Classification of Human Epithelial type 2 cell indirect immunofluoresence images via codebook based descriptors

Classification of Human Epithelial Type 2 Cell Indirect Immunofluoresence Images via Codebook Based Descriptors

Arnold Wiliem, Yongkang Wong, Conrad Sanderson, Peter Hobson, Shaokang Chen, Brian C. Lovell

NICTA, PO Box 6020, St Lucia, QLD 4067, Australia *
The University of Queensland, School of ITEE, QLD 4072, Australia
Sullivan Nicolaides Pathology, Australia

Abstract

The Anti-Nuclear Antibody (ANA) clinical pathology test is commonly used to identify the existence of various diseases. A hallmark method for identifying the presence of ANAs is the Indirect Immunofluorescence method on Human Epithelial (HEp-2) cells, due to its high sensitivity and the large range of antigens that can be detected. However, the method 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 (eg., 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. In this paper, we propose a cell classification system comprised of a dual-region codebook-based descriptor, combined with the Nearest Convex Hull Classifier. We evaluate the performance of several variants of the descriptor on two publicly available datasets: ICPR HEp-2 cell classification contest dataset and the new SNPHEp-2 dataset. To our knowledge, this is the first time codebook-based descriptors are applied and studied in this domain. Experiments show that the proposed system has consistent high performance and is more robust than two recent CAD systems.

1. Introduction

In recent years, there has been increasing interest in employing image analysis techniques for various routine clinical pathology tests [9, 10, 12]. Results produced by these techniques can be incorporated into subjective analysis done by scientists, leading to test results being more reliable and consistent across laboratories [10].

The Anti-Nuclear Antibody (ANA) test is commonly used by clinicians to identify the existence of Connective Tissue Diseases such as Systemic Lupus Erythematosus, Sjorgren's syndrome, and Rheumatoid Arthritis [16]. The hallmark protocol for doing this is through Indirect Immunofluorescence (IIF) on Human Epithelial type 2 (HEp-2) cells [16, 29]. This is due to its high sensitivity and the large range expression of antigens. Despite its advantages, the IIF method is labour intensive and time consuming [3, 20]. Each ANA specimen must be examined under a fluorescence microscope by at least two scientists. This also renders the test result more subjective, and thus has low reproducibility and large inter-/intra- personnel/laboratory variabilities [10, 24]. To address these issues, it is possible to use Computer Aided Diagnostic (CAD) systems which automatically determine the HEp-2 pattern in the given cell images of a specimen [6, 7, 10, 11, 19, 24, 25]. Examples of specimen images are shown in Figure 1.

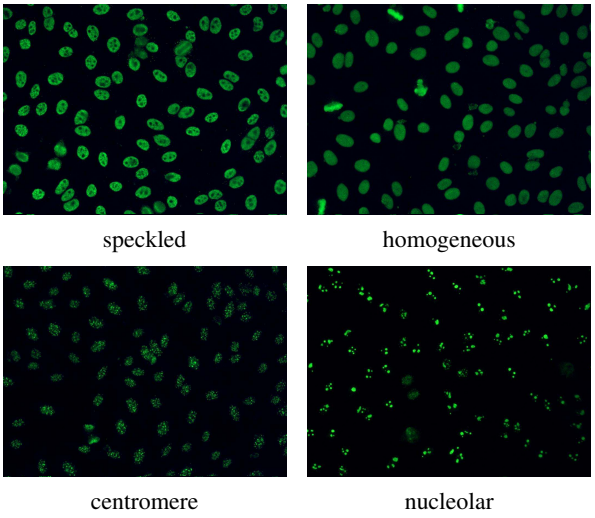

Figure 1. Examples of strong positive ANA specimens.

Properties of existing CAD systems in the literature are shown in Table 1. Most of these systems have a common trend: they use carefully handpicked features which may only work in a particular laboratory environment and/or microscope configuration. To address this, several approaches employ a large number of features and apply an automated feature selection process [10]. Another approach uses Multi Expert Systems to allow the use of a specifically tailored feature set and classifier for each HEp-2 cell pattern class [24]. Nevertheless, the generalisation ability of these systems is still not guaranteed since these were only evaluated in one particular dataset with a specific setup.

*Published in: IEEE Workshop on Applications of Computer Vision, pp. 95–102, 2013. http://dx.doi.org/10.1109/WACV.2013.6475005

Table 1. Existing CAD systems for HEp-2 cell classification.

