Project Title:
Haralick Texture Analysis for Stem Cell Identification

Student:
Nathan Loewke (noloewke@stanford.edu)

Specific Goals and Work Involved:
The field of stem cells and regenerative medicine has many basic questions still unanswered. Of particular interest is the analysis of cell culture and understanding what factors determine, or even predict, clinically viable cell cultures. Through prior analysis of time-lapse data, I’ve noticed that human embryonic and induced pluripotent stem cell (hESC and iPSC) colonies differentiate in particular patterns and in response to certain conditions such as media exchange. I’d like to understand these processes better, and find a way to determine differentiation status without the use of chemical reagents or fluorescent labels. I hypothesize that the pattern and rates of differentiation status change in stem cell colonies can be a predictive biomarker for cell culture viability. As these cells differentiate, they change visually in terms of packing density and shape. To quantify what I see in the lab, I’d like to use texture analysis and some simple machine learning to classify each frame of data into three categories: undifferentiated cells, differentiated cells, and background.

My project may be divided into the following specific goals:
1. Identification of clearly defined and easily distinguishable texture samples that may or may not be computer generated. These grayscale data represent an easier test case and will be used during the development of my texture analysis software.
2. Development of a simple graphical user interface (GUI) for identification and supervised learning of training data. The interface will allow the user to select sub regions of test data and label them as one of the identifying classes.
3. Calculation of Haralick textural features based on gray level co-occurrence matrices [1]. This is a well-known method for quantifying textural features, and gives information about the image region such as homogeneity, contrast, boundaries, and complexity. A recent paper used Haralick features to predict malaria based on image analysis of erythrocytes [2].
4. Rotation invariance. Because I don’t want to distinguish between cells based on direction orientation, I’ll need to convert these features into a rotation invariant form by, for example, calculating the mean, range, and deviation of each set of features calculated for each direction.
5. Dimension reduction and classification. Because the feature observation data will be of high dimensionality, I’ll most likely need to reduce the dimensionality of our feature
vectors using either subsampling, principle component analysis (PCA), or linear discriminant analysis (LDA). The exact method is yet to be determined, as I’ll need to figure out which features in particular will give the best results. I’ll then use the classification of every feature observation to form decision boundaries in feature space to classify test data. A recent paper used to a PCA-based approach in combination with Haralick texture features to classify synthetic aperture radar (SAR) images [3].

6. Testing on biological data. The data to be tested will be of hESCs and/or iPSCs imaged for 7-14 days using phase contrast microscopy (PCM) and/or quantitative phase microscopy (QPM). The imaging mode here is not important, but recognizing which modality is easier to recognize is crucial.

7. Finally, if time permits, I’d like to graph differentiation rates vs. media exchange rates to understand the mechanisms involved.

Android Device:
Seeing as this project is aimed at developing a tool for biologists in a laboratory-type setting, this project will not be implemented on an Android device.

References: