Published in final edited form as:

IEEE/ACM Trans Comput Biol Bioinform. 2014 ; 11(4): 714–726. doi:10.1109/TCBB.2013.151.

Automatic Myonuclear Detection in Isolated Single Muscle Fibers Using Robust Ellipse Fitting and Sparse Representation

Hai Su,

Division of Biomedical Informatics, Department of Biostatistics, Department of Computer Science, University of Kentucky, Rooms 230 A, B, C, Multidisciplinary Science Building, 725 Rose Street, Lexington, KY 40506

Fuyong Xing,

Division of Biomedical Informatics, Department of Biostatistics, Department of Computer Science, University of Kentucky, Rooms 230 A, B, C, Multidisciplinary Science Building, 725 Rose Street, Lexington, KY 40506

Jonah D. Lee,

College of Health Sciences, University of Kentucky, Room 105B Wethington Building, 900 South Limestone Street, Lexington, KY 40506

Charlotte A. Peterson, and

College of Health Sciences, University of Kentucky, Room 105B Wethington Building, 900 South Limestone Street, Lexington, KY 40506

Lin Yang

Division of Biomedical Informatics, Department of Biostatistics, Department of Computer Science, University of Kentucky, Rooms 230 A, B, C, Multidisciplinary Science Building, 725 Rose Street, Lexington, KY 40506

Abstract

Accurate and robust detection of myonuclei in isolated single muscle fibers is required to calculate myonuclear domain size. However, this task is challenging because: 1) shape and size variations of the nuclei, 2) overlapping nuclear clumps, and 3) multiple z-stack images with out-of-focus regions. In this paper, we have proposed a novel automatic detection algorithm to robustly quantify myonuclei in isolated single skeletal muscle fibers. The original z-stack images are first converted into one all-in-focus image using multi-focus image fusion. A sufficient number of ellipse fitting hypotheses are then generated from them myonuclei contour segments using heteroscedastic errors-invariables (HEIV) regression. A set of representative training samples and a set of discriminative features are selected by a two-stage sparse model. The selected samples with representative features are utilized to train a classifier to select the best candidates. A modified inner geodesic distance based mean-shift clustering algorithm is used to produce the final nuclei detection results. The proposed method was extensively tested using 42 sets of z-stack images containing over 1,500 myonuclei. The method demonstrates excellent results that are better than current state-of-the-art approaches.

Index Terms

Robust ellipse fitting; muscle; segmentation; sparse optimization

I. Introduction

Cell and organelle counting are fundamental measures in biological research. In the field of muscle biology, the number of nuclei is of particular interest given that muscle is organized as multinucleated myofibers, with each nucleus supporting a domain of cytoplasm. The ratio of myonuclei to cytoplasmic volume of the myofiber is termed the myonuclear domain. Contrary to dogma, recent evidence shows that the myonuclear domain is very flexible, changing during muscle growth, atrophy, and regeneration [1], [2]. The addition and subtraction of myonuclei to the existing myofiber presumably is the basis of skeletal muscle hypertrophy and atrophy. For example, myonuclei accretion usually increases during muscle hypertrophy, presumably through the fusion of muscle stem cells contributing myonuclei to the existing myofiber. Thus, accurate and efficient quantification of the number of myonuclei per fiber relative to fiber size, is an important factor to evaluate muscle adaptation, as well as its pathological conditions. Currently, manual counting is extremely time consuming and labor intensive. Computer assisted microscopic image analysis tools are better alternatives compared with manual quantifications with respect to efficiency.

There are many existing approaches for automatic cell or nuclear counting. The distance transform based method is commonly used to detect seeds in clustered objects. However, it does not work well for the tightly clustered or overlapping objects in which false positive error becomes nonnegligible. Some algorithms improved the distance transform method by combining geometric and intensity information to provide more accurate detection [3], [4]. An algorithm [5] was proposed to filter out the false seeds by minimizing the false detections based on mutual proximity. Watershed-based segmentation is a widely used method for splitting touching cells [6]. However, it suffers from over-segmentation. Various merging mechanisms based on hierarchical tree searching algorithm [7], anatomic characteristics of nuclei analysis [8], and topology dependence constraints [9] have been proposed. A cell segmentation algorithm using watershed-based cell splitting with photometric and shape information was proposed in [10]. A general segmentation framework for fluorescent images was proposed in [11]. The algorithm utilized both the image appearance information from both DNA and Actin channels. The intercellular interactions and the gradient information inside the cells were formulated as an energy functional, and minimized by a multiphase level set propagation. Watershed algorithms incorporating markers indicating object prior information were investigated in [12], [13], [14]. In [15], cell markers (used for marker controlled watershed) were detected by localizing the centers of the gravities in the cell clumps using kernel-based iterative radial voting. Jung and Kim [16] developed a modified watershed algorithm using H-minima transform that integrates the shape information of nuclei. The H-value is derived by evaluating a shape distortion function to measure the fitting accuracy between the ellipse and nuclear contour. In [17], an adaptive H-minima transform is proposed to detect cell markers (seeds) within a connected component obtained from inner distance transform.

