\mathbf{b}$ is the bias term, and $\mathcal{K}(\cdot, \cdot)$ is a combination kernel defined as

$$\mathcal{K}(\mathbf{a}, \mathbf{c}) = \sum_{m=1}^{M} w_m \mathcal{K}_m(\mathbf{a}, \mathbf{c})$$

(13)

where $\mathcal{K}_m(\cdot, \cdot)$ is the $m$-th kernel, with $w_m$ its corresponding mixing weight, under the constraints of $w_m \geq 0$ and $\sum w_m = 1$.

1 Here we have presented a binary classification problem. However, it can be easily generalised into a multi-class problem [30].
Without losing generality, $K_m(\cdot, \cdot)$ can be the kernel defined in Eq. (7) or (10).

In the MKL learning scheme, the importance weights and kernel mixing weights are learned together. In this work we employ the SimpleMKL method for learning [30], which employs a convex and smooth objective function.

4. Experiments and results

In this section we first compare the performance of six variants of the BoW descriptor, where each of the two low-level feature extraction techniques (SIFT and DCT) is coupled with three possible methods for generating the histograms of visual words (VQ, SA, and SC). The six variants are used within the framework of the DR, SPM and CPM spatial structures. We then show that by fusing the two approaches (DCT-SA CPM and DCT-VQ CPM) via the MKL framework leads to an increase in recognition rates. Finally, we compare the MKL based system against three recently proposed systems in the literature. The various systems were implemented with the aid of the Armadillo C++ library [38].

4.1. Datasets: ICPRContest and SNP HEp-2

For the experiments we used two publicly available datasets, briefly described below, in order to evaluate applicability of the various systems to differing assays and microscope parameters.

The ICPR HEp-2 Cell Classification Contest (ICPRContest) Dataset [39] contains 1457 cells extracted from 28 specimen images. It contains six patterns: centromere, coarse speckled, cytoplasmic, fine speckled, homogeneous, and nucleolar. Each specimen image was acquired by means of fluorescence microscope (40-fold magnification) coupled with 50W mercury vapour lamp and with a CCD camera. The cell image masks were hand labelled. See Fig. 2 for examples. We followed the ICPR contest evaluation protocol for this dataset which only has one pair of train and test sets.

The SNP HEp-2 Cell (SNPHEp-2) Dataset [15] was obtained between January and February 2012 at Sullivan Nicolaides Pathology laboratory, Australia. This dataset has five patterns: centromere, coarse speckled, fine speckled, homogeneous and nucleolar. The 18-well slide of HEp-2000 IIF assay from Immuno Concepts N. A. Ltd. with screening dilution 1:80 was used to prepare 40 specimens. Each specimen image was captured using a monochrome high dynamic range cooled microscopy camera, which was fitted on a microscope with a plan-Apochromat 20 ×/0.8 objective lens and an LED illumination source. 4′,6-diamidino-2-phenylindole (DAPI) image channel was used to automatically extract the cell image masks.

There are 1884 cell images extracted from 40 specimen images. The specimen images are divided into training and testing sets with 20 images each (4 images for each pattern). In total there are 905 and 979 cell images extracted for training and testing. Five-fold validations of training and testing were created by randomly selecting the training and test images. Both training and testing in each fold contain around 900 cell images (approx. 450 cell images each). Examples are shown in Fig. 2.

Due to possible varying filtering effects caused by image capture equipment, tuning, operator bias, and/or environmental conditions (all of which can result in low-pass filtering), cell images with the same pattern can simply differ due to gross mismatches in frequency spectra. In turn this can lead to a degradation in recognition accuracy [41]. To counteract this undesirable effect, and to ensure that a canonical image size is used, images from both datasets were downsampled by two to approximately 64 × 64 pixels.

4.2. Combinations of local features, histogram generation and spatial structures

We follow Lazebnik et al. [24] and Wiliem et al. [15] for SPM and DR implementations, respectively. The SVM classifier is used in all cases, with the kernels specified in Eqs. (7) and (10) for the SPM and DR methods, respectively. As noted in Section 3.4.3, a form of Eq. (7) is used as the SVM kernel for the CPM method.