Approach	Descriptors	Classifier
Perner *et al*. [19]	Textural	Decision Tree
Hiemann *et al*. [10]	Structural; textural (1400 features)	LogisticModel Tree
Elbischger *et al*. [7]	Image statistics; cell shape; textural (9 features)	Nearest Neighbour
Hsieh *et al*. [11]	Image statistics; textural (8 features)	Learning Vector Quantisation (LVQ)
Soda *et al*. [24]	Specific set of features (e.g. textural) for each class	Multi Expert System
Cordelli *et al*. [6]	Image statistics; textural; morphological (15 features)	AdaBoost
Strandmark *et al*. [25]	Morphological; image statistics; textural (322 features)	Random Forest

One of the most popular approaches for automatic image classification, here called the codebook approach, is to express an image in terms of a set of visual words, selected from a dictionary that has been trained beforehand [13, 23, 28]. In order to model an image, the codebook approach divides the image into small image patches, followed by patch-level feature extraction. An encoding process is then employed to compute a histogram of visual words based on these patches. Codebook-based descriptors often have higher discrimination power compared to the other image descriptors [13, 28, 30]. Thus, we argue that better classification performance can be achieved by employing such descriptors for CAD systems.

Contributions. In this work we propose the use of a dual-region codebook-based descriptor, specifically designed to exploit the nature of cell images, coupled with an adapted form of the Nearest Convex Hull classifier [17]. To our knowledge, this is the first time the codebook approach is applied and studied for the HEp-2 cell classification task. We evaluate two methods for low-level feature extraction from image patches, SIFT [15] and DCT [23], in conjunction with three methods for generating the histograms of visual words: vector quantisation [28], soft assignment [23] and sparse coding [30]. We furthermore propose a new HEp-2 cell image classification dataset, denoted as SNPHEp-2, which allows the evaluation of the robustness of CAD systems to various hardware configurations. The number of images is much larger than the existing ICPRContest dataset [8].

We continue this paper as follows. We first delineate the HEp-2 cell classification task in Section 2. In Section 3 we present the dual-region codebook-based descriptor. In Section 4 we overview the Nearest Convex Hull classifier. Section 5 is devoted to experiments and discussions. Main findings and future research avenues are given in Section 6.

2. HEp-2 Cell Classification System

Each positive HEp-2 cell image[1] is represented as a three-tuple $(\boldsymbol{I}, \boldsymbol{M}, \delta)$ which consists of: **(i)** the Fluorescein Isothiocyanate (FITC) image channel $\boldsymbol{I}$; **(ii)** a binary cell mask image $\boldsymbol{M}$ which can be manually defined, or extracted from the *4',6-diamidino-2-phenylindole* (DAPI) image channel [10]; and **(iii)** the fluorescence intensity $\delta \in \{\text{strong}, \text{weak}\}$ which specifies whether the cell is a strong positive or weak positive. Strong positive images normally have more defined details, while weak positive images are duller.

Let $\boldsymbol{Y}$ be a probe image $\boldsymbol{Y} = (\boldsymbol{I}, \boldsymbol{M}, \delta)$, and ℓ be its class label. Given a gallery set $\mathcal{G} = \{(\boldsymbol{I}, \boldsymbol{M}, \delta)_1^{\mathcal{G}}, (\boldsymbol{I}, \boldsymbol{M}, \delta)_2^{\mathcal{G}}, \ldots, (\boldsymbol{I}, \boldsymbol{M}, \delta)_m^{\mathcal{G}}\}$, the task of a classifier $\varphi : \boldsymbol{Y} \times \mathcal{G} \mapsto \widehat{\ell}$ is to produce $\widehat{\ell}$, where ideally $\widehat{\ell} = \ell$.

We consider six HEp-2 cell patterns [29] 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; nucloli 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

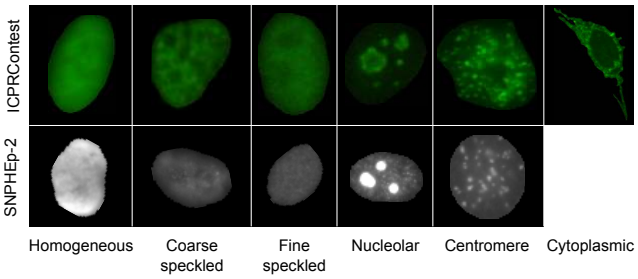

Figure 2. Sample images from ICPRContest dataset [8] and the proposed SNPHEp-2 dataset.

[1] It is assumed that the cell images have been extracted from specimen images via an approach such as background subtraction [21].

3. HEp-2 Cell Image Descriptor

The overall idea of the proposed HEp-2 cell image descriptor is shown in Fig. 3. Each cell is divided into small overlapping patches. The patches are then used to construct two histograms of visual words: *inner* and *outer* histograms, depending on whether the patches come from the the inside of the cell, or its edges, respectively. We first describe low-level patch-level features in Section 3.1, followed by presenting the dual-region structure in Section 3.2. In Section 3.3 we present several histogram encoding methods.