Object detection and segmentation can also be formulated into a graph cut problem [18], where pixels are modeled as the nodes. A graph cut segmentation algorithm minimizing an energy function was proposed in [19]. In [20], seed detections are implemented by dissecting the weighted graph using normalized cuts [21]. Some other graph based methods can be found in [14], [22], [23], [24], [25], [26], [27], [28].

A ridge detection based muscle fiber segmentation algorithm was proposed in [29]. Radial voting based methods [30], [31], [32] were proposed to locate the center of cells with a major assumption on round shaped cells, which does not hold in our case because myonuclei in muscle fibers contain both elongated ellipses and round circles. When voting based methods [29], [30], [31], [32] are applied to the cells with elliptical shapes, they tend to create false positive detections. Nuclei detection and segmentation methods based on Laplacian-of-Gaussian (LoG) with adaptive scale selection were reported in [33], [34]. Similarly, their method works better for touching cells with round shapes. Concave point detection based algorithm are proposed in [14], [35], however, it is often difficult to find a general rule for shape decomposition. Gradient vector diffusion [36] is proposed to detect blob-like objects in microscopic zebrafish images. The method is robust to noise, and can achieve accurate and consistent performance. In this paper, we have developed a novel algorithm for automatic myonuclear detection in isolated fixed skeletal muscle fibers stained with DAPI to visualize nuclei. The original z-stack images are presented in the left of Fig. 1. It can be observed that there exist the following challenges: 1) The myonuclei tend to exhibit a variety of circularity ratios, ranging from elongated ellipses to round circles. The round shape assumption therefore does not hold. 2) Some myonuclei have inhomogeneous staining intensities due to DAPI intercalating preferentially into heterochromatin, which creates some bright spots irregularly distributed in the myonuclei. This inhomogeneity tends to fail many current segmentation algorithms that require homogeneous intensity distribution of the object. In this paper, we have proposed an algorithm that can handle all these challenges. The contributions of this paper are:

- A novel robust ellipse fitting based on heteroscedastic error-in-variables (HEIV) regression algorithm is applied to generate a sufficient number of ellipse fitting hypotheses based on the concave points and contour segments.
- A novel two stage sparse optimization model is proposed for robust sample and feature selection.
- A modified mean-shift clustering algorithm based on geometric inner-distance is used to merge the seeds and generate the final myonuclear detection results.

The remainder of this paper is organized as follows. The myonuclear detection algorithm is explained in detail in Section 2. The experimental results are described in Section 3. Finally, Section 4 concludes the paper.

II. Method

Our algorithm consists of four steps: preprocessing, robust ellipse fitting, and refinement, and mean-shift clustering based on geodesic inner distance. The algorithm flowchart is shown in Fig. 2. In the preprocessing step, a multi-focus image fusion technique is applied

to derive an all-in-focus image from the original z-stack images, as shown in Fig. 1. Because there exist touching myonuclei, a concave point detection method is used to detect the concave points after thresholding using Otsu's method [37], contour extraction and smoothing. A sufficient number of hypotheses of potential fitting ellipses are calculated with heteroscedastic error-in-variable regression. In the ellipse refinement stage, two stage sparse optimization is applied to find the representative training samples and discriminative cellular features. Based on these selected sparse samples and their corresponding representative features, a support vector machine (SVM) classifier is trained to identify the best ellipse that represents the myonuclei. Finally, a mean-shift clustering algorithm based on geodesic inner distance is used to merge multiple ellipse candidates to produce the final myonuclei detection results.

2.1 Preprocessing

2.1.1 Multifocus Image Fusion—In z-stack imaging, the microscope collects lights from both the in-focus plane and the out-of-focus planes. Therefore, each image consists of in-focus objects as well as out-offocus objects. Motivated by [38], we have proposed a modified image fusion algorithm to compose a smooth multifocus image from the whole z-stack images. Let $i = 1, 2, \dots, N$ denote the z-stack images, the magnitude of gradient of each image is calculated as

$$M_i(x, y) = \sqrt{I_{X(i)}^2(x, y) + I_{Y(i)}^2(x, y)}, \quad (1)$$

where $I_{X(i)}^2(x, y)$ and $I_{Y(i)}^2(x, y)$ are the horizontal and vertical gradients of the image i , respectively. Different from [38] which picks the pixel $I_i(x, y) = \operatorname{argmax}_i M_i(x, y)$ as the synthesized image intensity $I_s(x, y)$, a linear combination of all the pixel values in the z-stack is used to calculate $I_s(x, y)$

$$g_i(x, y) = \frac{M_i(x, y)^k}{\sum_{i=1}^N M_i(x, y)^k}, \quad k=3, \quad (2)$$

$$I_s(x, y) = \sum_{i=1}^N g_i(x, y) I_i(x, y). \quad (3)$$

Our image fusion algorithm can provide the final composed image with smooth boundaries, which greatly increases the robustness of the subsequent automatic detection procedure.

2.1.2 Contour Smoothing and Concave Point Detection—Let $I_{sc}(x, y)$ represent the binary image containing the contours of the objects in the synthesized image $I_s(x, y)$. Before applying the concave point detection algorithm to $I_{sc}(x, y)$, a contour smoothing algorithm [39] is applied to increase the robustness for concave point detection. The contour of each connected component is represented by Freeman chain code with eight directions, elliptical Fourier transform [40] is then applied to smooth the contour by keeping the first m ($m=20$ in our case) Fourier coefficients.

A concave point detection algorithm [14] is applied on each point of the contour. As shown in Fig. 3, we first calculate an angle for each concave point using

$$\alpha = \arccos \frac{a^2 + b^2 - c^2}{2ab}. \quad (4)$$

The concave points with $\alpha < 165^\circ$ degrees and the midpoint of c outside the contour (Fig. 3b) are kept as candidates, and are further processed to suppress the local non-maxima points.

2.2 Ellipse Fitting and Refinement

2.2.1 Ellipse Fitting Using Heteroscedastic Errors-invariable (HEIV)

Regression—Heteroscedastic errors-in-variables regression is used to generate the robust fitting ellipses. Compared with some other traditional methods, HEIV has a weaker dependence on initialization and a faster convergence. Direct least-squares (DLS) method [41] is biased when the input data points are a short low-curvature segment of a whole ellipse as shown in Fig. 4. Geometric distance minimized with Levenberg-Marquardt (LM) [42] is sensitive to the initialization, and the initial value could be biased if it is obtained by DLS method [43]. We also include fitting result of Taubin's method [44] in Fig. 4 for comparison purpose. In our algorithm, an ellipse is modeled by its linear form

$$\varphi(x_{io}, \alpha, \theta) = \alpha + z(x_{io})^T \theta = 0, \quad (5)$$

where α and the five dimensional vector θ are the ellipse parameters, x_{io} is the noise-free data, and

$$z(x_{io}) = [x_{1io}, x_{2io}, x_{1io}x_{2io}, x_{1io}^2, x_{2io}^2], \quad (6)$$

where $z(x_{io})$ is used to represent the five dimensional carrier dependent on the data. The true data is corrupted by additive noise

$$x_i = x_{io} + \delta_i, \quad \delta_i \sim GI(0, \sigma^2 C_{x_i}), \quad (7)$$

where $GI(\mu, C)$ represents Gaussian distribution with mean μ and covariance C . Because the noise corrupting the true data x_{io} appears in the carrier vector, the carriers are not independent from each other. HEIV fitting algorithm employs the errors-invariables (EIV) model, and applies the HEIV algorithm to iteratively compute the maximum likelihood estimators in the linear model by solving

$$[\hat{\alpha}, \hat{\theta}, \hat{x}_i] = \arg \min_{\alpha, \theta, x_{io}} \sum_{i=1}^n (x_i - x_{io})^T C_{x_i}^{-1} (x_i - x_{io}), \quad (8)$$

subject to constraint (5) and

$$4\theta_4\theta_5 - \theta_3^2 = 1. \quad (9)$$

The HEIV algorithm exploits the dependency between α and θ , and iteratively computes the solution by solving a generalized eigenvalue problem

$$S(\hat{\theta})\hat{\theta} = \lambda C(\hat{\theta})\hat{\theta}, \quad (10)$$

where $S(\hat{\theta})$ and $C(\hat{\theta})$ are weighted scatter matrix and weighted covariance matrix, respectively. In each iteration a new estimate $\hat{\theta}$ is calculated by finding the eigenvector corresponding to the smallest eigenvalue, and a new α is computed from $\hat{\theta}$. The new parameter estimators are used to update the new estimates of the true data α , the noise variance $\hat{\sigma}^2$, and the moments of the carrier noise μ_{z_i} and C_{z_i} . The iteration stops when the smallest eigenvalue approaches 1. This algorithm works well for both homoscedastic and heteroscedastic noises. The estimators obtained are unbiased. The proof of its unbiasedness can be found in [45].

In order to segment the touching myonuclei, robust ellipse fitting based on HEIV regression model is performed on the contour pixels. Assume there are c concave points for a contour, the original object contour is intersected by the concave points into c segments. We fit candidate ellipses using the different combinations of the contour segments. Thus, there are $\binom{c}{k}$ segment combinations if we choose k segments to fit one ellipse. We empirically set $k = \{1, 2, 3\}$ to make sure that a sufficient number of ellipses using HEIV regression are generated as input to the subsequent refinement procedure. One robust ellipse fitting example using HEIV regression model is shown in Fig. 5.

2.2.2 Ellipse Feature Extraction—Given the ellipse fitting result calculated from Section 2.2.1, the next step is to generate representative image features. The following geometric features, summarized in Table 1, are defined as:

Set 1: The most intuitive geometric feature to evaluate the accuracy of ellipse fitting is to measure whether or not the fitting ellipse matches with the boundaries of myonuclei. A set of morphological ratios are defined for this measurement. Given the fitting ellipse and the contour of a myonucleus, the overlapping ratios are calculated based on the overlapping areas of the fitting ellipse and the contour. Let $r_{aoc} = \frac{A_o}{A_c}$ and $r_{aoe} = \frac{A_o}{A_e}$, where A_o denotes the overlapping area, and A_c and A_e denote the areas of the contour of myonucleus and the fitting ellipse, respectively. In addition, the ratio $r_{oace} = \frac{A_o}{A_c + A_e}$ is also calculated.

Considering robustness, the pixel-wise overlapping ratios are calculated. Let p_o denote the overlapping pixels between a fitting ellipse and the object, and p_c be the number of pixels of the contour. Then, we define contour pixel level overlapping ratio $r_{oc} = \frac{p_o}{p_c}$. Based on r_{oc} , we design an iterative procedure to assign a match-quality score to each fitting ellipse. In each iteration, we count the contour pixels overlapping with an ellipse, then sort the ellipses with respect to the number of their overlapping contour pixels. At the end of each iteration, the

contour pixels that overlap with the highest ranked ellipse are removed, and roc is updated for each remaining ellipse. The iteration stops when there is a small fraction (decided by a threshold) of contour pixels left. Note that in each iteration, the ellipse that has the most overlapping pixels with object contour pixels will be assigned the highest rank. The match-quality of an ellipse is defined as the ranks of this fitting ellipse calculated during each iteration. To improve efficiency, we only record the ranks of an ellipse in the first three iterations.

Set 2: The myonuclei are objects which have certainly biological meaningful areas. Therefore, the area of the ellipse A_e , the axis ratio r_{axis} , and the perimeter p_c and area A_c of the myonuclei contour are also considered as potential geometric features for classification.

Set 3: Concave point depth is a feature designed to distinguish the ellipses shown in Figs. 5b and 5e. This is based on the observation that an accurate ellipse fitting should not have concave points deeply inside the ellipse. Considering this, for a given ellipse $E_i, (i = 1, \dots, N)$ and a set of concave points $c_j, (j = 1, \dots, c)$, concave point depth d_i is defined as the sum of squares of the Euclidean distances from the concave points to ellipse E_i :

$$d_i = \sum_j dist^2(c_j, E_i), j \in \{j: c_j \text{ is in ellipse } E_i\}, \quad (11)$$

Where $dist(\cdot)$ denotes the Euclidean distance from c_j to E_i . Because the center of an accurate fitting ellipse should not locate near the boundary, the distance between the center of the fitting ellipse and object boundary d_{ecc} is also calculated. This feature can help to remove suboptimal fitting ellipses whose centers are close to the object boundaries as shown in Fig. 6.

Set 4: The irregularity of the boundaries are defined as $r_{irg} = \frac{n_c}{p_c}$, where p_c denotes the perimeter of a myonucleus contour, and n_c represents the number of concave points detected. Two examples are shown in Fig. 7.

Set 5: A set of statistical features are calculated to capture the relationships among the ellipses generated from the same contour. Assume N ellipse fitting candidates are generated from the segments of an object contour, and $f_i, i = (1, 2, \dots, N)$ represents one specific feature calculated for the i th ellipse E_i :

$$f_{1i} = f_i - \frac{1}{N} \sum_{i=1}^N f_i, \quad (12)$$

$$f_{2i} = f_i - \text{median}(f_1, f_2, \dots, f_N), \quad (13)$$

$$f_{3i} = f_i - \text{max}(f_1, f_2, \dots, f_N), \quad (14)$$

$$f_{4i} = f_i - \min(f_1, f_2, \dots, f_N), \quad (15)$$

where the functions **median**($\cdot$), **max**($\cdot$) and **min**($\cdot$) computes the median, maximum and minimum value of the input feature vector. These statistical features capture the intergroup variance of one ellipse fitting compared with the whole group of fitting ellipses for the same object contour. These statistical features are calculated for each feature in Set 1 and Set 3.

