A NOVEL FEATURE DESCRIPTOR BASED ON MICROSCOPY IMAGE STATISTICS

Neslihan Bayramoğlu [§]   *Juho Kannala* [§]   *Malin Åkerfelt* [‡]
Mika Kaakinen [*]   *Lauri Eklund* [*]   *Matthias Nees* [‡]   *Janne Heikkilä* [§]

[§] Center for Machine Vision Research, University of Oulu, Finland
[‡] Centre for Biotechnology, University of Turku, Finland
[*] Faculty of Biochemistry and Molecular Medicine, University of Oulu, Finland

ABSTRACT

In this paper, we propose a novel feature description algorithm based on image statistics. The pipeline first performs independent component analysis on training image patches to obtain basis vectors (filters) for a lower dimensional representation. Then for a given image, a set of filter responses at each pixel is computed. Finally, a histogram representation, which considers the signs and magnitudes of the responses as well as the number of filters, is applied on local image patches. We propose to apply this idea to a microscopy image pixel identification system based on a learning framework. Experimental results show that the proposed algorithm performs better than the state-of-the-art descriptors in biomedical images of different microscopy modalities.

Index Terms— local image descriptor, pixel labeling, cell detection, phase contrast imaging, electron microscopy, mitochondria, tumor, cell co-culture.

1. INTRODUCTION

In order to obtain quantitative cell analysis from microscopy data, there is need for computational methods that can extract measurements of complex cellular dynamics automatically. In the literature, there has been a great number of work for automatic analysis of microscopy images. However, the majority of papers in medical and biomedical image analysis describe an algorithm or solution that addresses a particular task. The diversity in the type of the microscopy, molecular labelling, resolution, quality of images, cell appearance and cell dynamics in image data is high. Therefore, usually, it is not applicable to utilize existing methods to address those challenges and needs. The challenge is to develop a generic algorithm that has a wide application area in automated microscopy image analysis.

In this study, we propose a novel local image descriptor that can be utilized in various applications in (bio)medical image analysis research. Our proposed feature descriptor is based on statistical models of images. In this work, performance of the descriptor is tested in a microscopy image pixel labeling framework. Pixel level identification scheme can be employed as a generic detection method and as a priori for subsequent segmentations [1] of different cell lines and microscopy modalities. We employ our feature descriptor for detection of tumour cell spheroids in phase contrast imaging of cell co-cultures and for detection of mitochondria in electron microscopy images. Our method works under heavy occlusions and clutter and therefore suitable for most of the biomedical images. Experimental results demonstrate significant improvements over strong baseline methods.

2. METHOD

In traditional computer vision problems, local image descriptors are dominant because of their proven successes. Various descriptors have been proposed in the literature [3]. The most widely used local region descriptor is Scale Invariant Feature Transform (SIFT) [2] and it has become a baseline method for comparison. Local binary pattern (LBP) [4] and local phase quantization (LPQ) [5] are notable feature descriptors mainly used for face and texture recognition. Local feature descriptors have shown to be very successful also in (bio)medical image analysis research such as live cell tracking [6], cell/object detection [7, 8, 9], mitosis event segmentation [10], cell classification [11], and image alignment and registration [12].

Our method is inspired by SIFT and a recent binary descriptor called Binarized Statistical Image Features (BSIF) [13]. Our descriptor computation has a similar step with the gradient orientation binning in SIFT. In SIFT, each pixel is assigned to the orientation bin, weighted by the gradient magnitude. However, in contrast to SIFT, we utilize a set of responses of linear filters for every pixel. On the other hand, BSIF method is based on binary code string representation of pixels. The code is computed by binarizing the response of linear filters that are learnt from independent component analysis of natural image patches. Similarly, we utilize a set of linear filters; however, our local patch descriptor is computed by accumulating patch filter response magnitudes within a histogram based representation. The proposed descriptor, which is called Natural Image Statistics based Feature (NISF), consists of two main steps: learning filters from image statistics
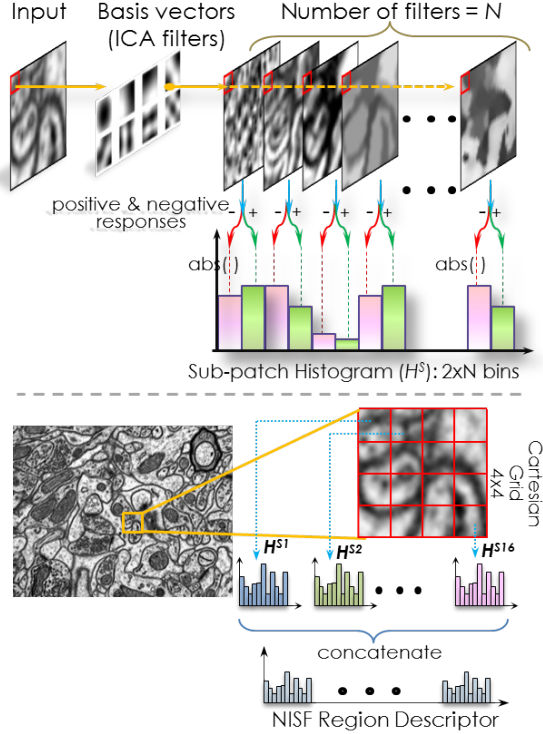
Fig. 1: NISF Descriptor. *(Top)* Illustration of histogram binning. *(Bottom)* Spatial encoding of a local image patch.