As there are three histogram encoding methods (i.e., VQ, SA and SC) and two patch-level features (i.e., SIFT and DCT), there are six variants of the BoW descriptor. For clarity, each variant is denoted by [patch-level features]–[histogram encoding method]. For example, the variant using DCT as its patch-level features and VQ as its encoding method is called DCT-VQ.

The results, presented in Table 2, indicate that in most cases the proposed CPM system obtains the best performance, suggesting that it is taking advantage of both the specialised spatial layout for cells inherited from the DR approach, and the pyramid match kernel inherited from the SPM approach. The results also show that in most cases the use of DCT based patch-level feature extraction leads to better performance than using SIFT based feature extraction. We conjecture that DCT obtains better performance as the SIFT descriptor needs a larger spatial support and is hence more likely to be affected by image deformations. Specifically, SIFT divides a given image patch into 4 × 4 subregions in which each has 4 × 4 bins, followed by extracting gradient information from each subregion [31]. Therefore, SIFT needs a spatial support of at least 16 × 16 pixels, which is relatively large when compared to the canonical cell image size of 64 × 64. In contrast, standard DCT requires a much smaller spatial support of 8 × 8 pixels, making it less susceptible to image deformations.

4.3. Fusion via Multiple Kernel Learning

Based on the results obtained in the previous experiment, we have selected the overall top three systems (DCT-SA CPM, DCT-VQ CPM, SIFT-SC CPM) and evaluated fusing them via the MKL framework. The results for various mixtures of the three systems are shown in Table 3.

By using the mixture that obtains the best overall performance across both datasets, i.e. DCT-SA CPM and DCT-VQ CPM, the recognition rate improves from 65.9% to 67.4% on the ICPRContest dataset, and from 81.2% to 82.4% on the SNPHEp-2 dataset.

Table 2

Performance comparison of BoW descriptor variants on the ICPRContest and SNPHEp-2 datasets, using various spatial configurations (DR, SPM, CPM). The scores for SNPHEp-2 dataset shown as average correct classification rate. DR = Dual Region; SPM = Spatial Matching Pyramid; CPM = Cell Pyramid Matching.

<table>
<thead>
<tr>
<th>Descriptor variant</th>
<th>ICPRContest</th>
<th>SNPHEp-2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DR</td>
<td>SPM</td>
</tr>
<tr>
<td>DCT-SA</td>
<td>64.9 64.3</td>
<td>65.9 79.5</td>
</tr>
<tr>
<td>DCT-VQ</td>
<td>54.5 57.1</td>
<td>61.2 80.7</td>
</tr>
<tr>
<td>DCT-SC</td>
<td>52.6 57.9</td>
<td>57.2 71.0</td>
</tr>
<tr>
<td>SIFT-SA</td>
<td>51.6 57.5</td>
<td>47.8 71.6</td>
</tr>
<tr>
<td>SIFT-VQ</td>
<td>55.6 53.8</td>
<td>59.0 64.9</td>
</tr>
<tr>
<td>SIFT-SC</td>
<td>60.8 59.9</td>
<td>62.1 76.2</td>
</tr>
</tbody>
</table>

2 It is assumed that the cell images have been extracted from specimen images either via a manual or automated approach such as background subtraction [40].

3 The SNPHEp-2 dataset is available for download at http://itee.uq.edu.au/~lowell/snphep2/.

Note that while it is possible to fuse information from more systems, there is no guarantee that this will always lead to better
performance [42,30,43]. In further experiments (not shown here) we have found that combining more systems can decrease performance. As the main aim was to show the possible advantage of using MKL, we leave a detailed study for future work.

4.4. Comparative evaluation of systems

In this section we compare the MKL based approach (where information from DCT-SA CPM and DCT-VQ CPM is fused) against three recently proposed systems in Wiliem et al. [15], Cordelli et al. [10] and Strandmark et al. [14].

The system in [15], denoted Wiliem, is based on the DCT-SA descriptor with the DR spatial structure and Nearest Convex Hull classifier. We denote the system in [14] by Strandmark, and used the code provided by the authors. The system employs various image statistics (e.g., mean, standard deviation) and morphological features (e.g., number of objects, area). The random forest classifier is used.

