HEp-2 Cells Classification Using Locally Aggregated Features Mapped in the Dissimilarity Space

Dimitrios Kastaniotis, *Student Member IEEE,* Ilias Theodorakopoulos, *Student Member IEEE*, George Economou and Spiros Fotopoulos

***Abstract*— Indirect Immunofluorescence (IIF) followed by manual evaluation of the acquired slides from specialized personnel is the preferred laboratory technique used for the detection of Antinucleolar Antibodies (ANAs) in patient serum. In this procedure, several limitations appear and thus several automatic techniques have been proposed for the task of ANA detection. In this paper we propose a system for automatic classification of HEp-2 staining patterns, inspired by a recently proposed method for aggregating local image (SIFT) features into a compact and fixed length representation. More specifically we present a novel framework in which aggregated features are mapped into feature vectors in the dissimilarity space where the dimensionality of the descriptors is "naturally reduced". The final descriptor is low dimensional, while evaluation on a recently published dataset yields state of the art results.**

I. Introduction

Several autoimmune diseases, such as dermatomyositis, autoimmune rheumatic diseases and primary biliary cirrhosis are connected with the presence of autoantibodies in patient serum. The preferred technique for detecting ANA's in serum is the indirect immunofluorescence (IIF) imaging, in which a specific fluorescence pattern occurs on a human epithelial cell line (HEp-2). The dominant identification procedure requires highly qualified physicians who manually identify these patterns by visually inspecting the slides acquired from a fluorescence microscope. This process on one hand lacks standardization and on the other hand is affected to the observer's limited classification ability. Furthermore, the presence of noise in acquired images and several other predicaments such as the photobleaching effect further hamper human ability. According to modern trends in other similar areas of medicine, automatic procedures aiming to provide a solution for performing these tests fast, reliably and economically have to be invoked. In the context of a Computer Aided Diagnosis (CAD) system, an automated method would work complementary to the physician's decision increasing the overall performance and keeping low error rates.

In this work, inspired by a method originally proposed for the task of large scale image retrieval [1], and the efficiency achieved by mapping features to the dissimilarity space, we propose a method for automatic classification of HEp-2 cells, based in staining patterns captured using Fluorescence Microscopy. More specifically, we focus on the automatic staining pattern classification of single-cell fluorescence images, which is the final stage of the IIF workflow. Our method starts by using a well known local image feature detector/descriptor [2] and an efficient method for aggregating information coming from several descriptors of an image into a fixed length feature vector. These feature vectors are then mapped into the dissimilarity space [3] using standard distance measures (here we used the L_1 norm). In the dissimilarity space the dimensionality of the features is "naturally reduced", while classification is performed using Sparse Based Classification techniques [4] [5].

The suggested framework is evaluated using a recently published HEp-2 Cells dataset under a relevant contest where the results obtained are very promising, reaching state of the art levels.

II. Related Work

Early methods targeting the task of HEp-2 cells classification used private datasets for evaluation of the results making performance comparisons extremely difficult. More precisely a method for HEp-2 cells segmentation and classification has been proposed by Perner et al. [6] where an automatic thresholding via Otsu's method was used for segmenting the individual cells and texture model estimation was obtained by various realizations of compact random sets. In another approach, Soda et al. [7] proposed an aggregation of binary classifiers operating on statistical and spectral textural features. Recently, a dataset suitable for evaluating several methods was made publicly available under a relevant contest [8]. Under this contest, several new methods were proposed. More specifically, Nosaka et al. [9] extracted textural features by utilizing an extension of the well known LBP descriptor named Co- occurent Adjacent LBP (CoALBP) .

Manuscript received July 30, 2013. This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

Dimitrios Kastaniotis, Ilias Theodorakopoulos, Spiros Fotopoulos and George Economou are with the University of Patras, Department of Physics, Electronics Laboratory, Patras, Greece (phone: +30 2610 997 463; fax: +30 2610 997 456; e-mail: {dkastaniotis, iltheodorako}@upatras.gr, (spiros, economou}@ physics.upatras.gr).

