CHARISMA: An Integrated Approach to Automatic H&E-stained Skeletal Muscle Cell Segmentation Using Supervised Learning and Novel Robust Clump Splitting

Thomas Janssens\textsuperscript{a,1,*}, Laura Antanas\textsuperscript{b,2}, Sarah Derde\textsuperscript{a,3}, Ilse Vanhorebeek\textsuperscript{a,4}, Greet Van den Berghe\textsuperscript{a,5}, Fabian Güiza Grandas\textsuperscript{a,6}

\textsuperscript{a}Department of Intensive Care Medicine, KU Leuven, O8N1 UZ Herestraat 49, 3000 Leuven, Belgium
\textsuperscript{b}Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium

Abstract

Histological image analysis plays a key role in understanding the effects of disease and treatment responses at the cellular level. However, evaluating histology images by hand is time-consuming and subjective. While semi-automatic and automatic approaches for image segmentation give acceptable results in some branches of histological image analysis, until now this has not been the case when applied to skeletal muscle histology images. We introduce CHARISMA, a new top-down cell segmentation framework for histology images which combines image processing techniques, a supervised trained classifier and a novel robust clump splitting algorithm. We evaluate our framework on real-world data from intensive care unit patients. Considering both segmentation and cell property distributions, the results obtained by our method correspond well to the ground truth, outperforming other examined methods.

Keywords: Histology, Skeletal Muscle, Image Segmentation, Machine Learning, Clump Splitting, Concavity Analysis

1. Introduction

The automation of histological image analysis plays a key role in advancing research techniques for disease diagnosis, minimizing time-consuming manual effort and providing vital non-subjective clinical information (Pellicciari (2010);
Pellicciari and Malatesta (2011)). In the case of tissue microscopy, this analysis is performed on images taken from a thin slice of dyed tissue. The size, shape, geometric arrangement and other morphological characteristics of tissue structures and cells provide crucial information to evaluate the clinical state of a patient.

In this work we focus on the automatic analysis of histology images of skeletal muscle, more specifically, of tissue biopsies from intensive care unit (ICU) patients. Skeletal muscle accounts for 40% of human body mass and is the most important protein reservoir of the body. Despite nutritional support, ICU patients can lose up to 10% of their muscle mass in one week (Reid et al. (2004)). The resulting muscle weakness is a severe complication of critical illness that delays rehabilitation, has long-term effects in quality of life and is associated with increased risk of death (Debaveye and Van den Berghe (2006)). It is important to find effective and safe treatments that are able to attenuate this wasting of muscle tissue. Therefore, understanding the response to treatment at the cellular level through analysis of histology images is instrumental. This analysis is however particularly difficult, as the tissue often exhibits pathological degradations related to the patient’s critical condition. Some examples of histology images for Hematoxylin and Eosin (H&E) stained skeletal muscle are illustrated in Figure 1.

The first step towards a comprehensive skeletal muscle tissue analysis is cell segmentation. Although some solutions are available, fully automated segmentation of H&E stained skeletal muscle tissue remains challenging. Existing frameworks, such as CellProfiler (Carpenter et al. (2006)), CellTracer (Wang et al. (2010)) and CellTracker (Shen et al. (2006)) have several critical shortcomings when used for histological cell segmentation in general. First, because they rely solely on general image processing techniques they are unable to deal simultaneously with imaging artifacts, irrelevant tissue, clumps of cells, and degraded cells. Second, important parameters for segmentation, such as expected cell size, need to be manually initialized. Additionally, these tools generate only basic statistics mainly related to cell areas. In a clinical setting, more detailed information about cell structure such as cell shape, texture, staining intensity, number of nuclei, integrity, etc. is needed to fully characterize the imaged tissue.

Consequently, because the quality of the segmentation is essential for correct interpretation, researchers often rely on manual or semi-automatic methods.
These include tracing cells by hand, or using a general purpose image editing tool such as Adobe Photoshop. Semi-automatic tools can be more efficient, but they still require pre-processing such as removing artifacts or irrelevant tissue, and post-processing, e.g., correcting segment boundaries (Mayachitra (2006)) or manually determining the center location of each muscle fiber (Klemencic et al. (1998)). These user interactions make analyzing a large number of images time-consuming and impractical.

In this paper we introduce CHARISMA (Cellular Histological Automated Robust Image Segmentation of Muscle Algorithm), a novel top-down approach for skeletal muscle cell segmentation in histology images. Our approach combines the efficiency of image processing with the adaptability of machine learning. In addition, we introduce a new clump splitting technique based on concavity analysis and shortest path detection. We also show that the improvement in segmentation performance translates to a better representation of cell property distributions.

We evaluate our technique and compare it to existing implementations on a challenging and extensive real-world dataset of histology images obtained from patients admitted to the ICU of the university hospital in Leuven, Belgium.

2. Related Work

Histological image analysis entails a large body of literature, as detailed by Loukas and Linney (2004); Wu et al. (2004). Here we refer to techniques that focus on cell detection in general, and muscle fibers in particular.

A well-known segmentation approach is intensity or color thresholding, where either a global or a local procedure can be employed, as in Francis et al. (2000). However, thresholding techniques by themselves rarely work well on complex images. Alternative general methods that were successfully applied to histological segmentation include perceptual grouping by Todman et al. (2008), region growing by Rogojanu et al. (2010); Reyes-Aldasoro et al. (2011) and fuzzy clustering by Bonnet et al. (2002). Advanced geometry-based methods exploit radial symmetries, convex hull properties, polygon approximations, distance properties and concavity analyses to identify and split cell clumps. They have been used to segment individual lymphoma cells in Kong et al. (2011) and to perform 3D cell cluster splitting in Indhumathi et al. (2011). Still, these techniques fail to accurately separate the complex cell clumps that typically arise in muscle tissue.

