Image Classification of Vascular Smooth Muscle Cells

Michael A. Grasso, MD, PhD
University of Maryland School of Medicine
Baltimore, Maryland
mikegrasso@umbc.edu

Ronil Mokashi
University of Maryland Baltimore County
Baltimore, Maryland
ronil1@umbc.edu

Darshana Dalvi
University of Maryland Baltimore County
Baltimore, Maryland
dalvi1@umbc.edu

Antonio Cardone, PhD
University of Maryland
College Park, Maryland
antonio.cardone@nist.gov

Alden A. Dima, MS
National Institute of Standards & Technology
Gaithersburg, Maryland
alden.dima@nist.gov

Kiran Bhadriraju, PhD
National Institute of Standards & Technology
Gaithersburg, Maryland
kiran.bhadriraju@nist.gov

Anne L. Plant, PhD
National Institute of Standards & Technology
Gaithersburg, Maryland
anne.plant@nist.gov

Mary Brady, MS
National Institute of Standards & Technology
Gaithersburg, Maryland
mary.brady@nist.gov

Yaacov Yesha, PhD
University of Maryland Baltimore County
Baltimore, Maryland
yayesha@umbc.edu

Yelena Yesha, PhD
University of Maryland Baltimore County
Baltimore, Maryland
yeyesha@umbc.edu

ABSTRACT
The traditional method of cell microscopy can be subjective, due to observer variability, a lack of standardization, and a limited feature set. To address this challenge, we developed an image classifier using a machine learning approach. Our system was able to classify cytoskeletal changes in A10 rat smooth muscle cells with an accuracy of 85% to 99%. These cytoskeletal changes correspond to cell-to-matrix interactions. Analysis of these changes may be used to better understand how these interactions correspond to certain physiologic processes.

Categories and Subject Descriptors
I.4.0 [**Image Processing and Computer Vision**]: General – *Image processing software*; J.3 [**Computer Applications**]: Life and Medical Sciences – *biology and genetics.*

General Terms
Algorithms, Design, Measurement

Keywords
digital image processing, machine learning, cell biology

1. INTRODUCTION
The traditional method of cell microscopy is through visual inspection of nuclei, organelles, and morphology, with observations made based on variations from expected appearance. This approach can be subjective, due to observer variability, a lack of standardization, and a limited feature set. Computational image classification can be used to provide more quantitative data for a number of applications in the field of molecular biology [1]. Our goal is to provide a set of quantitative measures which correlate well with visual appearance and allow for intracellular comparisons.

Our specific interest is to better understand how vascular smooth muscle cell proliferation is regulated by the mechanical stiffness of arteries. A study is currently in progress, which experimentally manipulates the stiffness of the collagen matrix for these cells in vitro (in a laboratory setting) [2,3]. The resulting morphologic changes in these cells correspond to its cell-to-matrix interactions [4]. It is these morphologic changes that we hope to quantify and classify.

IHI'10, November 11–12, 2010, Arlington, Virginia, USA.

2. BACKGROUND
Computational image classification is used to categorize an image into a finite set of classes based on their intrinsic features. When considering meaningful features for describing an image, the three fundamental patterns include spectral, textural, and contextual features. Spectral features describe the tonal variations, which can be measured as a distribution and represented as a histogram [5]. Textural features contain information about the spatial distribution of tonal variations and are represented as a co-occurrence matrix [6]. Context features are derived from the spectral and textural features in adjacent regions of interest.

The data extracted from these features are organized as vectors. These feature vectors can be compared to determine the similarity between any two images. Distance metrics, such as the Jeffrey Divergence, are commonly used for this comparison [7]. A machine learning approach, such as a support vector machine, can also be employed [8,9]. A support vector machine is a set of supervised learning methods. It can be used for image classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories, based on a set of image features.

Related efforts include the use of biomedical image classification to increase the diagnostic accuracy of prostate cancer biopsies [10], breast cancer biopsies [11], skin cancer detection [12], colonic polyps [13], hysteroscopy video segmentation [14], and echocardiogram analysis [15].

We have previously used this approach to segment surgical videos. Our image classifier recognized key events in laparoscopic cholecystectomy videos (a minimally invasive surgical procedure to remove the gall bladder). Our classifier had an accuracy of about 91% [16].

For this current effort, we have already reported on an initial image classifier [17]. We used this initial classifier to compare non-geometric image features in order to classify morphologic changes to the actin cytoskeleton of smooth muscle cells. This initial classifier used the Manhattan distance metric with an experimentally derived threshold, and had an accuracy of 87%. The remainder of this paper describes a second image classifier we developed. This new classifier applies a support vector machine to non-geometric image features in order to classify morphologic changes of smooth muscle cells.

