Unsupervised Segmentation for Myofiber Counting in Immunofluorescent Microscopy Images

K.L. Urish[1,2 *], J. August[3], and J. Huard[1,2,4]

[1] Department of Bioengineering, The University of Pittsburgh, Pittsburgh, PA
[2] Growth & Development Laboratory, Children's Hospital of Pittsburgh, Pittsburgh, PA
[3] Robotics Institute, Carnegie Mellon University, Pittsburgh, PA
[4] Department of Orthopedics, The University of Pittsburgh, Pittsburgh, PA

*klu6@pitt.edu

Abstract. The gold standard for measuring muscle regeneration in muscular dystrophy therapies is counting the number of dystrophin-positive muscle fibers on a cryostat muscle section immunostained for dystrophin. The standard process of manually counting a few thousand myofibers is tedious, time consuming, and limits quantitative analysis of a therapy's success. We present an unsupervised method for segmenting and counting the number of myofibers on an immunofluorescent microscopy image. The key threshold selection problem is resolved by maximizing the number of sub-threshold connected components. Components significantly smaller than the known lower bound myofiber area, the only input parameter, are ignored to reduce noise. Validation on a series of images (n=63) revealed that our algorithm varied by less than 10% from manual counts in the relevant range of operation. The algorithm allows us to quantify three-dimensional dystrophin expression and design experiments that address a major limitation in muscular dystrophy therapies, the limited distribution of dystrophin after treatment. Further we have extended this method to segment and count objects in other immunofluorescent images. The method was quickly developed and tested using the Insight Toolkit (ITK), an open source C++ library for the development of image analysis software.

1. Introduction

Immunohistochemistry is a valuable technique that measures protein expression in cell culture and tissue sections by labeling a protein, such as dystrophin, with an antibody attached to a fluorescent signal. Many research groups are interested in automating the measurement of muscle regeneration in therapies for muscular dystrophy. The gold standard for measuring muscle regeneration uses immunohistochemistry to stain dystrophin-positive myofibers on a muscle section so that the number of regenerated myofibers can be counted. Current therapies result in a few thousand myofibers being generated [1]. Manually counting such a large quantity is tedious and time consuming, but remains standard practice due to the lack of an accurate automated processes. Further, this analysis only assesses the local expression of dystrophin, but ignores the longitudinal distribution of dystrophin expression, ie, the distribution of dystrophin expression as a function of position along the muscle fiber. This is one of the main limitations of current therapies [2]. To design experiments that address this problem, we developed an algorithm

to automatically count the number of dystrophin-positive myofibers on an immunofluorescent image, allowing three-dimensional dystrophin expression to be quantified.

2. Unsupervised Thresholding

Thresholding is often used to segment an image. Ideally, the two populations to be classified (foreground and background) would have distinct ranges of image intensities with minimal overlap to form a bimodal histogram (Fig 1A). However, images of myofibers immunostained for dystrophin (Fig 1B) and immunofluorescent images in general have unimodal image histograms (Fig 1C) where an appropriate threshold is not obvious. We have developed an algorithm to automatically select a threshold for immunofluorescent images by searching for the threshold which maximizes the number of connected components of the thresholded image.

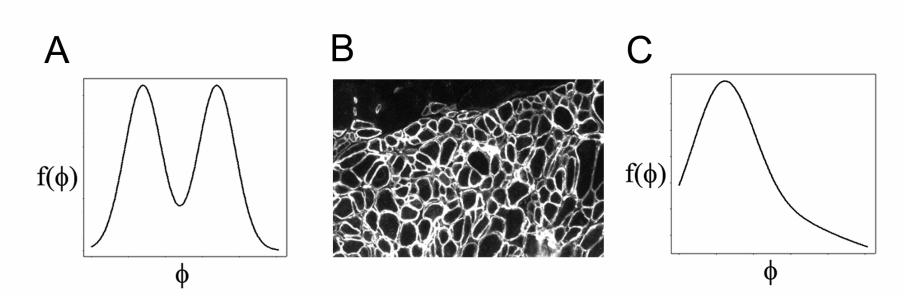

Fig. 1. The threshold used to separate two populations is apparent if the image histogram is bimodal (A). However, an immunofluorescent image of myotubes immunostained for dystrophin (B) has a unimodal image histogram (C), and the appropriate threshold is not clear. Note that (B) is only a small portion of the overall image of a regenerated muscle.

