MACHINE VISION FOR CLASSIFYING BIOLOGICAL AND BIOMEDICAL IMAGES

N.V. Orlov, D.M. Eckley, L.Shamir, and I.G. Goldberg
National Institute of Aging/National Institute of Health
251 Bayview Blvd., Suite 100, Baltimore, MD 21224, USA
norlov, dme, shamirli, igg { @nih.gov }

ABSTRACT
We present a method for automatic classification of biomedical images. The method exploits a machine vision approach to analyze image content based on a global set of image descriptors combined with image transforms. High dimensional feature spaces were used, and Fisher ranking was applied for weighting and truncating dimensions of the original feature space. Weighted nearest distance method was used as a basic classifier. The approach was applied to a variety of imaging problems including automatic classification of lymphoma malignancies and data from high-content screening experiments. The large degree of applicability of the proposed global feature set allows a search for patterns in a wide range of image types, across different domains. Importantly, we demonstrate certain capabilities of the machine vision approach in the biomedical domain.

KEY WORDS
Global features, image transforms, and biomedical images

1. Introduction

Automatic image classification is becoming a practical research tool in many traditional and emerging fields in contemporary bioinformatics. Over the last few decades image analysis grew from a subjective measure into a quantitative discipline; morphological patterns observed in microscopic images can be processed with computational algorithms measuring image content with a numerical value [1, 2]. Pattern recognition and automatic image classification can be used as practical tools in multiple image-driven applications.

In medical imaging [1, 7], computer vision methods [1-3] are emerging as a new tool bridging a gap between cancer diagnostics and pattern analysis. Existing research on cancer classification with machine vision methods focuses on individual cells, requiring segmentation of the pre-selected ROIs. The reported comparisons between computational approach and pathology expert evaluations demonstrate favorable advantage of machine vision methods [11-14].

In biology high-content screening (HCS) experiments with thousands of genes (phenotypes) result in high volumes of images reflecting the outcomes of the experiments. Automatic annotation of the output of HCS experiments is an important application of machine-vision approaches in biology and other fields.

Most techniques for automatic image classification, clustering and retrieval employ certain domain knowledge for constructing image descriptors. Such domain-specific feature sets may include low-level image descriptors, such as color [8], texture [9], histograms [10], and more.

Use of domain-specific features applies significant constraints on the range of applications – those image problems that do not meet domain criteria would not be a good fit for the constructed features. Biomedical applications remain an emerging area that presents many diverse problems and image modalities. Therefore, an approach that would work effectively for broader ranges

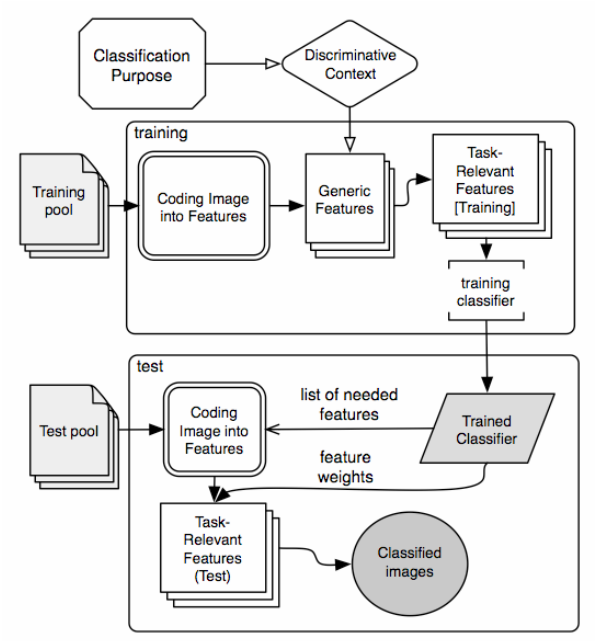

Figure 1. General classification scheme with feature extraction.

