Measuring visual structure in phase and fluorescence microscopy using image compression

Rohini Joshi[1], Sundar Ram Swaminathan[1], Mark Winter[1], Janmeet S. Saini[2], Timothy A. Blenkinsop[3], Jeffrey Stern[2], Sally Temple[2] and Andrew R. Cohen[1*]

Abstract **–Measuring visual properties of cell and tissue in biological microscopy images is a broadly important task. Typically, such measurements are accomplished by quantifying features extracted from the images by automated analysis algorithms. In some cases it is desirable to have approaches to measure visual characteristics from microscopy images that do not require feature extraction. We have applied a new approach using wavelet compression combined with a Kolmogorov complexity-based distance metric that enables robust classification on biological imaging data. Results are presented for two diverse applications: (1) measuring 'cobblestone' structure in RPE cells and (2) identifying mitotic events in time-lapse images of embryonic neural stem cells.**

Index Terms**—Retinal pigment epithelium, mitosis detection, cobblestone, normalized compression distance, wavelet similarity measure.**

I. Introduction

Identifying structures in biological microscopy images typically requires a segmentation or feature extraction step that is applied to the images. The resulting features are then used to identify or classify objects of interest in the images. In many cases, it is difficult or impossible to reliably extract features for object detection and classification, and non-feature based classification approaches are desirable. Here we present a non-feature based classification approach for identifying structure in biological images. We apply this technique to two diverse applications – mitosis detection and cobblestone morphology characterization in retinal pigment epithelial (RPE) cells.

The RPE is a pigmented cell layer in the retina that nourishes retinal visual cells and is responsible for providing vision. Healthy RPE is characterized by a distinctive "cobblestone" morphology, as shown in Figure 1. We present a novel technique to measure the degree of cobblestone structure in RPE images.

Mitosis detection is a fundamental task in the analysis of time-lapse microscopy images of proliferating cells. Mitosis detection provides a valuable measure of cell population growth on its own, and can also be used to improve segmentation, tracking and lineaging in the analysis of time- lapse movies showing clonal development [1, 2].

The work presented here uses the previously developed normalized compression distance (NCD) [3], together with a novel compression approach based on wavelets formulated specifically to work with the compression (NCD). This work is notable in the use of a single approach to robustly address these two diverse, important and challenging classification applications in biological microscopy.

There are other existing approaches that can be used to quantify cobblestone morphology. These approaches require feature extraction, computing characteristics from image pixels to identify cobblestone structure. Examples include methods were based on edge detection, extracting information from gradient maps [4]. Another approach involves a mathematical framework for steerable filters developed using parametric steerable template matching based methods [5]. Steerable designs provide an advantage over the other methods due to fact that they make no prior assumptions about the type of junction or angular separation. They are scale and rotation invariant and allow for a template design of any order of symmetry [6]. We have found steerable wavelet filters very effective at detecting cobblestone morphology in fluorescence RPE images, but less effective for phase contrast.

Many mitosis detection methods have been proposed for phase-contrast time-lapse microscopic images. Huh et al. [7] proposed a method involving Event Detection Conditional Random Fields. Ciresan et al. [8] used deep neural networks address the question. The approach described here is unique in that we have applied our technique to identifying over 3000 mitotic and non-mitotic events captured in phase contrast time-lapse images of embryonic neural stem cells. Our approach also differs in that time sequences of images were available for allowing us to incorporate temporal information. Another notable difference is that we achieve excellent classification accuracy (97%) using a minimum distance clustering-based classifier, using the distance measure to capture similarity for effective classification rather than requiring more sophisticated classification algorithms.

II. Approach

A. Normalized Compression Distance

In order to classify a cell culture, we measure its 'similarity' to a training set of images known to belong to a particular category. The NCD is an effective technique for measuring similarity among complex digital data. The NCD measures similarity between digital objects on a scale from zero (least similar) to one (most similar). The NCD uses file compression

[1] Electrical & Computer Engineering, Drexel University, Philadelphia, PA
[2] Neural Stem Cell Institute, Rensselaer, USA
[3] Developmental & Regenerative Biology, Mount Sinai Hospital, NY
Correspondence to acohen@coe.drexel.edu