We have observed that the number of connected components as a function of threshold is smooth in the vicinity of the maximum. The search for this optimal threshold is implemented using a modified bisection method. Cells typically have a pixel area greater than 100 pixels (Retiga 1300 camera at 200x magnification) so that components below 50 pixels, the only input parameter, were ignored to eliminate noise. This approach is based on the following logic. If the threshold is set too low, the number of connected components is too low because some of the myofibers will not be detected (Fig 2B). Conversely, if the threshold is too high, the number of connected components is too low because the white border separating the myofibers is eroded away connecting individual myofibers (Fig 2D). This suggests that a reasonable threshold maximizes the number of sub-thresholded connected components (Fig 2E). If this optimum threshold is used, the image appears to be correctly segmented (Fig 2C) compared to the original image (Fig 2A). This algorithm is a variant of a method termed topological stable-state thresholding [3].

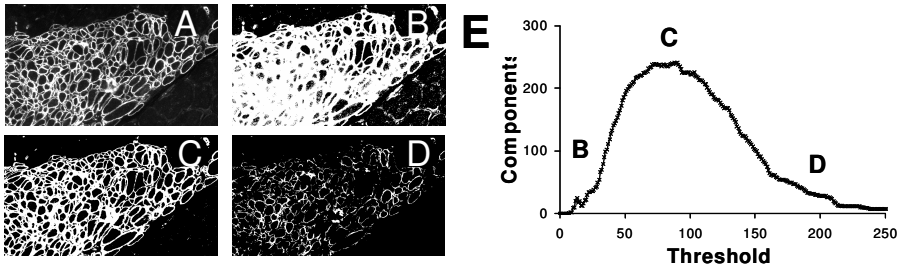

Fig. 2. The original image (A) is thresholded at a low (B), medium (C), and high (D) pixel intensity. The number of connected components above a user defined minimum area is plotted if the image was thresholded at all pixel intensities (E). Images B, C, and D are shown at their respective points on the curve in E. This suggests that an optimal threshold point can be identified based on maximizing the number of connected components in an image.

3. Proposed Method and ITK Implementation

The Insight Toolkit (ITK) is an open source C++ library used to design and implement algorithms for image analysis [4], and was used to rapidly assemble and test the feasibility of this method. Figure 3 outlines the pseudo-code of the algorithm using a modified bisection method. The minimum and maximum pixel intensities of the image are the set as the two endpoints. Each segment between an endpoint and the midpoint is bisected for the initial two evaluation points. The image is thresholded at each point, and the ITK classes, ConnectedComponentsImageFilter and RelabeledComponentsImageFilter, are used to count the number of connected components above the minimum object size on each binary image. The point with the higher number of objects is set as the new midpoint, and the point with the lower number of objects is set as the new endpoint. The process is repeated until the two endpoints are two pixel intensities apart. To decrease the step size of each iteration, the rate of bisection can be damped. Any number of other numerical methods for optimization without derivatives can be used to search for the maximum including Brent's and the golden ratio method [5].

This method of unsupervised segmentation has two inherent advantages. First, the only necessary input is the minimal area of the segmented objects; this parameter is known a priori, which we can exploit to help eliminate background noise. Second, the method employs a natural coarse-to-fine search strategy. Since the function is sampled at the three most widely separated intensities in the initial iteration, the function is, in effect, smoothed at a large scale and thus avoid local maximums. The sample selection has the effect of reducing the degree of smoothing with each iteration, i.e., decreasing as the search scale as the estimated maximum is approached. (We note that subsampling is a well-known strategy for regularizing ill-conditioned inverse problems, such as numerical differentiation.) It is possible that a local maximum near the global maximum may be identified; however this is an equally acceptable solution as little difference exists between the two answers in terms of number of components, which is the desired objective.