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

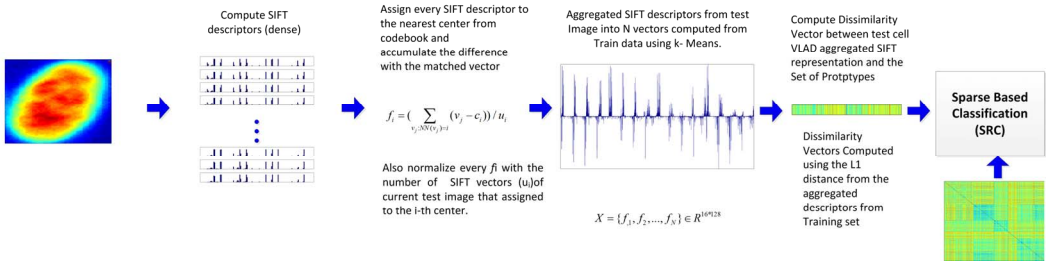

Fig. 1. Overview of the proposed method. Starting from the left, a HEp-2 cell is depicted and a total number of M dense SIFT descriptors is computed. Every descriptor is then encoded according to the VLAD framework using the N centers estimated from the training set. The aggregated SIFT descriptors are then depicted and a vector of dissimilarities is computed. At the last level, the dissimilarity vector is classified into one of the six categories using a set of prototype dissimilarity vectors.

This descriptor encodes information from co- occurrence of adjacent LBPs' thus dealing with the instability of luminance as well as the complex and diverse texture characterizing HEp-2 Cells. The final classification is performed using a linear SVM classifier.

XiangFei et al. [8], encoded textural and shape information using two histograms. In the first histogram images are encoded into a frequency histogram using a global texton dictionary while the second histogram encodes magnitude as well as direction of the gradient in order to accumulate information from shape features stemming from N blocks. Classification of Cells is performed using the x^2 distance using a k-NN classifier. Recently Kuan et. al [8] proposed a method that combines four textural descriptors, multiscale and rotation invariant uniform LBPs, statistical moments of Gabor wavelets and DCT on 4x4 non-overlapping blocks. By transforming the image to grayscale and binarizing the image using the Otsu method, a global appearance description consisting of ten statistical features was extracted. Also a combination of classifiers using adaBoost.M1 was used in order to improve classification performance.

III. Proposed Method

HEp-2 Cells are characterized by significant within class variability making the structure identification of individual cells extremely complex. In order to deal with this problem in this work we propose a framework based on the well known SIFT image descriptor [2] and Vector of Locally Aggregated Decsriptors (VLAD)- a recently proposed technique for image representation that aggregates information of several features into a compact and fixed length descriptor [1]. Originally, VLAD using the (SIFT) local descriptors computed from a given training set, first computes N centers forming a codebook which typical consists of N=16 vectors. Then given a set of M SIFT feature descriptors computed from a given image, this method encodes those M local features by assigning every descriptor to the closest vector of the codebook. All M SIFT vectors assigned to the each one of N centers are then aggregated into N vectors by summing the differences (using the L_1 distance) between the SIFT vector and its nearest codebook vector (the assigned center). In the next step, encodes these aggregated features into a code of B- bits by jointly optimizing a projection that reduces the dimension of the vector as well as a quantization that is used to index the resulting vectors.

In this work, we differentiate from VLAD at this point by selecting a more natural way to reduce the dimensionality of the aggregated descriptor. More specifically, we achieve this by mapping the aggregated features into the dissimilarity space using the L_1 distance as the dissimilarity function between aggregated vectors. The dimensionality of the features which is already reduced during the previous mapping is further reduced using Principal Component Analysis while preserving 90% of the total energy. An overview of the proposed method is given in figure 1.

a. Computation of SIFT features

As the cells exhibit similar sizes and images are acquired using a microscope with controlled magnification, the scale invariant character of the descriptor is not considered necessary. Thus in the context of modSIFT [10] which is a variant of SIFT descriptor, we skip the first stage of the SIFT algorithm in detecting key- points invariant to scaling and we directly compute features of a fixed predefined scale, on a dense grid which approximately includes all the pixels of a cell. Using 8-bins to represent gradient orientation, the final concatenated descriptor is 128-dimensional vector, each one computed in a single 4x4 block of a neighbor around each pixel.

Depending of the number of samples taken from every image, the number of SIFT features could vary significantly. In order to achieve an efficient low dimensional representation of the particular SIFT features in the next step we present a modified version of the VLAD descriptor [1].

b. Aggregation of SIFT features into a compact Descriptor

When the representation of an object is composed of an arbitrary sized set of feature descriptors, the most popular technique used to create a fixed-length vectorial representation is the Bag of Features (BoF) technique. This technique benefits from local descriptors, while the resulting histogram vectors can be compared using standard distances. In order to obtain a BoF representation initially a vocabulary is formed (usually using k-means) and then a sparse

frequency histogram of each local descriptor is computed. In the case of image retrieval this histogram tends to be extremely highly dimensional. The BoF framework imposes significant limitations due to the large amount of memory required to represent every image and the efficiency of the search due to the high dimensionality of the feature vector. Thus, the recently proposed VLAD method [1] which is appropriate for encoding information from local descriptors is adopted here. This method, inspired by both the Fisher Kernel and the BoF leads to a feature of a few bytes for every image by aggregating SIFT descriptors into a compact representation while lastly performs dimensionality reduction by jointly optimizing the representation, the dimensionality reduction of these vectors as well as the indexing algorithms.