3. METHODS

We obtained a set of vascular smooth muscle cell images from the A10 rat cell line, stained with an antibody for the actin cytoskeleton. The images were divided into four classes, based on whether they were cultured on a mechanically flexible or mechanically stiff fibrillar collagen extracellular matrix, and whether they were fixed one hour or 24 hours after plating.

A second set of A10 cell images was also obtained, which contained images stained for a specific cytoskeletal element (either actin, myosin, or focal adhesions). The images were divided into four classes, based on whether the culture matrix was mechanically stiff collagen fibrils, mechanically flexible collagen fibrils, monolayer collagen, or fibronectin. All cells were fixed 24 hours after plating.

The images were converted to 8-bit gray scale JPEG format before analysis. Representative cells, stained for different cytoskeletal elements, are shown in Figure 1.

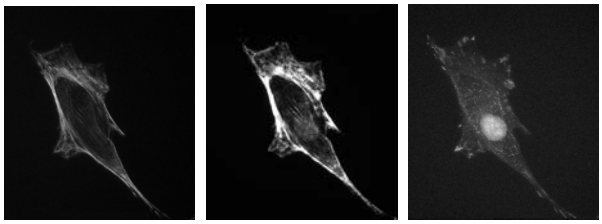

Figure 1. An A10 cell spread for 24 hours on a stiff matrix and stained for the actin (left), myosin (center), and focal adhesions (right).

We used the ImageJ image processing software to extract features from the cell images (http://rsbweb.nih.gov/ij). The image classifier was developed using the Weka machine learning software with a Bayesian network learning algorithm (http://sourceforge.net/projects/weka). The naïve Bayesian classifier uses training data to compute the probability of each image feature. The probability that a set of features belongs to a given class is done by applying Bayes' rule. An assumption is made about the independence of the image features, which although theoretically unrealistic, works reasonably well in practice.

As previously reported, we considered more than 60 spectral and textural features, and included the most promising features in our image classifier [17]. These included the distribution, medium, mode, homogeneity, energy, entropy, and inverse difference moment.

Both sets of images were categorized with the new image classifier, and analyzed using a 10-fold cross validation process.

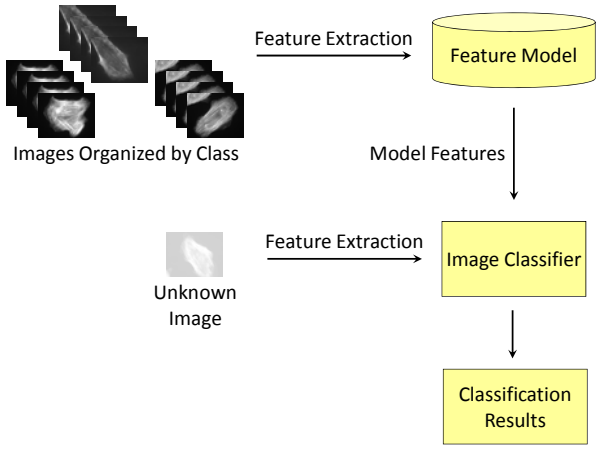

Figure 2. Image Classification Approach.

Our image classification approach is summarized in Figure 2. Training images were organized by class, based on the matrix properties employed or the type of cytoskeletal element that was stained. Features were extracted from the training images to create a feature model. An unknown image is then classified by comparing its features to the feature model. With 10-fold cross-validation, 90% of the images were randomly selected as training data, with the remaining images used for testing. This process was repeated four times for each image set.

Principal outcome variables were the classification results, identified as true positives, true negatives, false positives, and false negatives. These were used to identify the sensitivity (true positives / true positives + false negatives), specificity (true negatives / true negatives + false positives), and accuracy (true positives + true negatives / all positives + all negatives) of our image classifier.

4. RESULTS

For actin-stained images, our classifier was able to categorize images from the first image set with a sensitivity of 64%, specificity of 97%, and accuracy of 89%. The second image set was categorized with a sensitivity of 84%, specificity of 87%, and accuracy of 85%. For all images in the second set (actin, myosin, focal adhesions), our classifier was able to categorize the type of cytoskeletal element with a sensitivity of 99%, a specificity of 98%, and an accuracy of 99%. The results are shown in Tables 1, 2, and 3.

5. DISCUSSION

We developed an image classifier to categorize A10 rat smooth muscle cells according to morphologic changes in the cytoskeleton. This asymmetry in cell organization corresponds to cell-to-matrix interactions, and may be used to better understand how these interactions correspond to certain physiologic processes.

Our image classifier shows promise, with an accuracy of 89%, 85%, and 99% for the three experiments, respectively. We are currently working on ways to increase the classification accuracy by identifying new image features, applying different kernel functions to the support vector machine, and experimenting with feature clustering [18]. We are also studying the use of

classification and regression trees, which have the advantage of enabling understanding of how classification decisions depend on the image features [19].