and filter response binning.

2.1. Microscopy Image Statistics

We begin our descriptor computation pipeline with filter learning. We learn filters by applying Independent Component Analysis (ICA) [14] to image patches from a set of images. In ICA, it is assumed that the input signal is to be produced as a linear superposition of independent signals. Denote by $\mathbf{x} = [x_1, x_2, \ldots, x_{k \times k}]^T$ a vectorised image patch of size $k \times k$, then the model of the ICA is given by:

$$\mathbf{x} = \sum_{i=1}^{N} s_i \mathbf{a}_i = \mathbf{As} \qquad (1)$$

where N is the number of filters, s_i's are the independent components and basis images are the column vectors of $\mathbf{A}$. The aim of ICA is to find the statistically independent components s_i by estimating the inverse matrix $\mathbf{W}$ of matrix $\mathbf{A}$: $\mathbf{s=Wx}$ where the row vectors in $\mathbf{W}$ are the ICA filters that can be used to analyse images. But first, data centering and whitening by principal component analysis is applied. The dimensionality is decreased at the same time. ICA encoding provides a new basis of representation by reducing the information redundancy of input images by projecting the data on statistically independent signals. ICA learns filters from unlabelled image patches and the number of basis images can

be controlled in addition to the size of the image patches (i.e. $N \leq (k \times k)$). For a detailed algorithm explanation we refer reader to [14].

2.2. NISF Descriptor

Our natural image statistics based (NISF) descriptor utilizes responses of filters that are learnt from microscopy images in the previous step. First, filter responses, $\mathbf{y} = \{y_1, \ldots, y_i, \ldots, y_N\}$, are computed at each pixel by convolving N learnt filters with the input image. Then, we compute histogram of image patches based on their filter response magnitudes and their signs. The descriptor $\boldsymbol{f}_{p(x,y)}$ of a pixel $\mathbf{p}$ located at (x, y) is evaluated within a rectangular grid centered at pixel $\mathbf{p}$. Similar to SIFT, the supporting region is then divided into 4×4 sub-patches of sizes $m \times m$ (Figure 1). Denote filter responses within a sub region by $\mathbf{y}^t$ where t corresponds to pixel index. Then the histogram H^S for a sub-patch is computed by summing absolute values of filter responses. Positive and negative responses are accumulated in different bins. Therefore, the bin index j is determined by the filter index i and the sign of the filter response:

$$H^S_{bin_j} = \sum_{t=1}^{m \times m} abs(y_i^t), \quad j = \begin{cases} 2 \times i & y_i^t \geq 0 \\ 2 \times i - 1 & y_i^t < 0 \end{cases}$$

$$i = 1, 2, \ldots, N \qquad (2)$$

which results $(2 \times N)$ length sub-patch histograms. Finally, NISF descriptor is built by concatenating sub-patch histograms into a compact feature vector of length $(4 \times 4 \times 2 \times N)$: $\boldsymbol{f}_{p(x,y)} = \{H^{S1} H^{S2} \ldots H^{S16}\}$.

3. SOFT PIXEL LABELLING

In this study, we propose a learning based framework for soft pixel labelling that employs our NISF descriptor (Figure 2). Local image descriptors can be utilized to identify pixel data; for example to detect particular cells/objects in microscopy images. We employ it for detection of tumour cell spheroids in phase contrast imaging of cell co-cultures (Figure 5) and for detection of mitochondria in electron microscopy (EM) images within a learning based framework. We start our pipeline by training a *"random forest"* (RF) classifier on pixel level features [1]. Similarly, pixel level decisions are then made during test time. We tested our method on six different public datasets. Table 1 gives detailed information about databases used in this study.

4. EXPERIMENTS

In our initial experiments, we used 8 filters learnt from a set of EM images provided by the authors of [15]. Filters are learnt using 50000 image patches of sizes 7×7 sampled from

Fig. 2: Illustration of pixel classification framework: It depends on a random decision forest framework to provide a pixel wise probabilistic classification.

Table 1: Datasets used in RF classification

	Dataset	Number of Images			
		Train	Test	Total	Size (*: Scaled)
1	EPFL [15]	17	165	182	$410 \times 307*$
2	TrackEM2 [17]	6	24	30	$358 \times 358*$
3	UCSD [1]	9	42	51	$350 \times 350*$
4	Turku-Col [8]	18	167	185	400×300
5	Turku-Mat [8]	6	21	27	400×300
6	Turku-Mtg+Col [8]	6	21	27	400×300

20 images from the training set of EPFL dataset [15]. Data used during the RF classifier training step is shown in Table 1. Note that a subset of EPFL training images is utilized in this work. A RF classifier is trained on features that are extracted by using our proposed descriptor. Equal number of samples from background and foreground are selected from the training set. During test time, pixel level features are extracted as it is done in the training stage and pixels are assigned to probabilistic outputs (soft decisions) based on the average decision generated by the random trees. Binary decisions for pixel class IDs can be made by simply thresholding the probabilistic outputs. Precision-recall (PR) curves are obtained by varying the threshold from 0 to 1.

We compare our results with the state-of-the-art descriptors including SIFT [2], BSIF [13], LBP [4], and SURF [16]. Similar to NISF, a 4×4 patch is employed in constructing pixel based descriptors of all the features. Uniform LBP patterns are utilized in order to reduce the high dimensional features. BSIF descriptors are computed with the same set of filters employed in NISF (i.e. 8 filters learned from 7×7 patches of EPFL data). For RF classification, for SIFT and SURF feature extraction, *OpenCV* implementation is used. We extract features from image patches of sizes 41×41 at one scale. For training RF classifiers, 20 trees are employed with a maximum depth of 15.

The comparative results are displayed in Figure 3. The pixel classification performance of the proposed method achieves the best performance in each dataset. In Figure 5, visualization of soft pixel labeling of sample images from
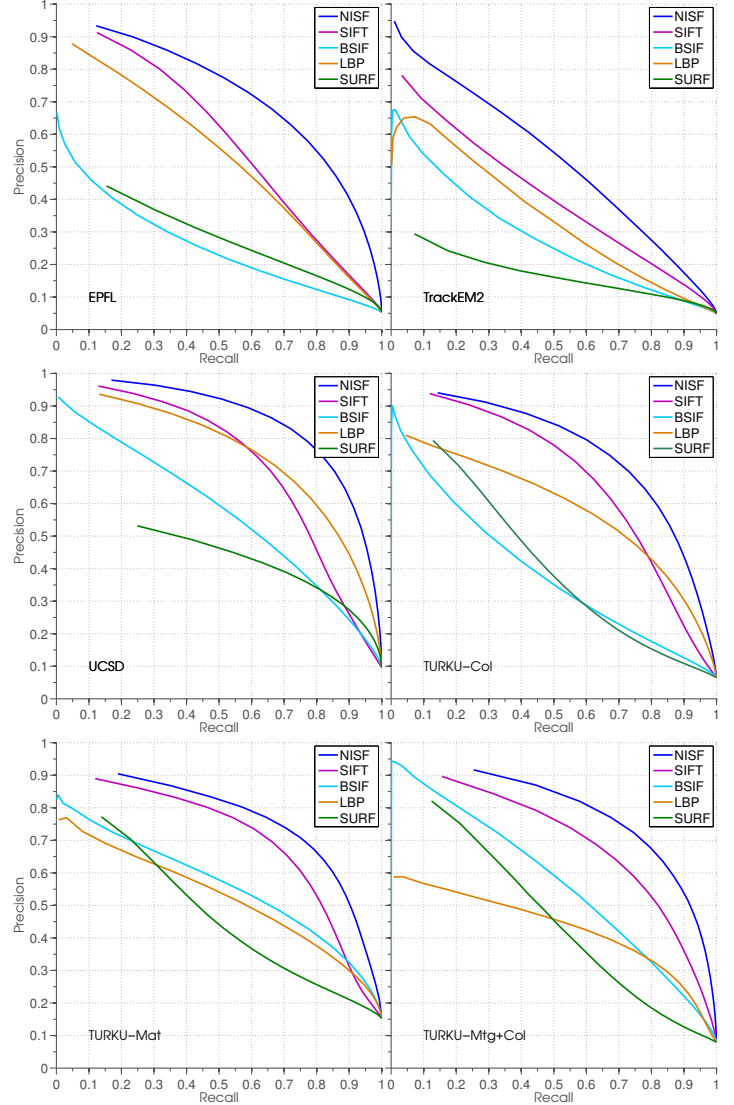
---

[1] ccdb.ucsd.edu, project ID: P2080, segmentation mask from http://cytoseg.googlecode.com

Fig. 3: Performance comparison of NISF, SIFT, BSIF, LBP and SURF descriptors on the datasets given in Table 1.