3.1. Patch-level Feature Extraction

Given a HEp-2 cell image $(\boldsymbol{I}, \boldsymbol{M}, \delta)$, both the FITC image $\boldsymbol{I}$ and mask image $\boldsymbol{M}$ are divided into small overlapping patches $\mathcal{P}_I = \{\boldsymbol{p}_{I,1}, \boldsymbol{p}_{I,2}, \ldots, \boldsymbol{p}_{I,n}\}$ and $\mathcal{P}_M = \{\boldsymbol{p}_{M,1}, \boldsymbol{p}_{M,2}, \ldots, \boldsymbol{p}_{M,n}\}$. The division is accomplished in the same manner of both images, resulting in each patch in the FITC image having a corresponding patch in the mask image. Let $\boldsymbol{f}$ be a patch-level feature extraction function $\boldsymbol{f} : \boldsymbol{p}_I \mapsto \boldsymbol{x}$, where $\boldsymbol{x} \in \mathbb{R}^d$. $\mathcal{P}_I$ now can be represented as $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$.

We selected two patch-level feature extraction techniques, based on the Scale Invariant Feature Transform (SIFT) and the Discrete Cosine Transform (DCT). The low-level SIFT descriptor is invariant to uniform scaling, orientation and partially invariant to affine distortion and illumination changes [15]. 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 [23, 30]. By using only the low frequency DCT coefficients (essentially a low-pass filter), each patch representation relatively robust to small alterations [23]. We follow the extraction procedures for SIFT and DCT as per [14] and [23], respectively.

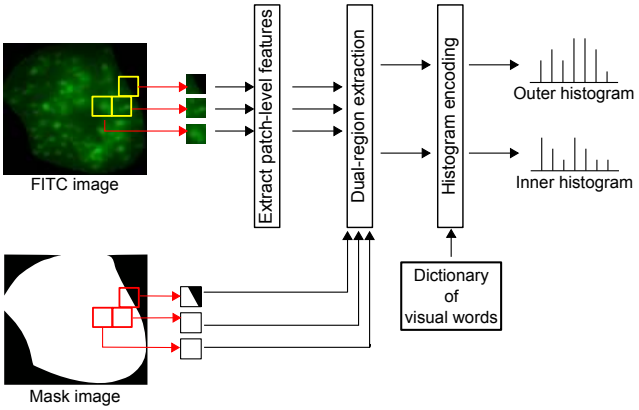

Figure 3. Conceptual diagram for the proposed HEp-2 cell image descriptor. Both the FITC image and its corresponding mask image are divided into small overlapping patches. Patch-level features are extracted from FITC patches. Each FITC patch is then classified into either outer or inner region by using information extracted from its corresponding mask patch. Finally, both inner and outer histograms are obtained by an encoder employing a learned dictionary of visual words.

3.2. Dual-Region Structure

We aim to model cell characteristics by building separate histograms for the inner region (which is often a uniform texture) and the outer region (which contains information related to cell edges and shape). To this end, each FITC patch is first classified as either belonging to the inner or outer region by inspecting its corresponding mask patch. Fig. 4 shows how the regions are imposed on a cell image.

Let $\mathcal{X} = \mathcal{X}_o \cup \mathcal{X}_i$, with $\mathcal{X}_o$ representing the set of outer patches, and $\mathcal{X}_i$ the set of inner patches. The classification of patch $\boldsymbol{p}_I$ into a region is done via:

$$\boldsymbol{p}_I \in \begin{cases} \mathcal{X}_o & \text{if } \tau_1 \leq \text{fg}(\boldsymbol{p}_M) < \tau_2 \\ \mathcal{X}_i & \text{if } \tau_2 \leq \text{fg}(\boldsymbol{p}_M) \end{cases} \quad (1)$$

where $\boldsymbol{p}_M$ is the corresponding mask patch; $\text{fg}(\boldsymbol{p}_M)$ computes the percentage of foreground pixels from mask patch $\boldsymbol{p}_M$; τ_1 is the minimum foreground pixel percentage of a patch belonging to the outer region; and τ_2 is the maximum foreground pixel percentage of a patch belonging to the outer region, as well as the minimum pixel percentage of a patch belonging to the inner region.