In this particular setting, more successful approaches exploit the fact that prominent boundaries between individual muscle fibers may provide more discriminative information. The work of Sertel et al. (2011) uses ridge detection for segmentation of ATPase stained muscle tissue. Similarly, edge-based active contour techniques have been used for individual muscle cell segmentation by Brox et al. (2006); Kim et al. (2007). However their clinical use is unclear as they have only been evaluated on very small datasets.

Recently, active contour algorithms based on a level set formulation have become popular for histological segmentation. Ali and Madabhushi (2012), pro-
pose a region-based active contour model that incorporates an energy function based on shape priors, image gradients and region statistics. The method is demonstrated to work well in segmenting noisy images. Qi et al. (2012) identify cells by detecting the geometric center using mean-shift-based seed detection. Segments containing multiple centers are split using a level set approach, by minimization of an energy function based on repulsion and competition. Fatak-dawala et al. (2010) segment lymphocyte images by initializing a geodesic active contour via expectation-maximization, level sets are employed for contour detection, and a heuristic concavity-based approach is used to split contours encompassing multiple objects. Though these methods perform well in their application domains, to our knowledge they have not been tested on images containing large and dense cell clumps characteristic of degraded skeletal muscle tissue. These complex clumps can lack deep enough concavities to allow for a level set split initialized by a contour enveloping the entire clump. Likewise, radial symmetry-based approaches would be limited in their ability to detect cell centers for contour initialization in these skeletal muscle clumps. Fatakdawala et al. (2010) specifically uses the single cell contour as fitness criterion, which limits a split to removing a single cell.

A separate line of work considers top-down, model-based techniques in which domain knowledge is encoded as a predefined model of the tissue structure of interest. Tutac et al. (2008) translates image features into a set of symbolic rules. These rules allow an algorithm to first detect single objects, i.e., nuclei and cells, and then evaluate their relationship. The work of de Vieilleville et al. (2009) models individual muscle fibers as deformable polygons and formulates segmentation as the internal energy minimization of the deformable model based on the minimum length polygon. Both works evaluate their method on histological breast tissue. A similar approach for skeletal muscle tissue is presented by Dryden et al. (2006). The segmentation problem is expressed as the maximization of the posterior probability based on the tessellation of the muscle tissue using Markov Chain Monte Carlo methods. These approaches differ from ours, as we do not assume any prior model of our domain.

Recent work on histological segmentation has focused on fully automated approaches employing machine learning techniques. Wang (2010) considers a Markov Random Field in a Bayesian formulation to segment cancerous structures by classifying individual pixels in H&E stained images. Yin et al. (2010) uses a bag of local Bayesian classifiers to determine which pixels in the image belong to cells. Zampirolli et al. (2010) starts by obtaining geometrical structures from the tissue using morphological techniques. Then a set of features is extracted from the segmented patches and used to train an SVM classifier to differentiate between epithelial and adipose cells. However, these methods have a restricted experimental setting and do not provide experimental evidence for skeletal muscle segmentation. The automated approach of Wang et al. (2012) is closely related to ours, and we will address the key differences it in detail throughout the text.

A different supervised learning technique for muscle fiber type classification is proposed by Karen et al. (2009). Their framework however requires significant
manual interaction from the user to perform the segmentation step.

To our knowledge, our work is the first to develop a robust fully automated system able to detect individual fibers in H&E stained skeletal muscle images.

3. Top-down Histological Segmentation

We propose a top-down approach to muscle cell segmentation, illustrated in Figure 2. We obtain the initial segments from the histology image by thresholding the saturation, which allows us to separate the cells from the background, and brightness, which helps to remove artifacts that arise as a result of the imaging process, such as dust spots or overlapping tissue. Most approaches commercially or freely available to researchers stop here and consider these segments as output for medical interpretation. However, such a result is insufficient for analysis of skeletal muscle tissue.

The initial segmentation returns image segments that can be classified as: a) an individual cell, b) a clump of cells, or c) remnant connective tissue (Figure 3). A 3-class SVM classifier was developed for this purpose using a set of training images with manually labeled cell segments. Segments classified as connective tissue are discarded, while those classified as clumps are split into constituent smaller image segments, until only individual cell segments remain. This overall approach allows for a fast, objective and accurate segmentation of skeletal muscle cell tissue.

Figure 2: Information flow in Charisma. Rectangles and diamonds represent algorithmic steps and data objects, respectively. The blue arrows show the steps taken during the model training phase, while the red ones during the classification/segmentation phase. The black arrows show initial segmentation steps. Note that the training step is executed only once.
4. Segment Classification

The classification serves two roles. First, it filters out the remaining unwanted tissue. Second, it distinguishes single cells from the clumps remaining after the initial thresholding step.

We use a Support Vector Machine (SVM), Boser and et al. (1992); Vapnik (1995) a very popular machine learning technique because of their good performance on noisy data and high-dimensional spaces. We use the implementation of Chang and Lin (2001) (available online at http://www.csie.ntu.edu.tw/~cjlin/libsvm/, accessed June 2013).

In order to provide our classifier with training examples, we extract a feature vector from each image segment obtained from the initial segmentation of a set of training images. Individual cells are approximately uniform in color, have a regular convex shape and relatively small variation in size. Cell clumps, on the other hand, can be arbitrarily large, contain internal empty areas as well as differently-colored regions between cells, exhibit concavities along their contours, and are generally less convex-shaped. Connective tissue is mostly sized and shaped very irregularly and has a lighter, less uniform coloration.