Set 6: The following two features are also considered in the classification algorithm: 1) Distance from fitting ellipse centroid to the object centroid. 2) Average distance from one ellipse centroid to the centroids of all the other fitting ellipses for the same object contour.

2.2.3 Learning Based Ellipse Refinement Using Two-Stage Sparse

Optimization—In the object space, not all the training samples are representative ellipseyonuclei cases that need to or should be used for learning. Similarly, in the feature space, a discriminative feature set can significantly improve the classification performance considering the “curse of dimensionality”. We have proposed a novel two-stage sparse optimization to 1) select the most representative training samples in the object space, and 2) identify the most discriminative features in the feature space.

Training example selection using dictionary learning by K-selection. In this step, we assume that all the training samples can be sparsely represented by a subset of representative samples (dictionary bases). Only these typical samples are needed to train the ellipse refinement classifier. Our recently proposed K-selection [46] algorithm is used to select a set of most representative training samples from the original training data set. For the given training sample set $E = \{e_i | i = 1, 2, \dots, N\}$, where e_i denotes the feature vector of the i th training sample, we assume a dictionary $B = \{b_i | i = 1, 2, \dots, K\}$ exists such that all the training samples can be represented by a linear combination of the selected bases in B . The basis samples in B can be found by performing dictionary update and sparse coding alternatively:

$$\min_{b_k \in B, x_i} \sum_{i=1}^N \|e_i - \sum_{k=1}^K b_k x_{ik}\|^2 + \lambda \|d_i \odot x_i\|^2, \quad s.t. 1^T x_i = 1, \forall i, \quad (16)$$

where b_k is k th feature vector selected as the basis vector from the original training sample set, x_i is the representation coefficients, and d_i is the distance between e_i and the basis vectors.

The dictionary is updated using a projection-based gradient descent method. Within each iteration, we choose sample e_l to replace basis b_t .

$$e_l^* = \arg \max_{e_l} \frac{-\nabla f_t(e_l - b_t)^T}{\|\nabla f_t\|_2 \|e_l - b_t\|_2}, \quad (17)$$

where ∇f_t is the gradient of cost function (16) with respect to basis b_t . We accept e_l if the replacement reduces the value of the cost function, that is the selected basis is in the

direction that has the largest correlation to the negative gradient direction. The last term in (16) models the locality constraint, which encourages each sample to be represented by its local bases and allows the optimization problem to be efficiently solved in a much smaller local system. Locality-constraint linear coding [47] is applied to (16) for sparse coding. In our algorithm, we perform this K-selection to the positive and negative samples separately. K-selection does not assume that the data is low-rank. For each sample, it does not require all the samples to be used during the optimization procedure, either. These characteristics are different from the recent selection-based dictionary learning methods [48], [49].

Feature selection using sparse representation. As shown in Table 1, in total we have extracted $P = 43$ morphological features for learning based fitting ellipse refinement. In order to improve the classification performance, a sparse representation model is employed to select the most discriminative features:

$$\min_{\alpha, \beta} \sum_{i=1}^N w_i \log(1 + e^{-z_i(\alpha^T f_i + \beta)}), \quad s.t. \|\alpha\|_1 \leq \rho, \alpha \geq 0, \quad (18)$$

where N is the number of training ellipses, w_i and $f_i \in \mathbb{R}^{P \times 1}$ represent the weight and the feature for the i th ellipse, respectively. The binary scalar z_i denotes the ellipse label: $z_i = +1$ for positive and $z_i = -1$ for negative, β is a scalar representing the intercept, and ρ is the parameter controlling the sparsity of α . Due to the l_1 norm constraint, the solution $\alpha^* \in \mathbb{R}^{P \times 1}$ to (18) is sparse with nonzero elements corresponding to the selected discriminative features. Based on α^* with L nonzero elements, all the features are projected onto a lower-dimensional, discriminative feature space, which can benefit the subsequent classification. The equation (18) is optimized with the widely used Sparse Learning Package (SLEP) [50].