3.3. Generation of Histograms

Let $\mathcal{X}_r$ be the set of patch-level features for either the inner or outer region (ie., $\mathcal{X}_r = \mathcal{X}_i$ or $\mathcal{X}_r = \mathcal{X}_o$). For each patch-level feature $\boldsymbol{x}_j \in \mathcal{X}_r$, a local histogram $\boldsymbol{h}_j$ is obtained by an encoding method. The overall histogram representation for region r is then obtained via averaging [23, 30]:

$$\boldsymbol{H}^{[r]} = \frac{1}{|\mathcal{X}_r|} \sum_{j=1}^{|\mathcal{X}_r|} \boldsymbol{h}_j \quad (2)$$

where $|\mathcal{X}_r|$ is the number of elements in set $\mathcal{X}_r$. In this work we consider three popular histogram encoding methods: (1) vector quantisation; (2) soft assignment; (3) sparse coding. The methods are elucidated below.

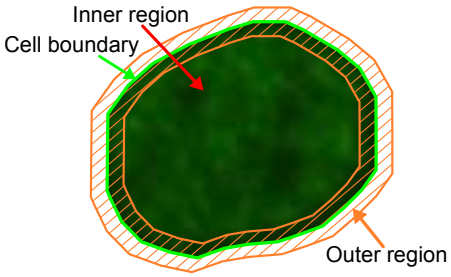

Figure 4. Conceptual diagram for the proposed dual-region structure. The green line is the cell boundary. The outer region is denoted by the striped patterns, and inner region is the area denoted by the red arrow.

3.3.1 Vector Quantisation (VQ)

Given set $\mathcal{D}$, a dictionary of visual words, the i-th dimension of local histogram $\boldsymbol{h}_j$ for patch $\boldsymbol{x}_j$ is computed via:

$$\boldsymbol{h}_{j,i} = \begin{cases} 1 & \text{if } i = \underset{k \in 1,\ldots,|\mathcal{D}|}{\arg\min}\ \phi(k) = \text{dist}(\boldsymbol{x}_j, \boldsymbol{d}_k) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $\text{dist}(\boldsymbol{x}_j, \boldsymbol{d}_k)$ is a distance function between $\boldsymbol{x}_j$ and $\boldsymbol{d}_k$, with $\boldsymbol{d}_k$ the k-th entry in the dictionary $\mathcal{D}$. The dictionary is obtained via the k-means algorithm [2] on training patches.

The VQ approach is considered as a hard assignment approach since each image patch is only assigned to one of the visual words. This hard assignment is sensitive to noise [28].

3.3.2 Soft Assignment (SA)

In comparison to the VQ approach above, a more robust approach is to apply a probabilistic method [23]. Here the visual dictionary $\mathcal{D}$ is a convex mixture of Gaussians. The i-th dimension of the local histogram for $\boldsymbol{x}_j$ is calculated as:

$$\boldsymbol{h}_{j,i} = \frac{w_i p_i(\boldsymbol{x}_j)}{\sum_{k=1}^{|\mathcal{D}|} w_k\ p_k(\boldsymbol{x}_j)} \quad (4)$$

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

$$p_i(\boldsymbol{x}) = \frac{\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \boldsymbol{C}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right]}{(2\pi)^{\frac{d}{2}} |\boldsymbol{C}_i|^{\frac{1}{2}}} \quad (5)$$

with w_i, $\boldsymbol{\mu}_i$ and $\boldsymbol{C}_i$ representing the weight, mean vector and covariance matrix of Gaussian i, respectively. The scalar d represents the dimensionality of $\boldsymbol{x}$. The dictionary $\mathcal{D}$ is obtained using the Expectation Maximisation algorithm [2] on training patches.

3.3.3 Sparse Coding (SC)

It has been observed that each local histogram produced via Eqn. (4) is sparse in nature (ie., most elements are close to zero) [30]. As such, it is possible to adapt dedicated sparse coding algorithms in order to represent each patch as a combination of dictionary atoms [5, 31].

A vector of weights $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_n]^T$ is computed for each $\boldsymbol{x}_j$ by solving a minimisation problem that selects a sparse set of dictionary atoms. As the theoretical optimality of the ℓ_1-norm minimisation solution is guaranteed [27], in this work we used:

$$\min \frac{1}{2}\|\boldsymbol{D}\boldsymbol{\alpha} - \boldsymbol{x}_j\|_2^2 + \lambda \sum\nolimits_k \|\alpha_k\|_1 \quad (6)$$

where $\|\cdot\|_p$ denotes the ℓ_p-norm and $\boldsymbol{D} \in \mathbb{R}^{d \times n}$ is a matrix of dictionary atoms. The dictionary $\boldsymbol{D}$ is trained by using the K-SVD algorithm [1].

As $\boldsymbol{\alpha}$ can have negative values due to the objective function in Eqn. (6), we construct each local histogram using the absolute value of each element in $\boldsymbol{\alpha}$ [30]:

$$\boldsymbol{h}_j = [\ |\alpha_1|,\ |\alpha_2|,\ \ldots,\ |\alpha_n|\ ] \quad (7)$$

Compared to both Eqns. (3) and (4), obtaining the histogram using sparse coding is considerably more computationally intensive, due to the need to solve a minimisation problem for each patch.