We use several geometric features to capture information about the size, regularity and shape of the segment. Some of these properties are straightforward (area, perimeter, convex area and perimeter, fitted ellipse major and minor axis lengths), and some are more complex (form factor, roundness, solidity and eccentricity (Zampirolli et al. (2010)). Texture features are derived from the gray level co-occurrence matrix (entropy, contrast, correlation, angular second moment and inverse difference moment), as defined in Cabrera (2005) and Clausi (2002). We also have own features derived from the bottleneck measure Wang et al. (2012), and luminosity-related features (Table 1). Our SVM classifier is trained to classify segment feature vectors consisting of these 27 features.
Feature Description

Bottleneck-related features
Characterize the two pixels on the contour that are the closest to each other relative to their arc length, as defined by Wang et al. Wang et al. (2012); we consider as features the smallest distance-arc length ratio, its corresponding direct distance and arc distance.

Radius vector variance
The magnitude variance of $n$ vectors from the segment centroid $C$ to random contour points $P_i$: $V = \frac{1}{n} \sum_{j=1}^{n} (r_j - \bar{r})^2$, where $\bar{r} = \frac{1}{n} \sum_{j=1}^{n} r_j$ and $r_i$ is the magnitude of $\overrightarrow{CP_i}$.

Relative segment luminosity
$\frac{l_{i,n}}{\sum_{i=0}^{n} l_i}$, $l_i$ is the mean segment luminosity, $l_i$ is the mean luminosity of segment $i$, $n$ is the number of remaining segments in the initial segmentation.

Segment luminosity relative to background
The mean luminosity value of the segment divided by that of the background.

Edge luminosity
Characterizes segment mean luminosity after Sobel edge detection and identifies segments that have internal color variation.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottleneck-related features</td>
<td>Characterize the two pixels on the contour that are the closest to each other relative to their arc length, as defined by Wang et al. Wang et al. (2012); we consider as features the smallest distance-arc length ratio, its corresponding direct distance and arc distance.</td>
</tr>
<tr>
<td>Radius vector variance</td>
<td>The magnitude variance of $n$ vectors from the segment centroid $C$ to random contour points $P_i$: $V = \frac{1}{n} \sum_{j=1}^{n} (r_j - \bar{r})^2$, where $\bar{r} = \frac{1}{n} \sum_{j=1}^{n} r_j$ and $r_i$ is the magnitude of $\overrightarrow{CP_i}$.</td>
</tr>
<tr>
<td>Relative segment luminosity</td>
<td>$\frac{l_{i,n}}{\sum_{i=0}^{n} l_i}$, $l_i$ is the mean segment luminosity, $l_i$ is the mean luminosity of segment $i$, $n$ is the number of remaining segments in the initial segmentation.</td>
</tr>
<tr>
<td>Segment luminosity relative to background</td>
<td>The mean luminosity value of the segment divided by that of the background.</td>
</tr>
<tr>
<td>Edge luminosity</td>
<td>Characterizes segment mean luminosity after Sobel edge detection and identifies segments that have internal color variation.</td>
</tr>
</tbody>
</table>

Table 1: Features used to characterize an image segment and their description.

5. Clump Splitting

The clump splitting method comprises four steps: concavity point generation, concavity pair selection, path selection and re-thresholding, each of which we discuss in turn.

5.1. Concavity point generation

Dividing a clump segment into two separate new segments means finding a good split path between the segments. For each clump segment, we want to find a path consisting of pixels $\{p_0, p_1, \ldots, p_n\}$ so that a) $p_{i-1}$ is a direct neighbor of $p_{i-2}$ and $p_i$ with $i = \{2 \ldots n\}$, and b) only $p_0$ and $p_n$ lie on the clump segment contour. Our aim is to find a path that follows as close as possible the natural boundaries that exist between the cells within the clump. Examples of such boundaries are illustrated in Figure 4.
Our initial goal is to detect good candidates for the end-points $p_0$ and $p_n$, for example the concave “wedges” along the clump contour that often indicate where two cells touch. We refer to the candidate end-points as concavity points. In this regard, our approach is similar to Kumar et al. (2006). However, we expand it by finding multiple concavity points per concave area, introducing a more flexible concavity point selection procedure and, later, a new split path routing method.

The candidate point generation procedure starts by identifying the different continuous concave parts of the clump contour. We denote the set of pixels on the contour as $P_{SC}$, the set of pixels on the convex hull as $P_{CH}$, and the set of pixels on the contour that are not on the convex hull as $P_{CC}$. Thus, $P_{CC} = P_{SC} \setminus P_{CH}$ is the set of pixels representing all concave sub-contours, from which the concavity points are a subset $P \subset P_{CC}$.

Typically, a concavity point is found where two cells touch and is marked by an abrupt change of the contour direction. For any point $p_i \in P_{CC}$ this change can be checked via the angle that the sub-contour forms in $p_i$, with $i$ indicating the $i^{th}$ pixel on the sub-contour. To this end, we examine multiple angles $\theta_{ij,i'}$ defined by contour points $(p_j,p_{j'})$ located at equal distances $d$ from the angle vertex $p_i$, such that $j = i - d$ and $j' = i + d$. $\Theta_i = \{\theta_{ij,i'}\}$ is the set of such angles with $d \in [d_{\text{min}}, d_{\text{max}}]$. The minimum distance $d_{\text{min}}$ cannot be zero or very small because angles with the immediate neighbors of $p_i$ are too noise-susceptible. Contour points beyond the maximum distance $d_{\text{max}}$ are no longer expected to be relevant for the concavity point $p_i$.

To get an accurate estimation of the angle for $p_i$, we consider the average of the angles in the set $\Theta_i$ (Figure 5a).

\[ \bar{\theta}_i = \frac{\sum_{\theta \in \Theta_i} \theta}{d_{\text{max}} - d_{\text{min}}} \]

The average angle $\bar{\theta}_i$ gives us an estimate of how sharply the contour bends at $p_i$. We impose a threshold $\bar{\theta}_{\text{max}} \in [0, \pi]$ on $\bar{\theta}_i$ to filter out shallow concavities unlikely to indicate a cell boundary.