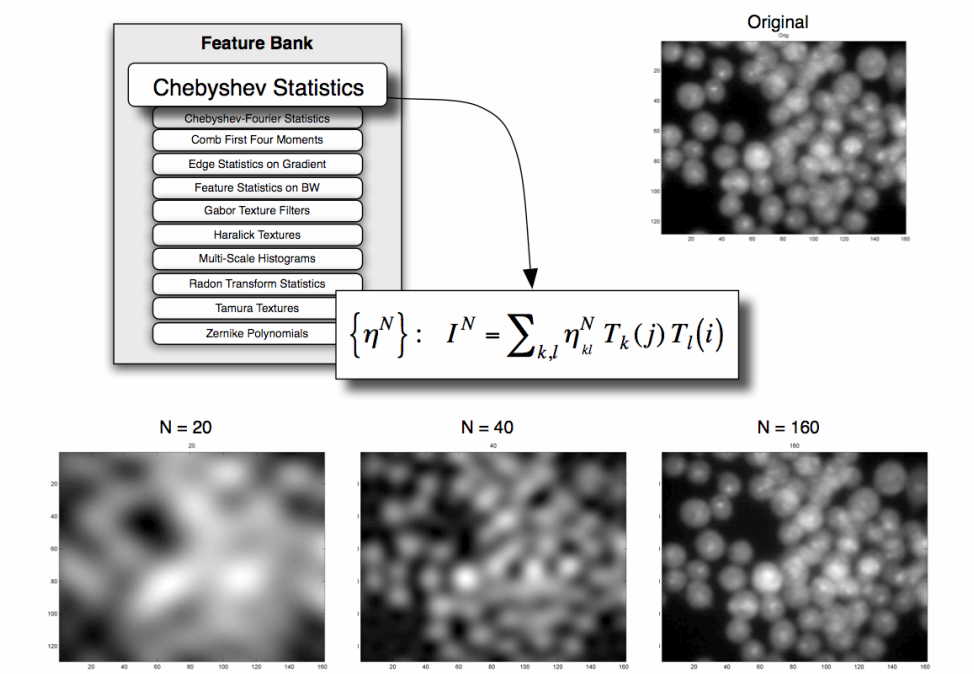

Figure 2. Chebyshev features are based on Chebushev expansion coefficients. Example demonstrates how three different level approximants of the original RNAi screen image report on different aspects of image content.

of applications remains an important task in the biological and biomedical domains. Such an approach is suggested in this report.

2. Feature Set for Biomedical Images

In image classification images are commonly represented through their quantitative content. An algorithm or set of algorithms is used to transform pixels into image features $I \xrightarrow{\Psi} \psi$, where I and ψ represent pixel and feature spaces.

2.1 Domain-Specific and Global Features

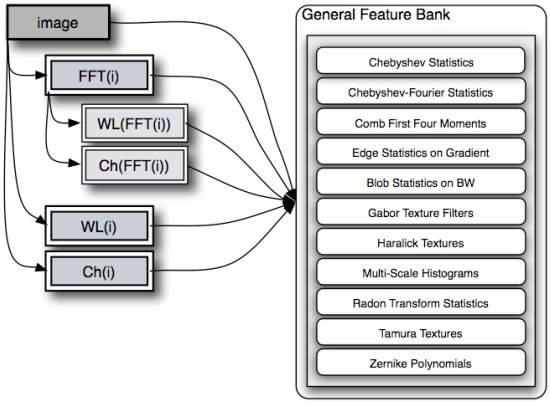

Figure 3. Image transforms and feature bank.

The choice of algorithms for coding images into features is usually predetermined [3] by the domain, scope of the problem, etc. There is often no rationale provided for selecting domain-specific predefined algorithms for image description [3, 4].

The image diversity found in biomedical applications makes it impossible to pre-select all relevant types of image descriptors. Descriptors of a general nature require no domain knowledge and pose no ambiguity. We chose global features having a broad range of applicability, and then selected those having the most discriminative power for the classification task given. No bias was applied when selecting the most relevant features for a particular problem.

Fig. 1 demonstrates the concept of using general-scope descriptors in image classification.

2.2 Global Features Used

Fig. 2 illustrates the use of Chebyshev polynomials for describing image content. Image approximation by Chebyshev polynomials has the form

$$I_{ij}^N = \sum_{k,l=0}^{N} \eta_{kl}^N \; T_k(x_j) \, T_l(y_i) \, .$$

Chebyshev polynomials $T_n(x) = \cos(n \operatorname{acos}(x))$ deliver approximation of the desired order, because the coefficients satisfy the condition