Equation (18) is a supervised method which aims to select discriminative features with respect to the subsequent classification as well as removing redundant information. In comparison with those unsupervised feature selection methods such as PCA [51], ISOMAP [52], and LLE [53], our feature selection method is designed to optimize the classification performance. Boosting [54] based feature selection may have performance similar to sparse representation due to their common logistic regression nature. However, boosting is relatively less efficient compared with sparse representation for feature selection [55]. SVM-RFE feature selection does not guarantee the selected features are most relevant with minimal redundancy [56]. Other methods, like Laplacian score [57], trace ratio [58], relief and relief [59], and SPEC [60], and HSIC [61], evaluate features individually and thus are not specifically designed to handle redundant features.

2.3 Classification and Inner Geodesic Distance Based Clustering

After two-stage sparse optimization, the selected training samples are used to train a SVM classifier. The classifier outputs are shown in Fig. 8. Note that the smaller markers are the centers of the ellipses selected by the classifier. In order to get the final position, we have designed an inner distance based mean-shift clustering algorithm to merge all the candidates. Our method consists of 1) inner-distance calculation to compute the inner-distance matrix, and 2) mean-shift based clustering to locate the final seeds.

The inner-distance is calculated as the length of the shortest connecting paths that only lay inside the object. For example, in Fig. 8g, the direct line connecting the two center points represents Euclidean distance because the paths are not entirely inside the object. On the contrary, the inner geodesic distance is calculated by first building a graph with the ellipse centers and the concave points as vertices and the links connecting these vertices inside the contour as edges, then running a shortest distance algorithm in the graph. An example of inner-distance graph is shown in Fig. 8h. Since the inputs of the mean shift clustering are coordinates of the points, the inner-distance matrix will be converted into a new coordinate system by harboring the origin at one of the two points and calculating the other relevant distances.

The reason that we do not choose standard Euclidean distance is demonstrated in Fig. 8. Inner distance is intuitively correct because the real distance between two seeds should be the paths within the object, instead of a direct line that might cross the cell boundaries [14]. As shown in Fig. 8a, Euclidean distance based clustering will face some serious challenges in selecting proper clustering bandwidth (Figs. 8b, 8f, and 8j show result of clustering with a smaller bandwidth and Figs. 8c, 8g, and 8k show results with a larger bandwidth). On the other hand, based on inner geodesic distance, correct clustering results have been obtained using one unified bandwidth, as shown in Figs. 8d, 8h, and 8l. Inner-distance is proven to be quite effective in natural shape classification [62]. Some clustering results are shown in Fig. 8. Note that all the clustering results based on inner geodesic distance are obtained with the same clustering bandwidth $bwdc = 14$.

III. EXPERIMENTAL RESULTS

All animal procedures were conducted in accordance with institutional guidelines for the care and use of laboratory animals as approved by the Animal Care and Use Committee at the University of Kentucky. Adult (four months of age) male C57Bl6 mice were injected intraperitoneally with ketamine (100 mg/kg) and xylazine (10 mg/kg) and euthanized by cervical dislocation. Plantaris muscles were fixed in situ at resting length by fixation in 4 percent paraformaldehyde for 48 hours. Single muscle fibers were subjected to 40 percent NaOH digestion and manually isolated as previously described [63]. Single fibers were stained with DAPI and nuclei from 15–25 fibers per animal counted. Fibers were imaged by z-stack analysis using the AxioImager M1 microscope.

In total we have collected over 500 multifocus z-stack images (over 1,500 myonuclei) from 42 individual isolated single muscle fibers. A set of image patches cropped from some randomly selected z-stack images and their corresponding all-in-focus fused image are shown in Fig. 9. A support vector machine classifier is trained using ellipses generated from seven fused images (1,000 positive samples and 3,000 negative samples), and the rest 35 cases are used for testing. The ground-truth ellipses and seeds are manually marked for each cell. In order to evaluate and validate the proposed algorithm, we have presented the experimental results on validating 1) the effectiveness of the geometric features described in Section 2.2.2 on ellipse refinement, 2) sparse representation based sample and feature selection methods described in Section 2.2.3, 3) the effectiveness of the inner-distance based clustering (2.3), and 4) the overall comparative myonuclei detection results.

We implemented our algorithm using MATLAB with the SVM package [64] on a PC machine with i3 CPU and 16 GB memory. The SVM classifier uses Gaussian kernel with precision $\gamma = 0.3$ and penalty 1 by cross validation. The bandwidth for mean-shift clustering is empirically set to 16 pixels. For the data with elongated or larger cells, a large window bandwidth is preferred. The selection of K in (16) interacts with the size of training samples. In practice a large K is chosen when the data have significant variations, and 15 percent of the training sample size is found to be sufficient to approximate the data. Usually the algorithm takes less than 1 minute to count one all-in-focus image.