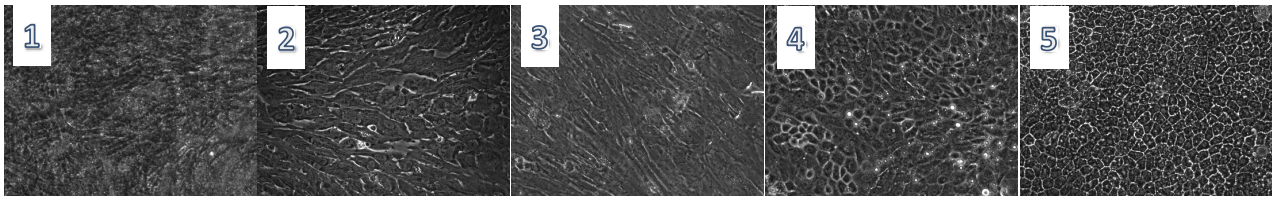

Figure 1. Cobblestone morphology. The images are labelled manually on a scale from (1) to (5) where 5 is a healthy cobblestone structure for the RPE culture. Our approach based on the multiset NCD with wavelet compression was able to classify 34 images with 100% accuracy into the five classes.

algorithms as a basis for approximating the relative Kolmogorov complexity [9],

$$NCD(X) = \frac{G(X) - \min_{x \in X} G(x)}{\max_{x \in X} G(X \setminus x)} \qquad (1)$$

where X is a multiset of digital objects and G is the size in bytes achieved by a compression algorithm. The most common approach to using the NCD has been with sets X of cardinality two. Recently, the NCD has been extended to compute distances among multisets with more than two elements, extracting information from the group of objects during the computation of the distance measure [3]. This "multiset" formulation has been found to be more accurate in many applications.

B. Wavelet representation with the NCD

The NCD requires a compressed object size in bytes. To obtain a compression measure that captures visual structural similarity, we apply a wavelet filter (SYMLET) to each image, and compute the L1 norm of the approximation sub-band. The L1 norm captures the sparsity in this sub-band, based on how similar pairs or multisets of images are to each other.

We use a similar approach for mitosis detection. We took 1550 mitotic events occurring in 123 clones of mouse embryonic neural stem cells. Images were captured every five minutes. We applied a naïve single frame segmentation [10] to identify candidate cells, and then used the fact that cells undergoing mitosis do not move to identify the potential daughter cells in the subsequent frame. A bounding box of 100 pixels around each detection was used to mask the cells. The masked cells were then compared to a training set of 24 mitotic and non-mitotic images.

C. Quantifying cobblestone formation in RPE cell cultures

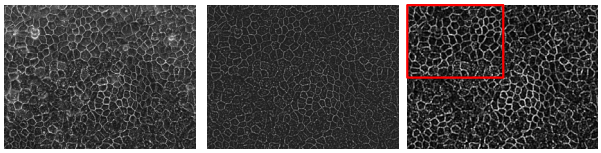

Figure 2. Preprocessing the RPE images. Original image (a), contrast enhancement filtering (b) and texture filter (c), are applied to remove variations in imaging conditions prior to classification.

RPE cells have a tendency to undergo epithelial to mesenchymal transition (EMT) spontaneously in culture, depending upon the passage number and density of passaging. RPE cells in their native epithelial form show cobblestone morphology, however, they become fibroblastic when they undergo EMT. We identified RPE cells at various stages of EMT in culture and manually classified them into 5 categories (Figure 1), with 5 being most cobblestone/epithelial and 1 most fibroblastic/mesenchymal. We took images for each category from multiple RPE lines.

Due to the large variability in images and inherent noise,

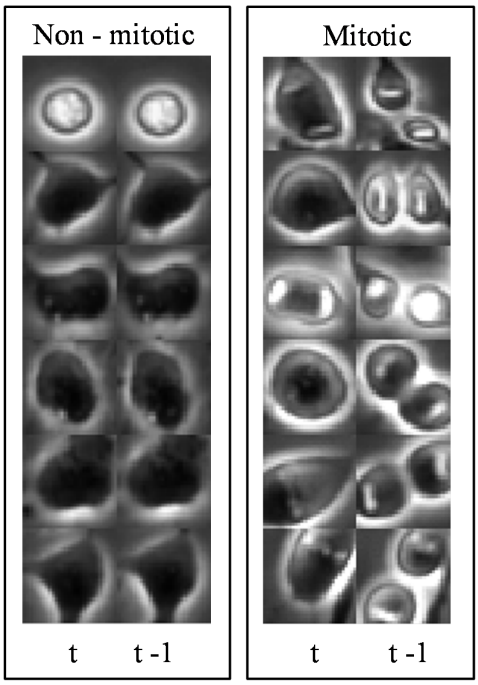

Figure 3. The training data used for mitosis detection. The data consists of pairs of $t-1$ and t frames of neural stem cells. On the left, shown are 6 training samples for mitotic and on the right are the 6 training samples for non-mitotic cells. Each of the $t-1$ and t images concatenated together are of the size 30x60 pixels.

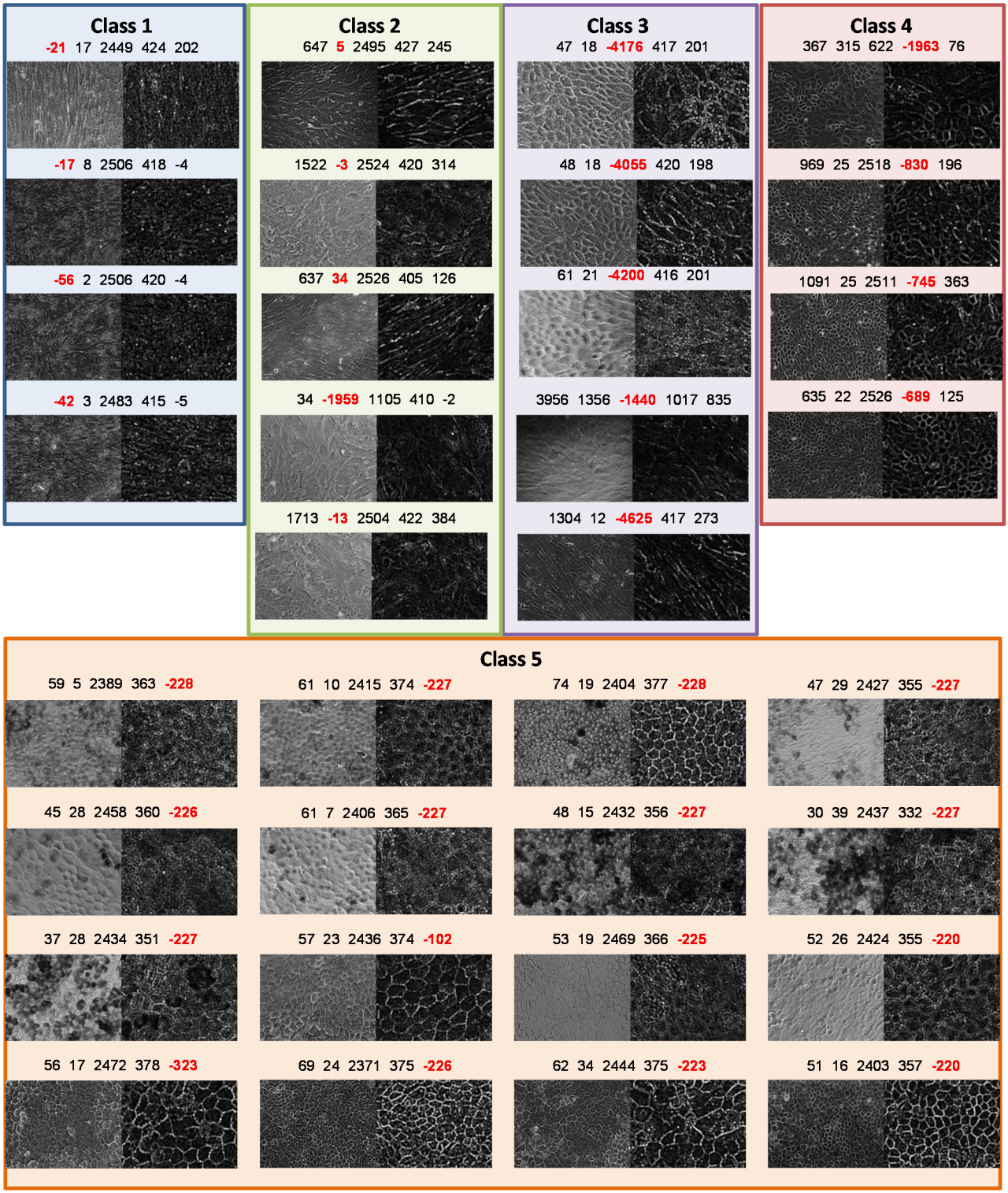

Figure 4: Training images (34) are divided into manually labeled classes 1 (least cobblestone) through 5 (most cobblestone). Numbers above the images are distances to each class measured using leave-one-out cross validation. All distances shown are multiplied by 10^5. All 34

preprocessing is a necessary step. The preprocessing is done in three steps. First, a contrast enhancement filter is applied. Next, a texture filter is applied, computing the standard deviation in a three pixel neighborhood around each pixel.

This enhances the edges and structure (due to the high standard deviation in the neighborhood of cell edges). Figure 2 illustrates the preprocessing steps. For classification we use the top left quarter of the filtered image as input to the compression distance (eqn. 1), and each image is assigned to the training set that achieves the minimum distance using a leave-one-out cross validation.