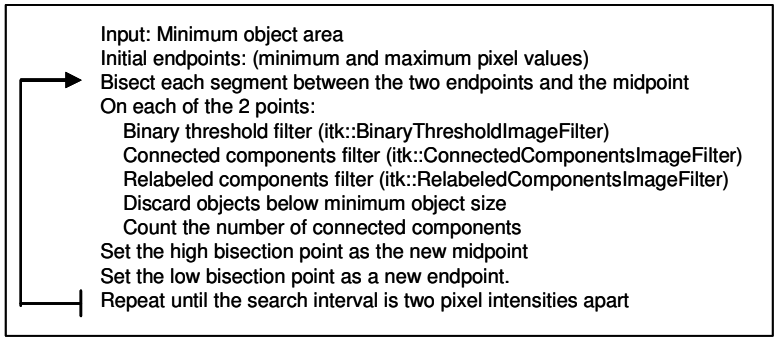

Fig. 3. The proposed method for unsupervised segmentation of immunofluorescent myofiber images uses a modified bisection method to search for the threshold point.

Further, ITK is a valuable tool to implement automated threshold-based segmentation of immunofluorescent images. Different numerical methods or other approaches can easily be tested to identify the best optimization or search strategy. The generic ITK pipeline allows the easy addition of edge enhancement features in situations where thresholding is not an entirely sufficient method for segmentation. In addition, standard image processing operations such as connected components that are challenging to code and debug are already reliably implemented in ITK.

4. Results and Discussion

The method was validated on a set of 63 images with a range of camera exposure, myofiber numbers, and immunostaining conditions. For images with more than 100 myofibers, there was less than a 10% difference between manual and our automated counts. For images with less than 100 myofibers, there was more variation; however this is below the needed operating range of our algorithm (Fig 4A).

To test the ability of this approach to measure three-dimensional regeneration, a series of serial cryostat sections of a regenerated muscle was stained, imaged, and measured with our algorithm. A longitudinal plot of the number of dystrophin-positive myofibers as a function of the length of the muscle, starting at the beginning of the engraftment, demonstrates the three-dimensional measurement of tissue regeneration (Fig 4B). A major obstacle in therapies for muscular dystrophy is the low migration of dystrophin gene vectors that result in a limited longitudinal distribution of dystrophin expression [2]. This automated algorithm will facilitate quantifying this measurement, allowing us to design experiments that address the limited distribution of dystrophin-positive myofibers.

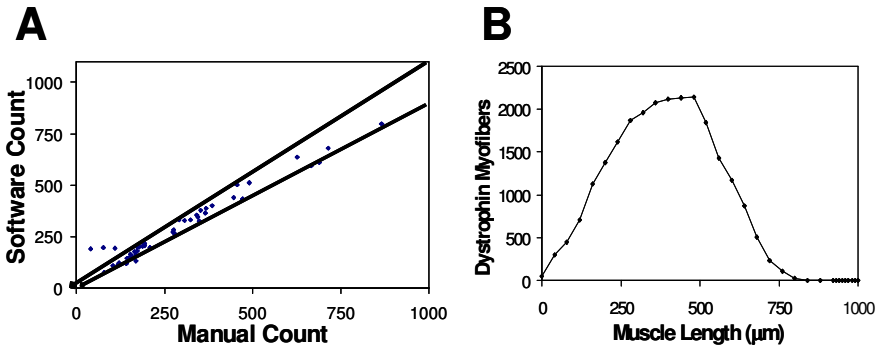

Fig. 4. Validation of this algorithm (n=63) revealed that images that contained more than 100 objects varied by less than ±10% (solid lines) from manual counts (A). After validation, this algorithm was used to measure the number of dystrophin-positive myofibers on a set of 30 serial cryostat cross sections of a single muscle, and a longitudinal plot of the counts displays the 3-D distribution of dystrophin expression (B).

References

1. Qu-Petersen Z, Deasy B, Jankowski R, Ikezawa M, Cummins J, Pruchnic R, Mytinger J, Cao B, Gates C, Huard J. Identification of a novel population of muscle stem cells in mice: potential for muscle regeneration. J Cell Biol, 157(5): 851-64, 2002.
2. Urish KL, Kanda Y, Huard J. Initial failure in myoblast transfer therapy has led the way toward the isolation of muscle stem cells: potential for tissue regeneration. Curr Top Dev Biol. 2002. In Press.
3. Pikaz A, Averbuch, A. Digital image thresholding based on topological stable-state. Pattern Recognition, 29(5): 829-843, 1996.
4. Ibanez L, Schroeder W, Ng L, Cates J.: The ITK Software Guide. Kitware Inc. 2003
5. Brent RP. Algorithms for Minimization without Derivatives. Engelwood Cliffs, NJ: Prentice-Hall. 1973.