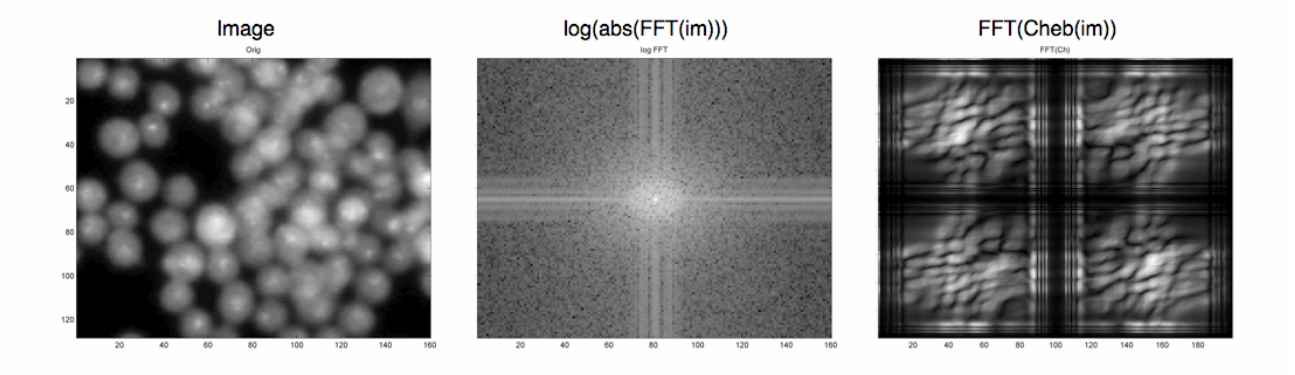

Figure 4. The image and its transforms exhibit absolutely different patterns, which makes efficient using the same filter bank on image and its transforms. Shown: image, its Fourier transform (absolute values), and a compound Fourier transform taken from Chebyshev transform.

$$\left\{\eta_{kl}^{N}\right\} = \arg\min_{\eta} \left\| I_{ij}^{N} - I_{ij} \right\|_{2},$$

and therefore are characteristics of the image. Fig. 3 demonstrates how different orders of approximation report on different elements of image content.

Our general feature bank contains three sections. The polynomial section calculates statistics on Chebyshev, Chebyshev-Fourier, as well as Zernike polynomials. Another part of the collection computes textures, including Tamura family (coarseness, directionality and contrast), Haralick set of statistics on co-occurance matrices, and Gabor filters to detect directionality for a certain spectral range. Radon transform histograms, Multi-scale histograms, First Four Comb moments, Edge statistics on image gradient, and blob statistics on thresholded grayscale images all belong to the third section of the feature bank. The feature bank is described in more detail in [5, 6].

3. Image Transforms and Their Meaning for Image Description

The general feature bank takes image pixels and converts them into a feature vector, as shown in Fig. 3. One aspect of our approach is that we use transforms to code images into feature space. Transforms remap image pixels to another set of pixels that is characteristic of the image, yet has a different image pattern (see FFT, Fig. 4). If several transforms are used, and k is a sequential index of the transform, we denote this mapping as $I \xrightarrow{F_k} f_k$. For compound transforms the mappings are $I \xrightarrow{F_k[F_j]} f_{k|j}$.

Fig. 4 demonstrates that an image and its transforms can exhibit different patterns (two transforms are shown: $\log\left(\left|FFT[I]\right|\right)$ and $\left|FFT\left\{\text{Cheb}[I]\right\}\right|$). Transforms allow a more effective use of the feature bank by extracting more patterns related to the same image. For example, an image might have no textural features (detected by the Tamura set), while its Fourier transform may exhibit them clearly. In this example, the benefit of Tamura features is revealed by the Fourier transform.

Raw pixels and pixels of image transforms (as well as the compound transforms shown in Fig. 3) are used as inputs to the feature bank. Wavelets (symlets, level-1 details), Chebyshev, Fourier and compound transforms were used to generate the final feature vector (1025 numerical values per image).

4. Pattern Classification

The entire set of features is computed, but only effective ones are used for the pattern search. The computed features undergo ranking [6], which reports discriminative power for different features, and identifies the weakest features that are likely to represent noise. The method for ranking utilizes the Fisher linear discriminant which assigns weights to features rewarding large between-class variation of data and penalizing data variation within the class.

The weighted neighbor distances method [6] was used in this study as a basic classifier. The classifier in this study uses the 660 highest-weighted features to construct a "weighted feature space", with each image represented by a point. Euclidian distances between a test image and all of the training images in the different classes are then computed, averaging the distances to each class. The test image is then assigned to the class with the lowest average distance to its training images.