In this work, we provide some early results focusing on the ability of the method to efficient aggregate information from several local descriptors into a dense feature vector. We differentiate though with [1] in the context of feature vector dimensionality reduction. More specifically, we achieve a low dimensional representation produced by mapping our samples into the Dissimilarity space using a natural distance metric (the L_1norm) between the aggregated features. More precisely, (as in the case of BoF) we take advantage of the fact that standard distances can again be used here in order to achieve a mapping in the dissimilarity space. Leaving other details for the next section we continue with the computation of VLAD [1] descriptor for every cell image with SIFT features *v* which can be expressed as follows:

$$f_i = \left(\sum_{v_j : NN(v_j)=i} (v_j - c_i) \right) / u_i \qquad (1)$$

In equation (1) *i* is an index to the *i*-th vector of the codebook $C = \{c_1, \ldots, c_N\}$ of total size N (as well as to the *i*-th aggregated feature vector), v_j is a feature vector (SIFT descriptor) from an arbitrary sized feature set and *j* is an index to the *j*-th feature vector (v_j) from a total number of *M* vectors such that $V=\{v_1, \ldots v_M\}$. The number of assigned SIFT vectors to every center *i* is u_i. Therefore, equation (1) aggregates every SIFT feature vector v_j to the vector f_i if v_j is closest to *i*-th center of the codebook c_i than any other vector *of C* (represented as $NN(v_j)=i$ in equation 1). A typical value for N is 16. The output of equation (1) is a concatenation of N aggregated and normalized with the number of assigned SIFT features (u_i), each one in the R^{128} feature space. Thus for a total of N=16 centers, the final aggregated descriptor is a R^{16x128} dimensional feature vector X:

$X = \{f_1, f_2, ..., f_N\}$, where $f_i \in R^{128}$ and $X \in R^{16x128}$.

For a more detailed description of the VLAD we refer the reader to [1].

IV. Dissimilarity Space Representation

Feature vectors computed in the previous section could be very high dimensional. For example consider that for a typical codebook of size N=16 we result into a 16x128=2048 dimensional feature vector.

In order to naturally reduce the dimensionality of the resulted feature descriptors (as well as to enhance discrimination) every sample is represented as a vector in the Dissimilarity Space as follows:

Given a set of *T* training cell images, let $X=\{f_1, \ldots, f_N\}$, where $f_i \in R^{128}$ be a vector of aggregated SIFT descriptors, where f_i represents the *i*-th feature vector of the aggregated SIFT features described in previous section. Given an appropriate dissimilarity function *d*, (which in our case is simply the L_1 norm), a mapping $D(\bullet, P) : X \to R^k$ is performed by selecting a set of *k* objects, namely prototypes where $k \le T$. As the training set consists of 721 training samples we used the aggregated feature vectors of all available training cell images as prototypes. Thus, the object size is equal to the training size $k = T$ which is significantly lower compared to the dimensional of the aggregated feature vector $X \in R^{16x128}$.

In the resulting space, called dissimilarity space, for every vector each dimension i describes the dissimilarity of the *i*-th sample with the k prototypes. The dimensionality of the final descriptor is further reduced to 75 dimensions using Principal Component Analysis, preserving 90% of the total energy. The aforementioned projection matrix is denoted as P_{pca} in equations (2) and (3).

More specifically, for every test cell represented as a dissimilarity feature vector Y projected on the orthonormal basis computed via PCA $Y_{proj} = YP_{pca}^T \in R^{75}$ and using a matrix D holding elements (codebook) computed from the training set, a vector containing sparse coefficients is computed. As shown in [5] this can be achieved by solving the following optimization problem:

$$x_{est} = \arg\min \|x\|_1, s.t., Y_{proj} = P_{pca}^T Dx \qquad (2)$$

In our case the dictionary consists from the training dissimilarity vectors themselves, thus *D* is $T \times k$ matrix (where *T* is the number of training samples and k is the number of selected prototypes) with dissimilarity vectors, and our goal according to equation 2 is to compute the vector of coefficients x_{est} with the minimum L_1 norm. Vector x_{est} will allow us to perform classification as we explain in the following section.