We obtain the point direction using the same pairs of equidistant points $(p_{i-d}, p_{i+d})$. We define the direction in point $p_i$ in terms of the vectors $\overrightarrow{p_{i-d}p_i}$, each starting at the halfway point $p_h$ of the segment $p_{i-d}, p_{i+d}$, and ending in the candidate point $p_i$, for $d \in [d_{\text{min}}, d_{\text{max}}]$. These directions can be represented as $\overrightarrow{p_hp_i} = (p_{i-d}p_i + p_{i+d}p_i)/2$. The concavity point direction then is the vectorial sum given by the equation:

\[ \overrightarrow{p_i} = \sum_{d \in [d_{\text{min}}, d_{\text{max}}]} \frac{\overrightarrow{p_hp_i}}{\|p_hp_i\|} \]
(a) Angles $\Theta_i$ are averaged to calculate $\bar{\theta}_i$ for point $p_i$. The contour parts marked in green delineate the points used to evaluate $\bar{\theta}_i$. Equidistant pairs on the green parts are illustrated in blue at distances $d_1, d_2$ and $d_3$ from $p_i$.

(b) The direction vector of the concavity point $p_i$ is marked in green. It is obtained by averaging the normalized vectors starting from the pair’s segment midpoint to the concavity point.

Figure 5: Angle-based and direction-based concavity point generation.

Figure 6: Concavity points with outward directions (in red) are filtered out.

Figure 7: Two opposite overlapping concavity points.

Figure 8: Fake concavity point (blue) introduced for a correct split and real concavity points (green).

Figure 5b illustrates the estimation of the direction of a concavity point. An approximation of this direction heuristic might be formulated in terms of the first derivative of the curve at different sub-contour ‘resolutions’ (i.e., taking points increasingly further apart) and taking the average of the orthogonal vectors to all derivatives.

As with the concavity point angle, we impose a constraint on the direction of a concavity point $p_i$ such that it is directed towards the interior of the clump segment, (figure 6).

5.2. Concavity Point Pair Selection

The next step is to find the concavity point pair $(p_i, p_j)$ most likely to produce the best split line. Parameters used to score the pairs of points include the (arc-) distance between the points, the depth and the orientation of the concavities. These properties can be expressed in terms of the following fitness measures: a) the Euclidean distance $D$ between the points in the pair; b) the distance between the points following the contour along the shortest connecting path, denoted as the arc distance $D_{arc}$; c) the concavity-concavity alignment.
Algorithm 1 Point pair selection algorithm (concavity points $P$ of the clump)

1: $\text{bestScore} \leftarrow 0$, $\text{bestPair} \leftarrow \emptyset$
2: for $(p_i, p_j), p_i \in P, p_j \in P \cup FP, i \neq j$ do
3:  $\text{score} \leftarrow 0$
4:  $(D, D_{\text{arc}}, A_{cc}, A_{cl}, \bar{\theta}_i, \bar{\theta}_j) = \text{EstimateParams}(p_i, p_j)$
5:  if $D = 0$ then
6:     $\text{score} = \infty$
7:  else if $D > \alpha$ and $D_{\text{arc}} > \beta$ and $A_{cc} \leq \gamma$ and $A_{cl} \leq \delta$ then
8:     $\text{score} = \left[ (A_{cc}^4 + 1)(A_{cl}^4 + 1)(D^2 + 1)\sqrt{\bar{\theta}_i \bar{\theta}_j} \right]^{-1} \cdot \frac{D_{\text{arc}}}{D}$
9:  end if
10: if $\text{score} > \text{bestScore}$ and $\overrightarrow{p_i, p_j}$ is interior w.r.t the clump contour then
11:     $\text{bestScore} \leftarrow \text{score}$, $\text{bestPair} \leftarrow (p_i, p_j)$
12: end if
13: end for
14: return $\text{bestPair}$

$A_{cc}$: d) the concavity-line alignment $A_{cl}$; and e) the average angles $\bar{\theta}_i$ and $\bar{\theta}_j$ as defined in section 5.1. $A_{cc}$ and $A_{cl}$ are defined for concavity points $p_i$ and $p_j$ as:

$$A_{cc} = \pi - \cos^{-1}(\overrightarrow{p_i, p_j}),$$

$$A_{cl} = \max(\cos^{-1}(\overrightarrow{\hat{p}_i, \hat{p}_j}), \cos^{-1}(\overrightarrow{\hat{p}_j, -\hat{p}_i})),\)$$

where $\hat{p}_i, \hat{p}_j$ and $\overrightarrow{\hat{p}_i, \hat{p}_j}$ are unit vectors (for more details see Kumar et al. (2006)).

The concavity point pair selection procedure is presented as Algorithm 1. It expresses our goal to minimize $A_{cc}$, $A_{cl}$, $D$, and the concavity angles $\bar{\theta}_i, \bar{\theta}_j$, giving preference to well-aligned, proximate and sharp concavities. The fraction $D_{\text{arc}}/D$ expresses the relation between the pair’s distance and the contour size (bottleneck, see Wang et al. (2012)). The exponents of the different terms in the formula were experimentally determined in order to give proper weight to each point pair fitness criterion. For example, alignment plays a larger role than distance in determining fitness. The effect of the measures is illustrated in Figure 9.

Instead of using the evaluation formula for all points pairs, to save computation costs we first filter out the most unlikely pairs, by thresholding measures applied to $D, D_{\text{arc}}, A_{cc}$ and $A_{cl}$ (lines 5 and 7):

- Each pair with points at a distance $D$ larger than $\alpha$ pixels apart is discarded. $\alpha$ is set to be a few times the average cell radius, so no needlessly long paths are calculated.
- To prevent a split between a pair of contour points lying in the same concavity, we enforce a minimum arc distance $\beta$.
- We constrain alignment within a maximum deviation from perfect alignment $A_{cc} = A_{cl} = 0$: $A_{cc} \leq \gamma$ and $A_{cl} \leq \delta$. 