5. Classification Results and Discussion

We applied the proposed method to several imaging problems, including automatic cancer categorization and

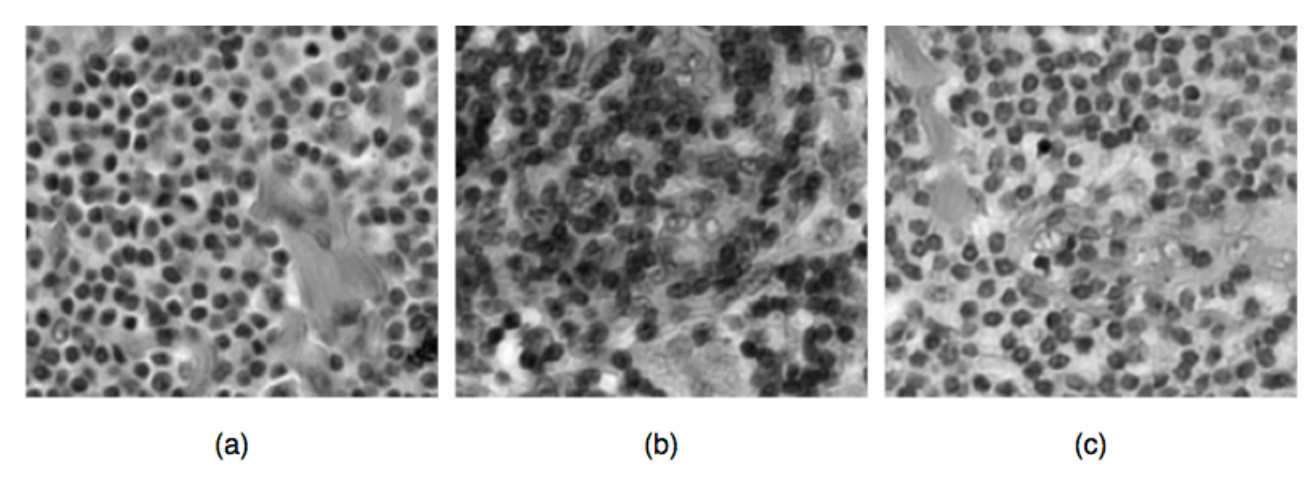

Figure 5. Representative tiles for three lymphoma cases in the grayscale. Panel (a) corresponds to CLL, (b) is FL, and (c) is MCL.

automatic phenotype classification resulting from HCS experiments.

Lymphoma is a cancer affecting lymph nodes [7]. The *Lymphoma* set is a collection of images taken from slides with hematoxylin and eosin stained biopsy samples. Three types of malignancies are present in the set: Chronic Lymphotic Leukimia (CLL), Follicular Lymphoma (FL) cells, and Mantle Cell Lymphoma (MCL). Ten slides per malignancy type were used in the analysis. Images were taken with an RGB camera on a Zeiss microscope using a 20x objective. The microscope was calibrated with a Zeiss calibration slide. AxioCam MRc5 camera was the imaging device; we used the full dynamic range of the camera (twelve bits per pixel) and stored it as double precision (sixteen bits) without distortions or reductions. No gamma-correction was used, and each channel was in the same dynamic range.

We tiled the images down to a manageable size for our feature bank (each image was split to 30 tiles, each was 208x231 pixels). Representative tiles for three malignancies are shown in Fig. 5. The feature bank works with one-channel data, therefore analyzing the RGB set results in a larger feature space. The separate RGB channels were also transformed into single grayscale images by the NTSC transform.

The trained classifier assigns a class for each tile; the average of all 30 tiles is the score for the whole image. We made four random splits of the image pool into training/test partitions and define the classifier accuracy as the per-split average of the performance score. The class probability is reported as the average of all test images in the class. Table I demonstrates per-class classification accuracy achieved for both image sets. The average score for the RGB set on five random data splits was 0.84 (plus/minus 0.03); while for the gray set the score was 0.79 (with deviation 0.15).

We also pursued another dimension in our research, implementing the automatic classification on data for two high-content screening experiments (*Binucleate* and *Lack of Centromere* data sets, shown in Fig. 6). The classifier distinguished the binucleate cellular phenotype from normal mononucleate cells, reporting the classification score as 1.0. The classifier performance was a notch lower for the *Lack of Centromere* data set, achieving an average classification rate of 0.84 (standard deviation 0.05).

6. Conclusions

We proposed a global set of image features suitable for description of a wide range of image modalities based on a machine vision approach. The flexibility of the set is a result of a) use of a thorough bank of image characterization algorithms, and b) an unconventional way of applying the algorithms to image transforms.

The approach was applied to several biomedical and biological problems including automatic classification of lymphoma malignancies and high-content screening experiments. The method demonstrated good performance for each of the image classification problems, despite the range of image types.