4. Classifiers

Let set $Q_X = \left\{\boldsymbol{H}_X^{[i]},\ \boldsymbol{H}_X^{[o]}\right\}$ represent the average inner and outer histograms for cell image X. Below we describe two classifiers that we have adapted to use both the inner and outer regions: (1) nearest neighbour (NN), and (2) nearest convex hull (NCH).

4.1. Nearest Neighbour (NN)

The NN classifier assigns the class of probe image A to be the class of the nearest training image B. For the purposes of this classifier, and inspired by [26], we define the distance between images A and B as:

$$\text{d}\left(Q_A, Q_B\right) = \gamma \left\|\boldsymbol{H}_A^{[i]} - \boldsymbol{H}_B^{[i]}\right\|_p + (1-\gamma)\left\|\boldsymbol{H}_A^{[o]} - \boldsymbol{H}_B^{[o]}\right\|_p \quad (8)$$

where $\gamma \in [0, 1]$ is a mixing parameter found during training.

4.2. Nearest Convex Hull (NCH)

In NCH, each training class is approximated with a simple convex model, or more specifically, the convex hull of the descriptors of the training images [17]. This reduces the sensitivity to within class variation, as "missing" samples can be approximated using the convex model [4].

Let $\lambda_C = \left\{\Omega_C^{[i]}, \Omega_C^{[o]}\right\}$ denote the model of class C, comprised of inner and outer components. In order to take into account both the inner and outer histograms, we define distance between image A and class C as:

$$\text{d}\left(Q_A, \lambda_C\right) = \gamma\, \text{d}_{\text{NCH}}\left(\boldsymbol{H}_A^{[i]}, \Omega_C^{[i]}\right) + (1-\gamma)\, \text{d}_{\text{NCH}}\left(\boldsymbol{H}_A^{[o]}, \Omega_C^{[o]}\right) \quad (9)$$

where $\gamma \in [0, 1]$ is a mixing parameter found during training, and $\text{d}_{NCH}(\boldsymbol{H}, \Omega_C)$ is the distance between histogram $\boldsymbol{H}$ and convex model Ω_C, defined as:

$$\text{d}_{NCH}\left(\boldsymbol{H}, \Omega_C\right) = \min \|\boldsymbol{H} - \boldsymbol{\omega}\|_1,\ \boldsymbol{\omega} \in \Omega_C \quad (10)$$

where Ω_C is a set of points generated by a linear combination of the training samples. Given a set of training histograms for class C, $\{\boldsymbol{H}_1,\ \boldsymbol{H}_2,\ \ldots,\ \boldsymbol{H}_m\}$, each member of Ω_C is defined as [17]:

$$\boldsymbol{\omega} = \sum\nolimits_{i=1}^{m} \beta_i \boldsymbol{H}_i,\ \text{ subject to } \sum\nolimits_{i=1}^{m} \beta_i = 1 \quad (11)$$

The above model implicitly treats any combination of histogram descriptors as a valid gallery example.

5. Experiments

In this section we first overview the two datasets used in the experiments. We then evaluate the six variants of the codebook-based descriptor, where each of 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). Finally we compare the best codebook-based variant against two recently proposed systems. The various systems were implemented with the aid of the Armadillo C++ library [22].

5.1. ICPR HEp-2 Contest Dataset

The ICPR HEp-2 Cell Classification Contest Dataset (ICPRContest) [8] contain 1,457 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.

As the official test set is not yet publicly available, we use only the official training set to create ten-fold validation sets. The available images are divided into training and testing sets with 14 specimens each. As such, in each set, each pattern class only has 1-2 specimen images, where a specimen image contains a set of cells having the same pattern (see Fig. 1 for a visual representation). In total there are 721 and 736 cell images extracted for training and testing, respectively. The validation sets were created by randomly selecting the images from the 721 cell images. Each fold contains 652 and 69 cell images for training and testing respectively.

Note that due to the abovementioned limitation of available images, each class can only have 1-2 specimen images. As such there is an assumed bias, as there is a high chance that cells extracted from the same specimen image exist in both the training and testing sets of each fold.

5.2. SNP HEp-2 Cell Dataset

The SNP HEp-2 Cell Dataset (SNPHEp-2) was obtained between January and February 2012 at Sullivan Nicolaides Pathology laboratory, Australia. The dataset[2] 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 20x/0.8 objective lenses and an LED illumination source. DAPI image channel was used to automatically extract the cell image masks.