D. Mitosis Detection

Mitosis detection was applied to 123 clones of embryonic neural stem cells that were cultured as described in Winter et al. [2]. Our mitosis detection approach applies a cell segmentation algorithm originally developed for retinal stem cells [10] and later modified to neural stem cells [2] to each frame in the image sequence. We rely on the fact that during mitosis cells become less motile and identify possible daughter cells in each subsequent frame by spatial overlap. Cells that are matched with two daughter cells candidates in the subsequent frame are then compared to a training set consisting of mitotic and non-mitotic images using the NCD (eqn. 1), and classified. We evaluated both the pairwise NCD [11] and the multiset NCD [3], and also both JPEG2000 compression and the same L1 norm applied to the wavelet filter as with the RPE classification. Results were computed against manually labelled ground truth information.

III. Results

We analyzed 34 phase contrast images of RPE monolayers that were manually labeled with a cobblestone measure from 1 (least cobblestone) to 5 (most cobblestone –viable culture). The resulting change in NCD for each multiset of training images is shown in Figure 3. We also analyzed 30 fluorescence images labelled with ZO-1 that marks tight junctions in RPE cultures. 95% confidence intervals for the classification were computed as described in [12] and are reported in square brackets following the classification result. All 34 phase images were correctly classified, 100% accuracy [0.9, 1.0]. For the ZO-1 images, the images were classified with 87% accuracy [0.71, 0.95]. The results were the same using both the pairwise and multiset NCD formulation.

For the mitosis detection, we analyzed 3,100 image pairs, half belonging to true mitotic events and half belonging to non-mitotic events. The pairwise NCD achieved 97% [0.96, 0.98] accuracy in classifying these using a nearest neighbor minimum distance approach. The multiset NCD did not perform as well as the pairwise, achieving only 93% [0.92, 0.94]. It is unusual for the pairwise NCD to outperform the multiset NCD. The large amount of appearance variation among the training data would suggest that choosing the training sets more carefully could improve performance even further.

IV. Conclusions

Measuring the amount of cobblestone in a microscopy image and detecting mitosis events are two important and complex image analysis tasks. The approach we describe here based on algorithmic information theory does not require feature extraction and is able to reliably process large amounts of image data with high accuracy. We see these approaches as being complementary to other feature-based image analysis techniques, and providing an effective new approach for measuring visual structure in biological images.

V. Acknowledgment

This work was supported in part by research grant R01NS076709 from the National Institute of Neurological Disorders and Stroke of the U.S. National Institutes of Health.

VI. References

[1] W. C. Mankowski, M. R. Winter, E. Wait, M. Lodder, T. Schumacher, S. H. Naik*, et al.*, "Segmentation of occluded hematopoietic stem cells from tracking," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014, pp. 5510-5513.

[2] M. Winter, E. Wait, B. Roysam, S. K. Goderie, R. A. N. Ali, E. Kokovay*, et al.*, "Vertebrate neural stem cell segmentation, tracking and lineaging with validation and editing," *Nature protocols,* vol. 6, pp. 1942-1952, 2011.

[3] A. R. Cohen and P. M. B. Vitanyi, "Normalized Compression Distance of Multisets with Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence (in press),* 2015.

[4] D. J. Beymer, "Finding junctions using the image gradient," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, 1991, pp. 720-721.

[5] M. Jacob and M. Unser, "Design of steerable filters for feature detection using canny-like criteria," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 26, pp. 1007-1019, 2004.

[6] Z. Puspoki, C. Vonesch, and M. Unser, "Detection of symmetric junctions in biological images using 2-D steerable wavelet transforms," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013, pp. 1496-1499.

[7] H. Seungil, D. F. E. Ker, R. Bise, C. Mei, and T. Kanade, "Automated Mitosis Detection of Stem Cell Populations in Phase-Contrast Microscopy Images," *Medical Imaging, IEEE Transactions on,* vol. 30, pp. 586-596, 2011.

[8] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," *Med Image Comput Comput Assist Interv,* vol. 16, pp. 411-8, 2013.

[9] M. Li and P. Vitányi, "A New Approach to Formal Language Theory by Kolmogorov Complexity," *SIAM Journal on Computing,* vol. 24, pp. 398-410, 1995/04/01 1995.

[10] A. R. Cohen, F. Gomes, B. Roysam, and M. Cayouette, "Computational prediction of neural progenitor cell fates," *Nat Methods,* vol. 7, pp. 213 - 218, Mar 2010.

[11] A. R. Cohen, C. Bjornsson, S. Temple, G. Banker, and B. Roysam, "Automatic Summarization of Changes in Biological Image Sequences using Algorithmic Information Theory," *IEEE Trans Pattern Anal Mach Intell,* vol. 31, pp. 1386-1403, Aug 2009.

[12] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques* Second ed., 2005.