EPFL and Turku-Col using our approach and SIFT, which performs second best among all others, are provided. Despite it's simplicity, BSIF demonstrate a weak performance in this context. This could be due to its high dimensional and sparse structure which makes the descriptor sensitive to small data variations. We have also tested BSIF without dividing the image patch into grids and using a single global histogram for the whole patch but it did not improve the performance.

The selection of ICA based filters in NISF experiments are not random. In order to observe the effects of filter parameters we tested various filters generated from different sized image patches. The effect of increasing sizes of image patches while keeping the number of filters fixed is shown in Figure 4a for Turku-Col dataset. The eight filters NISF descriptor give good results with 7×7 filters. If the size of filters is kept
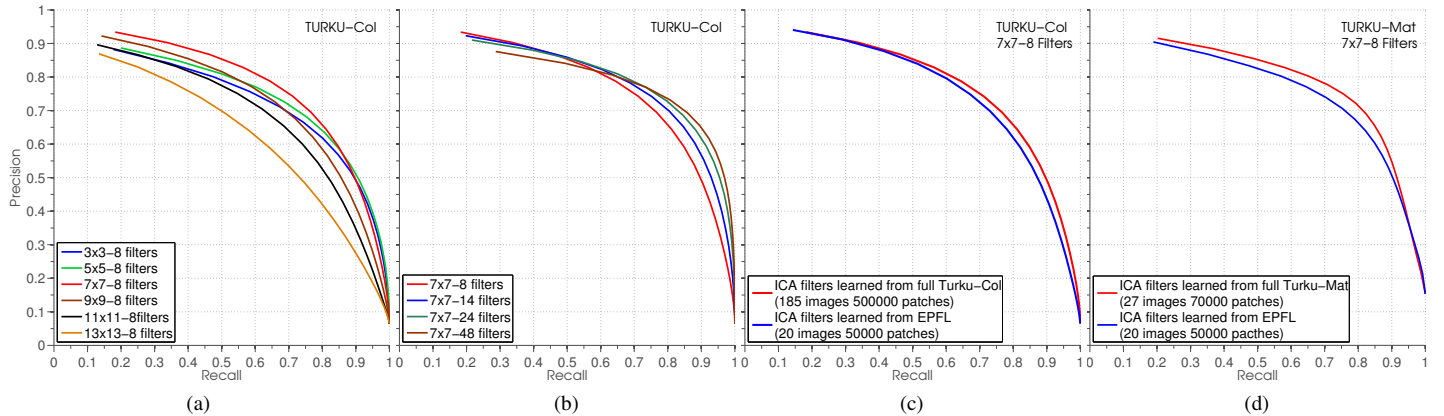
Fig. 4: Effects of (a) filter sizes, (b) number of filters, and (c-d) data on filter learning
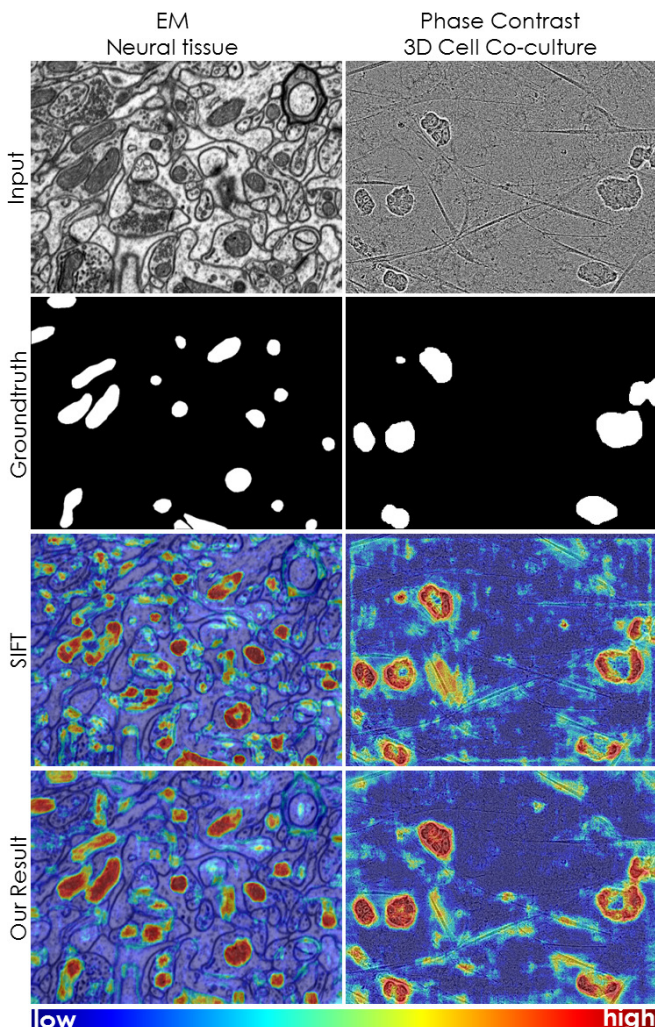