There are 1,884 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 validation 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 images each). Examples are shown in Fig. 2.

By using the SNPHEp-2 dataset in addition to the ICPRContest dataset, we obtain the following benefits: (1) the specimens of both datasets were not captured by the same microscope configuration (eg., the microscope's objective lens magnitude for ICPRContest is 40x, while 20x for SNPHEp-2), allowing us to test the robustness of CAD systems to variations in image capture conditions; (2) there is no bias compared to the ICPRContest experiment setup, allowing for a more thorough evaluation.

5.3. Codebook-Based Descriptor Variants

In this section we evaluate the discriminative power of the codebook-based descriptor, with and without the dual-region structure. When the dual-region structure is not employed, each cell image is represented by one histogram constructed from both the inner and outer patches.

As there are three histogram encoding methods (ie., VQ, SA and SC) and two patch-level features (ie., SIFT and DCT), there are six variants of the codebook-based descriptor. For clarity, each variant is styled as: *[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 NN classifier was employed in this comparison, in order to reduce the total number of combinations. Based on preliminary experiments, we selected ℓ_1-norm distance in Eqn. (8) to measure the distance between two images. All other hyperparameters of each approach were found in the training set of each cross-validation set.

Table 2 presents the average Correct Classification Rate (CCR) for each descriptor variant on the ICPRContest and SNPEHEp-2 datasets, using both single- and dual-region configurations. We can observe that all variants have higher CCR on ICPRContest than on SNPEHEp-2, which is consistent with the bias in the ICPRContest dataset setup.

The DCT-SA variant is more discriminative and robust to various hardware configurations, as it consistently outperforms the other variants on both datasets. This high performance can be partly attributed to effect of soft-assignment, which can be more expressive than the other variants [28].

Generally, DCT has better performance than SIFT on most codebook-based variants on both datasets. The only

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

Table 2. Performance comparison of codebook-based descriptor variants on the ICPRContest and SNPHEp-2 datasets, using the NN classifier. The scores are shown as average correct classification rate (in percentage) along with their standard deviations. SR = single region; DR = dual region.

Descriptor Variant	ICPRContest		SNPHEp-2	
	SR	DR	SR	DR
DCT-SA	**93.8 ± 2.2**	**94.9 ± 2.1**	**74.7 ± 3.6**	**76.7 ± 2.5**
DCT-VQ	89.5 ± 3.4	90.6 ± 3.2	72.3 ± 3.2	73.6 ± 2.1
DCT-SC	81.7 ± 3.4	84.8 ± 3.5	59.9 ± 2.7	63.6 ± 1.8
SIFT-SA	80.1 ± 3.8	82.6 ± 3.5	56.6 ± 3.1	59.2 ± 2.5
SIFT-VQ	86.4 ± 4.0	86.8 ± 3.7	64.7 ± 2.3	64.9 ± 2.5
SIFT-SC	79.9 ± 3.3	86.1 ± 2.6	66.4 ± 2.9	67.9 ± 2.5

exception is on SNPHEp-2, where the SIFT-SC outperforms DCT-SC. This suggests that for this application, DCT is more suitable than SIFT for representing low-level patch features.

The results also show that imposing a spatial structure (ie., using the dual-region setup in contrast to the single-region setup) increases the performance of all of the variants, while generally reducing the standard deviation.

5.4. Classifier Variants

In this section we compare the performance of the NN and NCH classifiers on both datasets. We use the most discriminative descriptor found in Section 5.3, ie., DCT-SA. In addition, we have also evaluated the performance of two baseline descriptors: (i) raw image, where a raw cell image is vectorised, and (ii) rotation invariant Linear Binary Patterns (LBP) [18], using a configuration of 8 neighbours and 1 pixel radius. For the baseline descriptors, the cell images were used without any further processing (eg., no spatial structure was imposed).

The results, presented in Table 3, show that NCH generally outperforms NN regardless of the descriptor being used. An exception is LBP on SNPHEp-2, where NN performs slightly better. In most cases the NCH classifier also provides the most improvement on the more difficult SNPHEp-2 dataset. with its performance on the ICPRContest dataset close to the performance of NN. The results also show that the proposed DCT-SA approach (in both single- and dual-region configurations) considerably outperforms LBP, especially on the SNPHEp-2 dataset.

5.5. Comparative Evaluation of Systems

In this section we compare the best performing codebook-based system found in Section 5.4, ie., dual-region DCT-SA coupled with the NCH classifier, against two recently proposed systems in Cordelli et al. [6] and Strandmark et al. [25].

We implemented the best reported descriptor in [6], which is comprised 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. We denote this system as *Cordelli*.