• $D = 0$ means two opposing concavities touch (as the $\beta$ constraint filtered out same-concavity pairs). This is always chosen as the best split (Figure 7).

![Figure 9: Examples of concavity point pairs evaluation. Each possible pair is represented by a colored line segment connecting the two concavity points. The quality of the scored pair is represented by hue: from red (worst) to dark blue (best); a), b), c) show that thin bottlenecks are preferred; d) shows that even for complex clumps the high-scoring pairs correspond to intuitively good splits. In this last example top-scoring pairs are too close together for the blue line to be visible.]

It is possible that only one end of the most desirable split line is detected as a concavity point. To handle cases like these, we introduce the notion of “fake concavity points”. These points, which are calculated for each concavity point $p_i$, are defined as the closest intersection of the point’s directional ray $\overrightarrow{p_i}$ with the opposite segment contour. (Figure 8). During the search for the best concavity point pair, each $p_i$ is also evaluated as a pair with its fake counterpart $p_i\cdot\text{fake}$. The only change in the evaluation is $\overline{\theta}_i\cdot\text{fake} = \pi$, representing a hypothetical concavity with a straight average angle.

Using these measures, the best-scoring pair with a connecting line completely inside the clump is selected as basis for creating the splitting path (lines 10 and 11).

5.3. Path Selection

Methods such as Kumar et al. (2006) use a straight line between two points on the contour to split the clump. Since this approach can be inaccurate for large clumps (Figure 9), we introduce a less naive, gradient-based, method. We explain how we can derive an end-to-end gradient path, and extend this method with the notion of waypoints.

**End-to-end.** We consider the clump pixel grid to be a directed graph $G(V, E)$, where a pixel $i$ is represented as a node $V_i \in V$. For every neighbor pixel $j$ we have an edge $E_{ij} \in E$ directed from $i$ to $j$. To every edge $E_{ij}$, we assign a weight based on the saturation value of $j$. More specifically, $W_{E_{ij}} = 1/(1 + |\text{sat}_{\text{peak}} - \text{sat}_{j}|)$, where $\text{sat}_{j}$ is the saturation of point $j$ and $\text{sat}_{\text{peak}}$ is the cell saturation peak of the saturation histogram. This function models our preference for the path to contain mostly non-cell or peripheral cell pixels.

The defined weighed graph $G$ is used as input for the A* algorithm (Hart et al. (1968)), which finds the shortest path with minimal weight $W_{E_{\text{opt}}}$ between
two vertices $V_s$ and $V_e$, corresponding to the selected pair of concavity points $p_s$ and $p_e$. A* requires a underestimating heuristic to approximate the expected path cost between any two vertices (here pixels) $i$ and $j$, which we formulate as $H_{ij}$:

$$H_{ij} = (|p_{i}.x - p_{j}.x| + |p_{i}.y - p_{j}.y|) \cdot W_{E,Q},$$

where $p_{i}.x$ and $p_{i}.y$ are the X and Y pixel coordinates of a point $p_i$; $W_{E,Q}$ represents an one-pixel step towards a pixel $Q$ with minimal cost, i.e. a saturation of 0 (or 255, if the saturation histogram peaks at a value below 128). This heuristic represents the Manhattan distance-length path between the pixels, such that every pixel in the path has the ideal saturation value (i.e. the most different from the typical cell saturation). This maximal saturation difference results in a minimal weight assigned to each edge in the path from $i$ to $j$, which guarantees the total estimated cost $H_{ij}$ will be an underestimation.

To exclude the low-cost path that simply traces around the clump, we exclude all vertices along the contour, except for the terminal vertices $V_s$ and $V_e$. A similar path detection procedure is used by Wang et al. (2012), with an edge operator as energy function. We believe our saturation-based method to be more appropriate here, as the proximity of the pixel color to the background color is more relevant.

**Waypoints.** Sometimes the selected concavity points are weakly aligned or quite far apart, making it difficult to find a good split path only by using the algorithm described above. We therefore introduce the notion of *waypoints*, which serve as intermediate points that guide the path routing process.

Waypoints are pixels forced to be included in the splitting path between two concavity points $V_s$ and $V_e$. To determine the pixels that are the most suitable waypoints, we examine the area around the concavity point $V_s$ generated by a radius $6 \times r_{cell}$. For each concavity point $p$, our goal is to select a pixel $c$ that has a high likelihood of belonging to an appropriate split path starting from $p$. We estimate this likelihood by examining i) the angle $\varphi_{p,c}$ between the vector $\vec{pc}$ and the concavity point direction $\vec{p}$ (Figure 10) and ii) the difference between the image cell saturation peak, $sat_{peak}$, and the saturation of $c$, $sat_c$. The angle component is expressed via an exponentially decaying function modeling the rapid likelihood decrease of $c$ being a waypoint as $\varphi_{p,c}$ increases:

$$s_1(c) = e^{-2\varphi_{p,c}},$$

The saturation component $s_2(c)$ is lower bounded by a minimum value in the saturation histogram for the entire image.

$$s_2(c) = \begin{cases} |sat_c - sat_{peak}|, & sat_{min} < sat_c \\ |sat_{min} - sat_{peak}|, & sat_c \leq sat_{min}. \end{cases}$$

The waypoint scoring procedure can be formally expressed as:

$$score(c) = s_1(c) \cdot s_2(c).$$
Figure 10: Point \(c\) is a candidate waypoint for concavity point \(p\), whose direction is indicated by the right-hand arrow and forms an angle \(\varphi_{p,c}\) with \(\overrightarrow{pc}\).

Figure 11: A three-cell clump (white contour) overlaid with waypoints scores around a concavity point (red dot). The quality of the waypoints is indicated by hue (worst orange, best blue). The highest-scoring waypoint is shown in white. For visualization purposes, the 10% lowest scoring pixels are left in grayscale. The directional term \(s_1(c)\) gives a high score to pixels lying close to the concavity direction (black line), while the saturation term \(s_2(c)\) gives a low score to cell pixels. The concavity point direction is the black arrow.