Fig. 5: *(From top to bottom)* Example images from datasets EPFL and Turku-Col. Segmentation mask of mitochondria in EPFL and tumour cell spheroids in Turku-Col data. Probability maps with no class labels assigned for SIFT and NISF respectively. Color map. (These results are obtained with 1000 RF trees.)

fixed at 7×7 and the number of filters is increased then the performance of NISF improves in the high recall region (Figure 4b). However, descriptor length increases as the number of filters increases. Therefore, we used 8 filters in all other experiments.

In Figure 3 experiments, a fixed set of filters learned only from EPFL EM images are used. We further investigate the effects of data used in filter learning stage. We tested NISF performance on phase contrast data by employing filters learned also from phase contrast data. The average PR curves of NISF descriptor employing filters learnt from EM and phase contrast images with fixed filter size of (7×7) and fixed filter number (8) are shown in Figure 4 (c-d) for Turku-Col and Turku-Mat. Learning filters from the same kind of test data performs better than learning filters from images of different content. ICA filter learning is a fully automatic step and does not require data annotation, therefore, all the available image data can be utilized in this stage.

In NISF, filter learning can be done off-line and only once. On the other hand, the number of convolutions performed during NISF construction and the increased vector length due to positive and negative bin partitioning increases the complexity of NISF descriptor. Although NISF is computationally more complex than SIFT, it performs better in all the experiments and the performance gain reaches up to 35% for example for UCSD dataset (Figure 3).

5. CONCLUSION

We introduce a new feature descriptor based on image statistics. Such a local image descriptor can be used in various applications in (bio)medical image analysis research from segmentation to tracking. The proposed method is tested in a microscopy image pixel classification application with several datasets. Our method NISF works under heavy occlusions and background clutter and performs better than the state-of-the-art baseline methods. Moreover, this work provides encouraging results for employing NISF in other domains.

6. REFERENCES

[1] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in *CVPR 2008*. IEEE.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 2004.

[3] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2005.

[4] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *PAMI*, 2002.

[5] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Image and signal processing*. Springer, 2008.

[6] R. M. Jiang, D. Crookes, N. Luo, and M. W. Davidson, "Live-cell tracking using sift features in dic microscopic videos," *Biomedical Engineering, IEEE Transactions on*, 2010.

[7] F. Mualla et al., "Automatic cell detection in bright-field microscope images using sift, random forests, and hierarchical clustering," *Medical Imaging, IEEE Transactions on*, 2013.

[8] N. Bayramoglu, M. Kaakinen, L. Eklund, M. Akerfelt, M. Nees, J. Kannala, and J. Heikkilä, "Detection of tumor cell spheroids from co-cultures using phase contrast images and machine learning approach," in *ICPR 2014*. IEEE.

[9] A. Farag, A. Ali, J. Graham, A. Farag, S. Elshazly, and R. Falk, "Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose ct scans of the chest," in *ISBI 2011*. IEEE.

[10] An-An Liu, Kang Li, and T. Kanade, "A semi-markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations," *Medical Imaging, IEEE Transactions on*, 2012.

[11] R. Nosaka and K. Fukui, "Hep-2 cell classification using rotation invariant co-occurrence among local binary patterns," *Pattern Recognition*, 2014.

[12] M. Toews and W. M. Wells III, "Efficient and robust model-to-image alignment using 3d scale-invariant features," *Medical image analysis*, 2013.

[13] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features," in *ICPR 2012*. IEEE.

[14] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, 2000.

[15] A. Lucchi, Y. Li, and P. Fua, "Learning for structured prediction using approximate subgradient descent with working sets," in *CVPR 2013*. IEEE.

[16] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in *ECCV 2006*.

[17] A. Cardona et al., "An integrated micro-and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy," *PLoS biology*, 2010.