Table 3. Performance comparison of the NN and NCH classifiers on the ICPRContest and SNPHEp-2 datasets. SR = single region; DR = dual region.

Descriptor	ICPRContest		SNPHEp-2	
	NN	NCH	NN	NCH
DCT-SA + DR	94.9 ± 2.1	**95.5 ± 2.2**	76.7 ± 2.5	**80.6 ± 2.1**
DCT-SA + SR	93.8 ± 2.2	94.3 ± 2.3	74.7 ± 3.6	78.5 ± 3.2
LBP	85.8 ± 2.6	86.4 ± 3.1	49.9 ± 3.8	47.6 ± 2.2
Raw Image	39.8 ± 3.2	57.7 ± 4.5	39.1 ± 3.3	43.1 ± 4.2

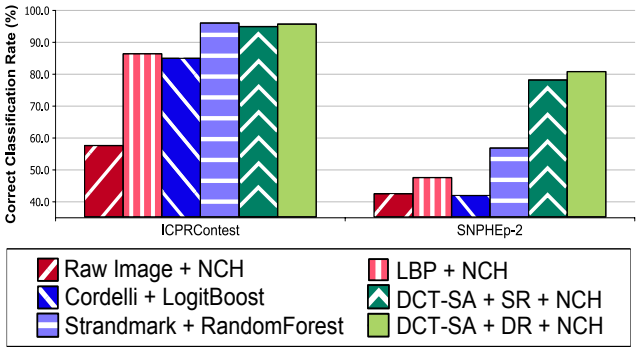

Figure 5. Performance comparison of various systems on the ICPRContest and SNPHEp-2 datasets. SR = single region; DR = dual region; NCH = Nearest Convex Hull Classifier.

We denote the system in [25] as *Strandmark*, and used the code provided by the authors. The system employs various image statistics features (eg., mean, standard deviation) and morphological features (eg., number of objects, area). The random forest classifier is used.

The results are presented in Fig. 5. Both the Cordelli and Strandmark systems have reasonable performance on the ICPRContest dataset. Strandmark has slightly better performance than the proposed DCT-SA system (96.1% vs 95.5%). However, both the Strandmark and Cordelli systems perform poorly on the more challenging SNPHEp-2 dataset, while the proposed system has considerably better performance. This indicates that the descriptors used by Cordelli and Strandmark systems are sensitive to hardware configuration variations. The poor performance of the Cordelli system on SNPHEp-2 can be partly explained from the observation that LBP also performs poorly on SNPHEp-2. As mentioned before, Cordelli uses the LBP channel to compute some of its features.

6. Main Findings

The Indirect Immunofluorescence method on Human Epithelial (HEp-2) cells is a hallmark method for identifying the presence of Anti-Nuclear Antibodies in clinical pathology tests. Despite its high sensitivity and the large range of antigens that can be detected, it has numerous shortcomings, such as being subjective as well as time and labour intensive. Computer Aided Diagnostic (CAD) systems have been recently developed to address these problems, which automatically classify a HEp-2 cell image into one of the known patterns (eg., 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.

In this paper we have proposed a cell classification system comprised of a dual-region codebook-based descriptor combined with the Nearest Convex Hull Classifier. The system splits a cell image into small patches, which are then grouped into sets representing the inner and edge regions of the cell. Each region is the described as a histogram of visual words. To our knowledge, this is the first time codebook-based descriptors are successfully applied and thoroughly studied in the domain of cell classification.

We evaluated numerous variants of the descriptor on two publicly available datasets: ICPR HEp-2 cell classification contest dataset and the new SNPHEp-2 dataset. We found that DCT patch-level features in conjunction with soft-assignment/probabilistic encoding of histograms leads to the highest discrimination performance. We also found that imposing the dual-region spatial structure increases discrimination performance of all codebook-based descriptor variants. Furthermore, the experiments show that the proposed system has consistent high performance and is more robust than two recent CAD systems presented in [6, 25].

We note that the proposed dual-region spatial structure used in this work is intuitive and lacks a theoretical explanation. Given the encouraging results, a more complete model of spatial structure could be developed to further increase performance.

Acknowledgements

This research was supported by Sullivan Nicolaides Pathology, Australia and NICTA. 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.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing*, 54(11):4311–4322, 2006.

[2] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] N. Bizzaro, R. Tozzoli, E. Tonutti, A. Piazza, F. Manoni, A. Ghirardello, D. Bassetti, D. Villalta, M. Pradella, and P. Rizzotti. Variability between methods to determine ANA, anti-dsDNA and anti-ENA autoantibodies: a collaborative study with the biomedical industry. *Journal of Immunological Methods*, 219(1-2):99–107, 1998.