The k in Equation (2) determines the sharpness of the fused image. A smaller k will generate a fused image with higher contrast and vice versa. In our experiment, $k = 3$ produces enough contrast. In order to avoid the local false concave points, the contours are smoothed by elliptical Fourier descriptor, and its first m coefficients are kept. $m = 20$ is chosen in our experiments.

3.1 Geometric Features

In this section we illustrate the effectiveness of the geometric features for ellipse refinement in Fig. 10. The first row of Fig. 10 is the original ellipse fitting results obtained from Section 2.2.1, and the results after the learning based ellipse fitting refinement are demonstrated in the fourth row. Set 1 evaluates the extent of matching between object boundaries and fitting ellipses. In Fig. 10a, it is clear that the classifier can successfully reject some false fitting ellipses that have lower matching ratios.

The features A_e and A_c in Set 2 help to differentiate the valid fitting ellipse from false fittings by measuring their areas. As we show in Fig. 10b, although the largest (longest) ellipse exhibits better overlapping ratios, it also encloses a larger area. In addition, the axis ratio r_{axis} and contour perimeter p_c also indicate that the largest ellipse is suboptimal. Considering these regularizations, the ellipse refinement classifier has rejected this candidate, as shown in Fig. 10b.

In Set 3, the concave point depth is designed for the touching myonuclei. We put one typical case of the touching myonuclei in Fig. 10c. The red circles in the third row are concave points. It can be observed that this feature enables the classifier to reject the two ellipses containing concave points. The effect of feature $decc$ is shown in Fig. 10d. The ellipses whose centers are close to the contour boundary are removed.

The contribution of the irregularity measurement is shown in Fig. 10e. This feature is designed to enable a more robust classifier in the presence of extremely irregular contours. Compared with the myonuclei with smooth contours, these irregular ones can generate higher values with respect to irregularity measurements of the fitting ellipses. This enables the classifier to differentiate the ellipses having lower overlapping ratios with higher irregularity measurements, which may still be valid ellipse fitting results.

The Set 5 features are designed to capture the statistical properties of fitting ellipses within one contour. The effect of the two distances features in Set 6 is shown in Fig. 10f. These

features force the classifier to select the ellipses whose centroids locate near to the centroid of the contour and away from the contour boundary.

The parameter $\rho = 180$ is used in the sparse representation-based feature selection model in (18), and a set of most discriminative features is presented in Fig. 11. Those features with zero coefficient values are non-discriminative ones, which contain the $f2i$ (except the median feature for *matchqual2*) and $f3i$ in Table 1. The mean features $f1i$ for the *raoc*, *raoe* and *roace* in Table 1 are not considered as important features either. On the contrary, the ranking scores *matchqual1*, *matchqual2*, *matchqual3*, and their corresponding min features $f4i$ are considered as the most significant features, since they iteratively evaluate the accuracy of a fitted ellipse based on the pixel-wise contour/fitting ellipse overlapping ratio. The selected features based on sparse representation are denoted with bold font in Table 1: Feature Index.

3.2 Sample and Feature Selection

3.2.1 Sparse Representation for Training Sample—Selection In our algorithm, the purpose for training sample selection is to learn a compact dictionary whose element is directly selected from the original training data set. This approach is different from the widely used bag-of-words (BoW) model [65], [66] and spatial pyramid matching (SPM) [67] method for object recognition. In BoW model, the elements (visual words) in the codebook are not selection based, and are often learned through some clustering methods such as Kmeans. In order to compare these two different dictionary learning approaches, these two methods were evaluated based on the classification accuracy in finding correct ellipse fitting candidates for myonuclei detection.

In the comparative experiment, 4,000 ellipse fitting candidates are randomly divided into training (1,600, 400 positive and 1,200 negative) and testing (2,400, 600 positive and 1,800 negative) data sets, respectively. 15 percent of the training samples were selected by Kselection and the same amount of visual words were calculated in BoW model for fair comparison. The same classifier (linear kernel SVM) using the same parameters are trained for evaluation. In order to avoid bias introduced in the feature selection step, all features are used to evaluate the two different dictionary learning methods. The ROC curve is shown in Fig. 14, and the areas under the ROC curves (AUCs) are also presented in the legend. As one can tell, sparse representation based sample selection provides better results than BoW for our application with an average $AUC = 0.85$ compared with $AUC = 0.84$ using BoW model.

3.2.2 Sparse Representation for Feature Selection—In order to justify the proposed sparse representation based feature selection method, our approach was compared with three state-of-the-arts, including PCA, Ada-Boost [54], and SVM-RFE [56]. PCA achieves linear dimension reduction by removing the principal components corresponding to smaller eigenvalues. In this way the original data vector can be represented in a lower dimension space. In AdaBoost, each weak classifier corresponds to one single image feature, and the final strong classifier not only represent the ensemble of all weak learners, but can also be used for feature selection by observing each weak learner's weight. SVM-RFE is an iterative method [68], [69]. In each iteration it trains a linear SVM classifier and ranks the features

with respect to their weights. The feature with the lowest weight is removed. The final output is a ranking of all the features in terms of their importance.