V. Classification

The coefficient vector might have small non-zero entries in positions corresponding to categories other than those associated with the true class. Thus instead of looking at the nonzero entries of the coefficient vector (e.g. x) we perform classification of every vector y_{est} based on the ability of coefficients in vector x_{est} to reconstruct y_{est} in terms of smallest residual as described in the following equation:

$$class(y_{est}) = \arg\min_{cl} \left\| y_{est} - P_{pca}^T Dx\delta_{cl}(x_{est}) \right\|_2 \qquad (3)$$

In equation (3), δ_{cl} is a vector who's non- zero entries correspond to the indices of x that are associated with class *cl*. Thus δ_{cl} selects for each class *cl* the appropriate vector in order to estimate the reconstruction error.
For more details regarding the classification in the dissimilarity space we refer the reader to [5].

VI. Dataset

In order to evaluate the proposed framework, we used a publicly available datasets that was proposed as a common benchmarking dataset for the HEp-2 Cell classification methods. The dataset consist of six categories namely Centromere, Fine Speckled, Coarse Speckled, Nucleolar and Homogeneous. According to the contest [8], the training set consists of 721 cells and the test set of 734 cells. For more details on the dataset the reader can refer to [8]. Evaluation here is performed in cell level, following the experimental protocol released during the contest.

TABLE I
Distribution of cells per category

Category	Number of available cells in train set
Centromere	388
Fine Speckled	225
Coarse Speckled	239
Nucleolar	257
Homogeneous	345
Cytoplasmatic	128

VII. Experimental Results

Following the evaluation protocol used in the [8] contest we achieved state of the art results as the following Table indicates. While our method overpasses the one proposed by Nosaka et.al [9] it is important to notice that both methods are very close to the results obtained from a human expert.

TABLE II
Experimental Results

Method	Overall Classification Rate %
Human Expert	73.33
Ours	**70.57**
Nosaka et. al.	68.7
Xiangfei et al.	65.8
Kuan et. al.	64.2

VIII. Conclusions and Future Work

In this work we proposed a novel framework for the task of HEp-2 cell classification. More specifically, in order to achieve a robust and discriminative description of each individual HEp-2 Cell we utilized only local image descriptors (SIFT features) that we aggregated into compact and fixed length descriptors in the context of VLAD representation**.** We differentiate with the authors–of [1] by selecting the Dissimilarity Representation [3] as a way to reduce the dimensionality of the final feature vector. Experimental results prove the effectiveness of the proposed framework, as the achieved overall classification rate is close to the human expert performance.
In the future we plan to provide a more detailed analysis by studying the benefits and the effects of the use of a mapping in the dissimilarity space as a "natural way" to achieve a low dimensional feature representation in comparison with the coding provided by the VLAD framework. Furthermore we intent to study the effect of forcing the use of class specific vocabularies in the selection of prototypes (during the first step of the VLAD algorithm) with respect to the discriminative performance of the VLAD aggregating framework.

Acknowledgment

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

References

[1] Jegou, H.; Douze, M.; Schmid, C.; Perez, P., "Aggregating local descriptors into a compact image representation," Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on , vol., no., pp.3304,3311, 13-18 June 2010 VLAD
[2] Lowe, D.G., Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vision, 2004. 60(2): p. 91-110.
[3] E. Pekalska, E.P.W. Duin, The dissimilarity representation for pattern recognition, World Scientific (2005) .
[4] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation,PAMI, 31, (2009), 210- 227.
[5] I. Theodorakopoulos, G. Economou and S. Fotopoulos, Classification of Dissimilarity Data via Sparse Representation, to appear on ICPRAM (2013).
[6] P. Perner, H. Perner, and B. Muller. "Mining knowledge for HEp-2 cellimage classification," Journal Artificial Intelligence in Medicine, vol. 26, pp. 161–73, September- October 2002.
[7] Soda, P. and G. Iannello, Aggregation of Classifiers for Staining Pattern Recognition in Antinuclear Autoantibodies Analysis. Information Technology in Biomedicine, IEEE Transactions on, 2009. 13(3): p. 322-329.
[8] Foggia, P., Percannella, G., Soda, P., Vento, M. (2013): Benchmarking HEp-2 Cells Classification Methods. In: Medical Imaging, IEEE Transactions on, PP (99), pp. 1-1, 2013,
[9] Nosaka, R., Y. Ohkawa, and K. Fukui, Feature extraction based on co-occurrence of adjacent local binary patterns, in Proceedings of the 5th Pacific Rim conference on Advances in Image and Video Technology - Volume Part II. 2012, Springer-Verlag: Gwangju, South Korea. p. 82-91.
[10] Tommasi, T., F. Orabona, and B. Caputo, Discriminative cue integration for medical image annotation. Pattern Recogn. Lett., 2008. 29(15): p. 1996-2002.