We implemented the best reported descriptor in [10], denoted by Cordelli, which is composed of features such as image energy, mean and entropy, calculated from intensity and LBP channels. The LBP channel is computed by computing the local pattern code for each pixel in the intensity channel. We selected Logistic Boosting (LogitBoost) as the classifier instead of AdaBoost as the former obtained better performance.

The results are presented in Fig. 5. On the ICPRContest dataset, the Cordelli and Strandmark systems obtain comparable performance. However, the performance of Cordelli is considerably lower than Strandmark on the SNPHEp-2 dataset, indicating that the Cordelli system is not able to generalise to various recording conditions. The Wiliem system obtains better performance than Cordelli and Strandmark on both datasets, with a marked increase over Wiliem on the ICPRContest dataset.

Table 3

<table>
<thead>
<tr>
<th>System mixture</th>
<th>ICPRContest</th>
<th>SNPHEp-2</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>(DCT-SA CPM) + (DCT-VQ CPM) + (SIFT-SC CPM)</td>
<td>66.9</td>
<td>82.5</td>
<td>74.70</td>
</tr>
<tr>
<td>(DCT-SA CPM) + (DCT-VQ CPM)</td>
<td>67.4</td>
<td>82.4</td>
<td>74.90</td>
</tr>
<tr>
<td>(DCT-SA CPM) + (SIFT-SC CPM)</td>
<td>66.3</td>
<td>81.2</td>
<td>73.75</td>
</tr>
<tr>
<td>(DCT-VQ CPM) + (SIFT-SC CPM)</td>
<td>64.0</td>
<td>79.7</td>
<td>71.85</td>
</tr>
</tbody>
</table>

Fig. 5. Performance comparison of various systems on the ICPRContest and SNPHEp-2 datasets.

4.5. Cell level and image level performance on the ICPRContest dataset

Using the proposed MKL-based system from Section 4.4, Fig. 6 shows the confusion matrix for the classification results on the ICPRContest dataset. We also present the image level classification in Fig. 7. In image level classification we simply determine the label of an image based on the most frequent cell pattern. In this setting, the MKL-based system achieves 71.4%.

We also report Leave-One-Out validation results for ICPRContest in Table 4 as well as Figs. 8 and 9. In this setting, the validation constructs 28 splits of train and test images, where for each split cells belonging to a particular specimen image are used as the test images, and the rest as training images.

5. Conclusions

In this paper we have proposed a cell classification system composed of a Cell Pyramid Matching (CPM) descriptor combined...
We also found that the proposed CPM spatial layout is more effective than SPM and DR structures. The proposed CPM also has an advantage of not having heuristic parameters and leads to a much shorter descriptor length. The experiments show that the proposed system consistently delivered high performance and is more robust than three recent CAD systems presented in [10,14,15].

**Conflict of interest statement**

None declared.

**Table 4**

Cell level classification performance for each cell image. Ce=Centromere; Ho=Homogeneous; Nu=Nucleolar; Co=Coarse speckled; Fi=Fine speckled; Cy=Cytoplasmic.