[4] H. Cevikalp and B. Triggs. Face recognition based on image sets. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 2567–2573, 2010.

[5] A. Coates and A. Y. Ng. The importance of encoding versus training with sparse coding and vector quantization. In *Int. Conf. Machine Learning*, 2011.

[6] E. Cordelli and P. Soda. Color to grayscale staining pattern representation in IIF. In *International Symposium on Computer-Based Medical Systems*, pages 1–6, 2011.

[7] P. Elbischger, S. Geerts, K. Sander, G. Ziervogel-Lukas, and P. Sinah. Algorithmic framework for HEp-2 fluorescence pattern classification to aid auto-immune diseases diagnosis. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 562–565, 2009.

[8] P. Foggia, G. Percannella, P. Soda, and M. Vento. Early experiences in mitotic cells recognition on HEp-2 slides. In *International Symposium on Computer-Based Medical Systems*, pages 38–43, 2010.

[9] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2:147–171, 2009.

[10] R. Hiemann, T. Bttner, T. Krieger, D. Roggenbuck, U. Sack, and K. Conrad. Challenges of automated screening and differentiation of non-organ specific autoantibodies on HEp-2 cells. *Autoimmunity Reviews*, 9(1):17–22, 2009.

[11] T. Hsieh, Y. Huang, C. Chung, and Y. Huang. HEp-2 cell classification in indirect immunofluorescence images. In *Int. Conf. Information, Communications and Signal Processing*, pages 1–4, 2009.

[12] R. Khutlang, S. Krishnan, R. Dendere, A. Whitelaw, K. Veropoulos, G. Learmonth, and T. S. Douglas. Classification of mycobacterium tuberculosis in images of ZN-stained sputum smears. *IEEE Trans. Information Technology in Biomedicine*, 14(4):949–957, 2010.

[13] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2169–2178, 2006.

[14] C. Liu, J. Yuen, and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33(5):978–994, 2011.

[15] D. G. Lowe. Distinctive image features from Scale-Invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[16] P. L. Meroni and P. H. Schur. ANA screening: an old test with new recommendations. *Annals of the Rheumatic Diseases*, 69(8):1420 –1422, 2010.

[17] G. Nalbantov, P. Groenen, and J. Bioch. Nearest convex hull classification. Report EI 2006-50, Erasmus University Rotterdam, Econometric Institute, 2006.

[18] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[19] P. Perner, H. Perner, and B. Mller. Mining knowledge for HEp-2 cell image classification. *Artificial Intelligence in Medicine*, 26:161–173, 2002.

[20] B. Pham, S. Albarede, A. Guyard, E. Burg, and P. Maisonneuve. Impact of external quality assessment on antinuclear antibody detection performance. *Lupus*, 14(2):113–119, 2005.

[21] V. Reddy, C. Sanderson, and B. C. Lovell. Improved foreground detection via block-based classifier cascade with probabilistic decision integration. *IEEE Transactions on Circuits and Systems for Video Technology*, (in press).

[22] C. Sanderson. Armadillo: An open source C++ linear algebra library for fast prototyping and computationally intensive experiments. Technical report, NICTA, 2010.

[23] C. Sanderson and B. C. Lovell. Multi-region probabilistic histograms for robust and scalable identity inference. In *Lecture Notes in Computer Science (LNCS)*, volume 5558, pages 199–208, 2009.

[24] P. Soda and G. Iannello. Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis. *IEEE Trans. Information Technology in Biomedicine*, 13(3):322–329, 2009.

[25] P. Strandmark, J. Ulén, and F. Kahl. Hep-2 staining pattern classification. In *Int. Conf. Pattern Recognition*, 2012.

[26] T. Tommasi, F. Orabona, and B. Caputo. Discriminative cue integration for medical image annotation. *Pattern Recognition Letters*, 29(15):1996–2002, 2008.

[27] J. Tropp and S. Wright. Computational methods for sparse solution of linear inverse problems. *Proceedings of the IEEE*, 98(6):948 –958, 2010.

[28] J. van Gemert, C. Veenman, A. Smeulders, and J. Geusebroek. Visual word ambiguity. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(7):1271–1283, 2010.

[29] A. S. Wiik, M. Hier-Madsen, J. Forslid, P. Charles, and J. Meyrowitsch. Antinuclear antibodies: A contemporary nomenclature using HEp-2 cells. *Journal of Autoimmunity*, 35:276–290, 2010.

[30] Y. Wong, M. T. Harandi, C. Sanderson, and B. C. Lovell. On robust biometric identity verification via sparse encoding of faces: Holistic vs local approaches. In *IEEE International Joint Conference on Neural Networks*, pages 1762–1769, 2012.

[31] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 1794–1801, 2009.