Figure 11 illustrates the application of this formula on a concavity point.

The number of waypoints included in the split line depends on the spatial configuration of the concavity and waypoints. More specifically, we examine if the line segments \(\overrightarrow{pa}\) and \(\overrightarrow{pb}\) intersect. Figure 12 enumerates the possible cases.

The splitting step requires several parameters together with their threshold values: \(\overline{\theta}_{max}, d_{min}, d_{max}\) for concavity point selection, and \(\alpha, \beta, \gamma, \delta\) for pair evaluation. They were all set by experimental validation on the training set.

5.4. Re-thresholding

Splitting the clump segment might result in a region of non-cell pixels in the newly created segment touching the contour. These non-cell parts are easily filtered out by re-applying our initial threshold to the segment (Figure 13).

6. Recursive Segmentation

The two segments resulting from a split are re-classified and processed in turn. We opt for this recursive approach to splitting multi-cell clumps for several reasons. First, there is a substantial reduction of the split search space: finding one single path bisecting the clump is faster than finding a full segmentation for all cells at once. This way, the complexity of computing individual paths is lowered sufficiently to allow for segmentation of large cell clumps. Second, the intermediary classification step acts as a filter. The segment class re-evaluation
Figure 12: Examples of waypoints. Concavity points are marked in green, waypoints in blue, adjusted points in red. The concave contour is a continuous line, the path determined by waypoints is a dotted line, and a possible splitting path is gray; a) the path passes through two waypoints; b) the waypoints are replaced by the concavity-waypoint lines’ intersection point $I$; c) the two waypoints are close enough to allow replacement by their middle point $M$; d) the waypoint $c_0$ is ignored as it falls outside the original contour; e) both waypoints are ignored.

Figure 13: Re-thresholding preformed on a cell clump: an initial split line (left) and its re-thresolded split line (right).

may avoid unnecessary splits if, for example, connective tissue is encountered inside the clump (Figure 14).

Figure 14: The segmented clump is separated into discarded connective tissue (blue) and a cell (red) coinciding with the ground truth (black).
7. Experiments

We perform several types of experiments to evaluate different aspects of CHARISMA, relating to SVM performance, segmentation performance and the accuracy of derived cell properties.

7.1. Parameter values

Table 2 lists all parameter values that need to be set. We use these experimentally verified values in all our experiments.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>0.05</td>
<td>Brightness threshold</td>
</tr>
<tr>
<td>$\bar{\theta}_{max}$</td>
<td>$\frac{3\pi}{4}$</td>
<td>Maximum average concavity angle to be considered for a split path</td>
</tr>
<tr>
<td>$d_{min}$</td>
<td>3px</td>
<td>Minimum neighboring contour point distance for $\bar{\theta}$ calculation</td>
</tr>
<tr>
<td>$d_{max}$</td>
<td>7px</td>
<td>Maximum neighboring contour point distance for $\bar{\theta}$ calculation</td>
</tr>
<tr>
<td>$A_{cell}$</td>
<td>35px</td>
<td>average cell area</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$8 \cdot \bar{r}_{cell} \text{px}$</td>
<td>Max. point pair distance</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$0.01 \cdot 2\pi \cdot \bar{r}_{cell} \text{px}$</td>
<td>Min. point pair arc distance</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1.83</td>
<td>Max. Concavity-concavity misalignment</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1.22</td>
<td>Max. Concavity-line misalignment</td>
</tr>
<tr>
<td>$\bar{\theta}_{i \text{ fake}}$</td>
<td>$\pi$</td>
<td>Preset fake concavity point average angle</td>
</tr>
</tbody>
</table>

Table 2: Empirically established parameter values. $\bar{r}_{cell}$ is the average cell radius approximated using the average cell area $A_{cell}$ in our training set: $\bar{r}_{cell} = \sqrt{\frac{A_{cell}}{\pi}}$. 
7.2. Dataset

For experimental evaluation we consider a dataset of 111 images of skeletal muscle tissue (right musculus rectus abdominis and musculus vastus lateralis of the quadriceps femoris) from biopsy samples taken immediately after death and in vivo from surgical and medical ICU patients, patient controls and healthy volunteers admitted to the university hospital of Leuven, Belgium as described in Derde et al. (2012); Vanhorebeek et al. (2011). Written informed consent is obtained from the patients or next of kin and from the healthy volunteers. A medical expert manually traces the cells in each image, providing the ground truth. A subset of 11 images is used for training the classifier, the remaining 100 images (with 33402 ground truth segments) are used for testing. The histology images, having a resolution of 2592 by 1944 pixels, are taken at 5-fold magnification using a Leica DM3000 microscope. They characterize tissue with various degrees of skeletal muscle cell degradation, representing a challenging segmentation task (Figure 1).

7.3. SVM Classifier Performance

We report the performance of the SVM classifier separately in order to assess the behavior of the classification step.

A segment correctly classified as belonging (not belonging) to the class of interest is denoted true positive $TP$ (true negative $TN$). Likewise, a segment incorrectly classified as belonging (not belonging) to the class of interest is denoted false positive $FP$ (false negative $FN$). We evaluate the classifier in terms of accuracy (A), the true positive rate ($TPR$), the false positive rate ($FPR$) and the area under the ROC curve (AUC), standard in supervised machine learning settings. To train the SVM we use a dataset of 2127 manually labeled segments from 11 images. Training was done on segments resulting from the initial segmentation step to ensure similarity between training and testing instances. We employ a polynomial kernel, normalized features and perform 10-fold cross-validation on the training data to select the SVM parameters. We obtain $A = 93.3\%$, $TPR = 93.4\%$, $FPR = 5.2\%$, and $AUC = 0.94$. Some misclassifications correspond to segments that are difficult to label even for human experts (Figure 15).