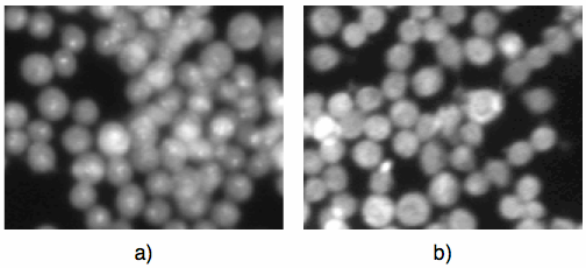

Figure 6. RNAi screening experiment: a) control cells, b) cells lacking centromere.

Gray set (0.79 ± 0.15)			
True Class	Reported Class		
	CLL	**FL**	**MCL**
CLL	0.84	0.07	0.09
FL	0.05	0.91	0.04
MCL	0.19	0.20	0.61
RGB set (0.84 ± 0.03)			
True Class	Reported Class		
	CLL	**FL**	**MCL**
CLL	0.86	0.06	0.08
FL	0.06	0.90	0.04
MCL	0.08	0.15	0.77

Table I. Per-class accuracy for two lymphoma sets. Classification rates are average of four random data splits into training/test partitions.

Any application producing large amounts of images needs an automatic method for annotation/analysis, which is provided by the proposed method. The method has demonstrated multi-purpose capabilities for a range of biomedical problems. The impact of using a multi-purpose classifier is significant: the approach can be applied to a wide range of image types across different biomedical domains.

Acknowledgements

This research was supported by the Intramural Research Program of the NIH, National Institute on Aging.

References

[1] S. P. Awate, T. Tasdizen, N. Foster, & R. T. Whitaker, Adaptive Markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification, *Medical Image Analysis, 10*, 2006, 726-739.

[2] M. V. Boland & R. F. Murphy, A Neural Network Classifier Capable of Recognizing the Patterns of all Major Subcellular Structures in Fluorescence Microscope Images of HeLa Cells, *Bioinformatics, 17*, 2001, 1213-1223.

[3] K. Rodenacker & E. Bengtsson, A feature set for cytometry on digitized microscopic images, *Analytical Cellular Pathology*, *25(1)*, 2003, 1-36.

[4] I.B. Gurevich & I.V. Koryabkina, Comparative analysis and classification of features for image models, *Pattern Recognition and Image Analysis*, *16*, 2006, 265-297.

[5] N. Orlov, J. Johnston, T. Macura, C. Wolkow, & I. Goldberg, Pattern recognition approaches to compute image similarities: application to age related morphological change, *Proc. Int.l Symp. on Biomedical Imaging*, Arlington, VA, 2006, 1152-1156.

[6] N. Orlov, J. Johnston, T. Macura, L. Shamir, and I. Goldberg, Computer vision for microscopy applications, *Vision Systems – Segmentation and Pattern Recognition* (ARS Pub., Vienna, Austria, 2007), 221-242.

[7] M. Cuadros, S. S. Dave, E. S. Jaffe, E. Honrado, R. Milne, J. Alves, J. Rodriguez, M. Zajac, J. Benitez, L. M. Staudt, & B. Martinez-Delgado, Identification of a proliferation signature related to survival in nodal periipheral T-cell lemphomas, *Journal of Clinical Oncology, 25*, 2007, 3321-3329.

[8] B.V. Funt & G.D. Finlayson, Color Constant Color Indexing, *IEEE Trans. on Pattern Analysis and Machine Intelligence, 17(5)*, 1995, 522-529.

[9] H. Tamura, S. Mori, & T. Yamavaki, Textural features corresponding to visual perception, *IEEE Trans. on Systems, Man, and Cybernetics, 8*, 1978, 460-472.

[10] E. Hadjidementriou, M. Grossberg, & S. Nayar, Spatial information in multiresolution histograms, *Proc. IEEE Conf on Computer Vision and Pattern Recognition*, Kauai, HI, 1, 2001, 702.

[11] D. Foran, D. Commaniciu, P. Meer, & L. Goodel, Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy, *IEEE Trans. on Information Technology in Biomedicine, 4(12)*, 2000, 265-273.

[12] O. Tuzel, L. Yang, P. Meer, & D. Foran, Classification of hematologic malignancies using texton signatures, *Pattern Analysis and Applications, 10(4)*, 2007, 277-290.

[13] B. Nelsen, F Albregtsen, & H. Danielsen, Low dimensional adaptive texture feature vectors from class distance and class difference matrices, *IEEE Trans. on Medical Imaging, 23*, 2004, 73-84.

[14] R. Tagaya, N. Kurimoto, H. Osada, & A. Kobayashi, Automatic objective diagnosis of lymph nodal disease by B-mode images from convex-type echobronchoscopy, *Chest, 133*, 2008, 137-142.