<table>
<thead>
<tr>
<th>Specimen image number</th>
<th>True class</th>
<th>Cells assigned to each class (in absolute number)</th>
<th>Cells assigned to each class (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ce Ho Nu Co Fi Cy</td>
<td>Ce Ho Nu Co Fi Cy</td>
<td>Ce Ho Nu Co Fi Cy</td>
</tr>
<tr>
<td>1</td>
<td>Homogeneous 0 43 2 0 16 0</td>
<td>0.0 0.0 70.5 3.3 0.0 26.2 0.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Fine speckled 4 8 10 19 6 1</td>
<td>8.3 16.7 20.8 39.6 12.5 2.1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Centromere 87 0 1 1 0 0</td>
<td>97.8 0.0 1.1 1.1 0.0 0.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Nucleolar 23 3 32 0 5 3</td>
<td>34.8 4.5 48.5 0.0 7.6 4.5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Homogeneous 8 26 7 0 5 1</td>
<td>17.0 55.3 14.9 0.0 10.6 2.1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Coarse speckled 44 9 0 15 0 0</td>
<td>64.7 15.3 22.1 0.0 0.0 0.0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Centromere 54 1 1 0 0 0</td>
<td>96.4 1.8 1.8 0.0 0.0 0.0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Nucleolar 25 0 30 0 0 1</td>
<td>44.6 0.0 53.6 0.0 0.0 1.8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Fine speckled 0 20 1 2 23 0</td>
<td>0.0 43.5 2.2 43.0 0.0 0.0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Coarse speckled 16 0 0 7 10 0</td>
<td>48.5 0.0 0.0 21.2 30.3 0.0</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Homogeneous 1 0 1 7 32 0</td>
<td>2.4 0.0 2.4 17.1 78.0 0.0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Coarse speckled 4 0 5 35 0 5</td>
<td>8.2 0.0 10.2 71.4 0.0 10.2</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Centromere 13 0 3 30 0 0</td>
<td>28.3 0.0 65.2 0.0 0.0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Centromere 4 19 1 9 30 0</td>
<td>6.3 30.2 1.6 14.3 47.6 0.0</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Fine speckled 9 7 0 36 3 8</td>
<td>14.3 11.1 0.0 57.1 4.8 12.7</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Centromere 36 0 2 0 0 0</td>
<td>94.7 0.0 5.3 0.0 0.0 0.0</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Coarse speckled 0 2 0 3 14 0</td>
<td>0.0 10.5 0.0 15.8 73.7 0.0</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Homogeneous 0 19 0 3 14 0</td>
<td>0.0 45.2 0.0 7.1 47.6 0.0</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Centromere 56 0 7 2 0 0</td>
<td>86.2 0.0 10.8 3.1 0.0 0.0</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Nucleolar 3 2 41 0 0 0</td>
<td>6.5 4.3 89.1 0.0 0.0 0.0</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Homogeneous 1 50 3 0 7 0</td>
<td>1.6 82.0 4.9 0.0 11.5 0.0</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Homogeneous 1 78 2 7 20 11</td>
<td>0.8 65.5 1.7 5.9 16.8 9.2</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Fine speckled 0 20 0 3 28 0</td>
<td>0.0 39.2 0.0 5.9 54.9 0.0</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Nucleolar 1 1 11 45 15 1 0</td>
<td>1.4 15.1 61.6 20.5 1.4 0.0</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Cytoplasmic 10 0 0 0 6 8</td>
<td>41.7 0.0 0.0 25.0 33.3</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Cytoplasmic 0 0 1 0 0 33</td>
<td>0.0 0.0 2.9 0.0 0.0 0.0</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Cytoplasmic 1 1 0 1 0 35</td>
<td>2.6 2.6 0.0 2.6 0.0 92.1</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Cytoplasmic 0 0 2 0 1 10</td>
<td>0.0 0.0 15.4 0.0 7.7 76.9</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 8.** Cell level confusion matrix of the proposed MKL-based system on the ICPRContest dataset using Leave-One-Out validation protocol. The average accuracy is 56.8%.

**Fig. 9.** Image level confusion matrix of the proposed MKL-based system on the ICPRContest dataset using Leave-One-Out validation protocol. The average accuracy is 64.3%.

Please cite this article as: A. Willem, et al., Automatic classification of Human Epithelial type 2 cell Indirect Immunofluorescence images using Cell Pyramid Matching, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.10.014
Acknowledgements

This research was partly funded by Sullivan Nicolaides Pathology, Australia and the Australian Research Council (ARC) Linkage Projects Grant LP130100230. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy, as well as the Australian Research Council through the ICT Centre of Excellence program. NJSZU Sensor-Enhanced Social Media (SeSaMe) Centre is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the Interactive Digital Media Programme Office.

References


Arnold Wiliem is a Research Fellow at the University of Queensland, Australia. He received his PhD in 2010 from Queensland University of Technology, Australia, and his Bachelor of Computer Science in 2006 from University of Indonesia. He has previously worked on computer vision and automated medical image analysis at NICTA before joining the University of Queensland to pursue research in the digital pathology domain. His current research interests are in the areas of machine learning, pattern recognition, and computer vision. He is a member of the IEEE.