(a) Possible connective tissue classified as cell  
(b) Clump classified as connective tissue  
(c) Connective tissue classified as clump

Figure 15: Examples of (potential) misclassification.
7.4. Comparison to Related Work

In this experiment, we compare CHARISMA with a recent method proposed in the literature, Wang et al. (2012), and CellProfiler, a widely-used and freely available tool (Carpenter et al. (2006)). We chose these because of the conceptual similarity of the first and public availability of the latter.

7.4.1. Wang et al.

Instead of an explicit initial segmentation step, this method starts from a collection of segments that it then classifies as either cells or clumps. We use CHARISMA’s initial segmentation to provide the required collection of initial segments. The classifier in Wang’s method is not trained to recognize connective tissue or artifacts. In order to assess the effect of this limiting factor we train two classifiers using their method: one trained only on cell and clump segments (2-class), as done in their paper, and one also trained on connective tissue and noise segments (3-class).

7.4.2. CellProfiler

We create an image processing pipeline with several empirically optimized parameter values, as would happen in a typical usage scenario. A thresholding of the green image channel is performed with the threshold value determined by a mixture of Gaussians algorithm. Second, muscle cells are identified with the help of user-supplied fixed parameters such as size constraints, an estimate of relative tissue coverage of the image, and others. These parameters remain fixed for the entire image batch.

7.5. Evaluation Measures for Final Segmentation Performance

The overall segmentation task is in essence a skeletal muscle cell detection problem, therefore we adopt evaluation measures used in information retrieval: we evaluate segmentation performance in terms of the number of true and false detections of cells in the test dataset.

We compare the area of the detected cell $D$ to the area of the ground truth cell $G$, using a minimum overlap threshold $T$. If $\frac{\text{area}(D \cap G)}{\text{area}(D \cup G)} > T$, then $D$ is a true positive detection $TP$, otherwise the detection is a false positive $FP$. Ground truth cells that are not covered by detected cells are considered false negatives $FN$. We evaluate the following performance measures: precision $P = TP/(TP + FP)$, recall $R = TP/(TP + FN)$ and $F1$-score. Precision is the proportion of correctly detected cells to all detected cells and Recall is the proportion of correctly detected cells to all ground truth cells. The $F1$-score measures detection accuracy and is the harmonic mean of precision and recall, $F1 = (2PR)/(P + R)$. These performance measures range in value from 0 (worst) to 1 (best).

To assess the impact of the required overlap on detection performance, we vary the threshold $T$ in the interval $[0.5, 1]$. 

17
7.6. Statistics of Cell Properties

The need for histological cell segmentation is motivated by the fact that the collection of individual cell properties has a great significance in biomedical research. We evaluate statistics for several properties of interest, such as cell area, major axis length and radius vector variance (as defined in section 4). To illustrate how they are influenced by the choice of segmentation method, we compare the distributions of cell properties for each method with those of the ground truth cells.

7.7. Implementation

Charisma is implemented as an extension to ImageJ (Abramoff et al. (2004)) using Java. For the SVM implementation and evaluation, we use the Weka toolkit Frank et al. (2004), all other statistical analyses were performed using R (R Core Team (2012)). All experiments were performed on a machine with an Intel i7 2.40 GHz processor.

8. Results

The results of the final segmentation for all methods on all test segments are shown in Table 3 and as performance distributions over the 100 test images in Figure 16. Unless specified otherwise, reported results are for an overlap threshold $T = 0.75$.

Charisma outperforms the other methods on all evaluation measures considered. Regarding performance distributions the differences of medians with respect to all other methods are statistically significant with $p$-values $< 0.0001$ for precision, recall and F1-score. It is noteworthy that the minimum F1-score and recall of Charisma is above the 25th percentiles of all remaining methods. Similarly the minimum precision of charisma is above the 25th percentile of all methods except for Wang (3-class). The differences in medians between performance distributions between Wang (3-class) and Wang (2-class) are statistically significant.
Table 3: Segmentation performance of Charisma compared to Wang et al. (2012) and CellProfiler (Carpenter et al. 2006) Each absolute number is also shown as a percentage of the true number of cell segments (33402). These values are for $T = 0.75$. Precision, Recall and F1 are calculated for the full set of respective segments, collected over all images.

significant for all measures, though this difference is less pronounced for recall. Cellprofiler has the worst performance with distribution medians significantly lower than all other methods on all measures with $p$-values $< 0.0001$.

Charisma has higher F1-scores than all the other methods on the vast majority of test images. A qualitative comparison of segmentation performance between methods is presented in Figure 19. With few exceptions, the cells detected by Charisma approximate the ground truth very well. Also noteworthy is the fact that Charisma is the only method that correctly discards irrelevant tissue instead of falsely detecting it as cells. An example of a fully segmented image is shown in Figure 20.

Figure 17 shows the effect on detection performance when varying the overlap threshold $T$ in the $[0.5, 1]$ interval. Charisma scores higher that all other methods for the entire $T$ range. A similar behavior is observed for precision and recall. Charisma also has a slower performance decline than the other methods as the threshold increases. For completeness, Figure 17 evaluates performance up to $T = 1$, but at this threshold value all methods have virtually no true positive results as this would require a pixel-perfect match with the ground
Figure 18: Kernel density estimations of clinically relevant properties: area, radius vector variance and major axis length measured for ground truth and detected cell segments at $T = 0.75$

<table>
<thead>
<tr>
<th>Processing time (minutes)</th>
<th>Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t &lt; 5$</td>
<td>60</td>
</tr>
<tr>
<td>$5 &lt; t &lt; 10$</td>
<td>16</td>
</tr>
<tr>
<td>$10 &lt; t &lt; 15$</td>
<td>12</td>
</tr>
<tr>
<td>$t &gt; 15$</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 4: Processing time per image used in the experiment

...truth.