Table 1. Actin-stained images from image set 1, classified by differences in the cell matrix.

Image Set 1 - Actin	Value
n	11
Sensitivity	64%
Specificity	97%
Accuracy	89%

Table 2. Actin-stained images from image set 2, classified by differences in the cell matrix.

Image Set 2 - Actin	Value
n	57
Sensitivity	84%
Specificity	87%
Accuracy	85%

Table 3. All images from image set 2, classified by differences in cytoskeleton element (actin, myosin, or focal adhesions).

Image Set 2 - Cytoskeleton	Value
n	171
Sensitivity	99%
Specificity	98%
Accuracy	99%

When interpreting these results, it is important to consider several limitations. The image datasets were small in size, which is likely the cause of the low sensitivity in the first image set. As our data-acquisition efforts continue, we expect to have a dataset of images from more than 500 cells. In addition, the images were restricted to a single cell line and specific cytoskeletal structures, which may not be generalizable to other cell lines or cell structures.

The focus of this effort was on non-geometric spectral and textural features. In a separate effort, we are working on new ways to segment cells beyond typical thresholding techniques, in order to identify clinically significant subcellular regions of interest. This information is being used to extract geometric features from cells, and quantify the size, orientation, distribution, and correlation among these regions of interest.

6. ACKNOWLEDGMENTS

This work was funding through contract 00005357 from the National Institute of Standards and Technology.

7. REFERENCE

1. Elis R, Athale C. Computational imaging in cell biology. J Cell Biol. 2003 May 12;161(3):477-81.
2. Plant AL, Bhadriraju K, Spurlin TA, Elliott JT. Cell response to matrix mechanics: focus on collagen. Biochim Biophys Acta. 2009 May;1793(5):893-902.
3. McDaniel DP, Shaw GA, Elliott JT, Bhadriraju K, Meuse C, Chung KH, Plant AL. The stiffness of collagen fibrils influences vascular smooth muscle cell phenotype. Biophys J. 2007 Mar 1;92(5):1759-69.
4. Théry M, Racine V, Piel M, Pépin A, Dimitrov A, Chen Y, Sibarita JB, Bornens M. Anisotropy of cell adhesive microenvironment governs cell internal organization and orientation of polarity. Proc Natl Acad Sci U S A. 2006 Dec 26;103(52):19771-6.
5. Kunttu1 I, Lepistö1 L, Rauhamaa J, Visa1 A. Binary histogram in image classification for retrieval purposes. Journal of WSCG. 2003; 11(1):269-273.
6. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Transaction on Systems, Man, and Cybernetics. 1973 Nov; 3(6):610-621.
7. Bugatti PH, Traina AJ, Traina C. Assessing the best integration between distance-function and image-feature to answer similarity queries. Proceedings ACM SAC. 2008;:1225-1230.
8. Burges, C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. 1998;2(2):121-167.
9. Boland MV, Murphy RF. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. Bioinformatics. 2001;17:1213–1223.
10. Doyle S, Shah K, Madabhushi A, Feldman M, Tomaszeweski J. Automated grading of prostate cancer using architectural and textural image features. Biomedical Imaging: from Nano to Macro, 2007;:1284-1287.
11. Doyle S, Agner S, Madabhushi A, Feldman M, Tomaszeweski J. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. Biomedical Imaging: from Nano to Macro, 2008;:496-499.
12. Yuan X, Yang Z, Zouridakis G, Mullani N. SVM-based texture classification and application to early melanoma detection. Conf Proc IEEE Eng Med Biol Soc. 2008;2008:50-3.
13. Li B, Meng MQ, Xu L. A comparative study of shape features for polyp detection in wireless capsule endoscopy images. Conf Proc IEEE Eng Med Biol Soc. 2009;2009:3731-4.
14. Scharcanski J, Gaviao Neto W, Cunha-Filho JS. Diagnostic hysteroscopy video summarization and browsing. Conf Proc IEEE Eng Med Biol Soc. 2005;6:5680-3.
15. Roy A, Sural S, Mukherjee J, Majumdar AK. State-based modeling and object extraction from echocardiogram video. IEEE Trans Inf Technol Biomed. 2008 May;12(3):366-76.
16. Grasso MA, Finin TW, Zhu X, Joshi A, Yesha Y. Video summarization of laparoscopic cholecystectomies. AMIA Annu Symp Proc, 2009, Nov 14-18.
17. Grasso MA, Mokashi R, Dima AA, Cardone A, Bhadriraju K, Plan AL, Brady M, Yesha Y. Image classification in cell biology. AMIA Annu Symp Proc, 2010.
18. Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. Applied Statistics. 1979;28(1):100-108.
19. Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. Chapman & Hall/CRC, 1984.