Conrad Sanderson is a Senior Research Scientist and Team Leader at NICTA, Australia. He received the PhD degree in 2003 from Griffith University, and the MBA degree in 2012 from the University of Queensland. He has previously worked on robust speech recognition at the Advanced Telecommunication Research Laboratories (Japan), audio-visual based face learning in a non-frontal face recognition at the IDiap Research Institute (Switzerland), and image classification in infra-red images at the Centre for Sensor Signal and Information Processing (Australia), as well as natural language processing, bioinformatics and automated surveillance at NICTA. His work on innovative surveillance technologies has led to several industry awards. He has served as a reviewer for major international conferences and scientific periodicals (such as the Proceedings of the IEEE journal). His current interests are in the areas of machine learning, pattern recognition, computer vision and data mining, as well as adaptation of new technologies towards industrial use.
Yongkang Wong is a Research Fellow at the Sensor-enhanced Social Media (SeSaMe) Centre at Interactive and Digital Media Institution (IDMI), National University of Singapore. He received the Bachelor of Engineering degree in 2006 from the University of Adelaide, and the PhD degree in 2012 from the University of Queensland, Australia. He has worked as a summer research scholar at NICTA’s Canberra laboratory in 2004 and 2005, as well as a graduate researcher at NICTA’s Queensland laboratory from 2008 to 2012. His current research interests are in the areas of computer vision, machine learning, and video surveillance. He is a member of the IEEE.

Peter Hobson graduated from Queensland Institute of Technology with a Bachelor of Applied Science in Medical Technology in 1983. Peter has worked at Sullivan Nicolaides Pathology (SNP) his whole career in various disciplines, including histopathology and haematology. He is currently the Laboratory Manager of Immunology, Infectious Serology and the Bone Marrow Transplant departments. Peter has collaborated with biotech company PanBio in assay development, the University of Queensland Environmental Toxicology (EnTox), clinical trials with Esoterix laboratories and the Queensland Institute of Medical Research. He is responsible for numerous innovative IT projects within SNP. Currently he is collaborating with the University of Queensland to develop innovative digital imaging applications for microscopy.

Rodney F. Minchin received his PhD from the Department of Pharmacology at the University of Western Australia before undertaking post-doctoral studies at the National Cancer Institute in Washington DC. His major work was on pulmonary toxicology of anti-cancer drugs and chemical carcinogenesis. He returned as a QELF fellow to the University of New South Wales to investigate polyamine transport in the lung. Eventually, he returned to the University of Western Australia as a teaching and research academic, where his interest in arylamine N-acetyltransferases and gene therapy developed. Currently, he is a Professor and the Chair of Molecular Pharmacology at the University of Queensland, Australia.

Brian C. Lovell was born in Brisbane, Australia, in 1960. He received the BE in electrical engineering in 1982, the BSc in computer science in 1983, and the PhD in signal processing in 1991: all from the University of Queensland (UQ). Professor Lovell was the President of the Australian Pattern Recognition Society (1995–2000), Research Leader at NICTA (2006–2012) and Research Director of the Security and Surveillance Research group at the School of ITEE, UQ. He was Technical Co-chair of ICPR2006 in Hong Kong (Computer Vision and Image Analysis), and Program Co-chair of ICPR2008 in Tampa, Florida. He is a Senior Member of the IEEE, Fellow of the World Innovation Forum, Fellow of the IEAust, and voting member for Australia on the governing board of the International Association for Pattern Recognition since 1998. He serves on the Editorial Board of Pattern Recognition Letters and reviews for many of the major journals in the fields of Computer Vision and Pattern Recognition. In March 2005, he was awarded Number 1 author at UQ with almost 35,000 copies of his papers downloaded from the UQ library archive. His research interests are currently focused on intelligent surveillance techniques, optimal image segmentation, real-time video analysis, and face recognition.

Please cite this article as: A. Wiliem, et al., Automatic classification of Human Epithelial type 2 cell Indirect Immunofluorescence images using Cell Pyramid Matching, Pattern Recognition (2013), http://dx.doi.org/10.1016/j.patcog.2013.10.014