The measured distributions of clinically relevant cell properties area, major axis length and radius vector variance are shown as kernel density estimations in Figure 18. Of all the methods CHARISMA delivers the best approximation of the ground truth distributions, and as such provides a far better estimate of clinically relevant cell properties.

The processing time distribution for the experiment’s images can be seen in Table 4. Our algorithm took some minutes to process a typical image, resulting in a runtime of several hours for the entire experimental set.

9. Discussion

To the best of our knowledge, CHARISMA is the most complete automatic solution for individual skeletal muscle cell segmentation, capable of dealing with the pathological degradation of muscle tissue observed in biopsies from the critically ill.

CHARISMA combines image pre-processing with a classifier using a comprehensive feature set. This allows for the encoding of in-depth skeletal muscle domain knowledge through machine learning. It also makes use of novel clump splitting techniques balancing flexibility with domain knowledge, while its recursive segmentation approach allows for robust evaluation of large clumps.

The importance of incorporating a third category in the classifier to encode non-cell tissue can be seen most clearly in the improvement of precision
Figure 19: Qualitative comparison of the different detection methods evaluated with respect to ground truth cells for various image sections. From left to right, the images show the ground truth, CellProfiler, the Wang method (2-class) and CHARISMA. The ground truth cells are outlined in black, detected cells in green and non-cell segments in blue.
and F1-score from Wang(2-class) to Wang(3-class). The further improvement of Charisma over Wang(3-class) is partially due to the use of the extensive feature set for its classifier. Shape and color-based features are found to be discriminative, though texture features contribute relatively little. The second row of Figure 19 clearly shows the inability to deal with non-cell tissue of a two-class approach, such as Wang(2-class) or of the general-purpose approach of CellProfiler.

Another important aspect of Charisma is its complex but robust clump splitting methodology. Simple approaches are not sufficient when large clumps of several cells are encountered (Figure 19). CellProfiler’s general-purpose segmentation solution is not able to properly handle complex cell clumps, while Wang’s method often fails by relying solely on the bottleneck measure, which does not scale well with the number of concavities. Charisma overcomes these limitations by using a comprehensive set of measures to evaluate a split point pair, creating robustness over a diverse set of cases. These features include, besides the bottleneck measure, the concavity-concavity and concavity-line alignments, the Euclidean and arc distances and the ‘sharpness’ of the concavity angle.

Several other helpful concepts have been introduced. Fake concavity points enable Charisma to split clumps that have less pronounced concavities. The
inclusion of waypoints efficiently constrains the search for split paths to valid areas, thereby resulting in a further search space reduction. Re-thresholding is crucial in segmenting larger clumps, as it prevents large interior empty parts from being subsumed into the contour of clumps in intermediate splitting stages, which would lead to accumulating segmentation errors. All these additions were found to either provide a decrease in execution time or an increase in accuracy. Their application here results in split lines that are an accurate representation of the true cell boundaries (Figure 19).

We have introduced several concavity point selection thresholds (in Algorithm 1) that heavily reduce the point pair selection problem search space. Most notably, the neighborhood criterion reduces the search from a quadratic to a linear problem. This is absolutely necessary for these segment contours, often containing hundreds of candidate points.

It is important to mention that there is an inherent uncertainty to the segmentation problem (Figure 15). The differences between CHARISMA and the ground truth are best interpreted as a lower bound on F1, precision and recall: while our experiments show how accurately we can approximate a valid segmentation, there might exist an equally valid one that is closer to our result. To a certain extent we account for this by using an overlap threshold \( T < 1 \).

In Figure 18, CHARISMA is shown to make the best approximation to several clinically important morphological characteristics of tissue structure which provide crucial information to evaluate the clinical state of the patient Dubowitz and Sewry (2007). While these are only a subset of relevant statistics that can be collected, they are easily accessible as they coincide with features used by the SVM classifier. CHARISMA detects a small secondary peak to the right of the main distribution peak for area and major axis length, which are absent in the ground truth and the other methods. This can be explained on one hand by a slight tendency of CHARISMA to misclassify clumps consisting of two cells as single cells. On the other hand it is an indication of the actual presence of elongated cells in the image which were excluded from the ground truth by the medical experts. These elongated cells are observed in parts of the biopsy that were cut in the longitudinal direction of the muscle fiber.

The methodology presented here was specifically developed for high performance in skeletal muscle segmentation, most notably by making use of features specific to this tissue type and of heuristics designed to resolve the type of cell clumps encountered in this domain. Although beyond the scope of the study presented here, interesting future work is the extension of general methods such as active contours and level sets, which have been shown to perform well across tissues, to deal with the particularities of segmenting skeletal muscle fibers.

On the practical level, an important property of CHARISMA is its large degree of automation. Users only need to label a selected set of segments for training purposes, after which arbitrarily large batches of images can be segmented without further parameter tuning.
10. Conclusion

This paper introduces Charisma, a new, fully automatic method to segment and detect individual cells in H&E stained histology skeletal muscle cell images. Charisma is able to detect and remove image artifacts and connective tissue and to correctly identify muscle fibers in challenging skeletal muscle images containing large and complex cell clumps.

As part of the framework, we introduce a multi-class classifier component robust enough to reliably separate cells, clumps and connective tissue. We propose a novel method to segment very large clumps via concavity analysis and the notions of fake concavity points, waypoints and re-thresholding.

Charisma is a fully automatic solution as it infers all its required parameters from the training data. To the best of our knowledge, we are the first to provide a fully integrated solution to this problem setting, and to validate our methodology on a large dataset of real-world images of biopsies from ICU patients. The results indicate that Charisma achieves an accurate segmentation on such difficult data, both qualitatively and quantitatively, compared to manual human segmentation. This is a crucial step towards fully automatic and accurate analysis of patient pathology based on tissue morphology.