For fair comparison, the same set of features listed in Table 1 is used in these algorithms. Similar to Section 3.2.1, the performance is measured based on the classification accuracy using the 2,400 testing samples. The ROC curves are shown in Fig. 15. It can be observed that the proposed feature selection method outperforms the other three methods. The performance of AdaBoost is close to sparse feature selection method. SVM-RFE can provide satisfactory results but is still outperformed by the proposed method. PCA is an unsupervised method, and thus it is not optimized to provide best classification results. The AUCs of four methods are listed in Fig. 16.

In addition, the statistics (sensitivity, specificity, and accuracy) of the true positive rate (TPR) with the false positive rate being fixed at 0.2 are also presented in Table 2. All these detailed quantitative experiments confirm that sparse feature selection algorithm provides the best results classifying ellipse fitting candidates for robust myonuclei detection.

3.3 Inner-Distance Based Clustering

The performance of inner-distance based clustering is shown in Fig. 12. The first row denotes the classifier's initial output. The second row demonstrates the inner-distance based clustering results. Through the whole experiments, we set the clustering bandwidth to $bwdc = 16$. Some myonuclei detection results of nine fused image after inner-distance based clustering is shown in Fig. 13.

3.4 Comparative Experiments

The proposed algorithm is compared with three recent state-of-the-art cell detection methods: Laplacian-of-Gaussian filters [33], iterative radial voting (IRV) [30], and single-pass voting (SPV) [32]. Several representative image patches using different algorithm are shown in Fig. 17. For all the comparison experiments, we tried many possible parameters and configurations to generate the best results for the other methods used for comparison. As one can tell, it is difficult for LoG to handle touching cells and inhomogeneous intensity, while IRV and SPV may fail on cells with elongated ellipse shape, because they all assume round shape cells. Meanwhile, it is often difficult for IRV and SPV to create a general rule for parameter settings on an image containing cells with different sizes and shapes. However, the proposed algorithm is more robust with respect to touching cells exhibiting a large variations of scales, shapes, and especially inhomogeneous intensity distributions. This can be attributed to the robust ellipse fitting using HEIV regression, two stage sparse optimization model, and inner distance based mean-shift clustering in this proposed algorithm.

For quantitative comparison, we have measured the Euclidean distances between the manually annotated seeds and those created by the proposed automatic algorithm. Pixel-wise detection accuracy is listed in Table 3. It is obvious that the proposed method produces best accuracy with respect to both mean errors and standard deviations, representing high detection accuracy and robustness.

Table 4 shows the seed detection evaluation metrics including missing rate (MR), over-detection rate (OR), effective rate (ER), precision, recall, and $F1$ score. The MR or OR means no seeds or more than one seed are detected for one ground-truth cell, respectively. The ER is calculated by evaluating the ratio between the number of detected seeds and the ground truth seeds, which measures the robustness to background clutter with $ER = 1$ indicating the strongest robustness. Precision (P), recall (R) and $F1$ score are defined as $P = \frac{TP}{TP+FP}$, $R = \frac{TP}{TP+FN}$ and $F1 = \frac{2*P*R}{P+R}$. True positive (TP) is defined as a detected seed that is inside a circular region with a radius $r = 8$ pixels centered at the ground truth seed. If it is outside this circular region, the detection is treated as false positives (FP). False negatives (FN) are defined as the missing detections.

As we show in Table 4, LoG is sensitive to image background noise, SPV may have higher FN , and IRV gives relatively low P and R values. On the contrary, the proposed method produces the highest P value without sacrificing the accuracy of R . For fair comparisons, the MR and OR are excluded when we evaluate the pixel-wise accuracy in Table 3.

IV. CONCLUSION

We presented a touching myonuclei detection approach. Our approach tackles the challenges in splitting clusters consisting of myonuclei with diverse shapes and sizes using a robust ellipse fitting based on heteroscedastic errors-in-variables regression, sample and feature selection by two-stage sparse optimization model, and learning based refinement and mean-shift clustering using geodesic inner-distance. The proposed algorithm is resilient to the variations in nuclear shapes and sizes. The experimental results demonstrate that this algorithm outperforms many existing state-of-the-art approaches. Previously, it is very common for a biology lab to analyze hundreds of isolated fibers from various muscle perturbations, requiring an incredible amount of time- and man-power for accurate quantification. This automatic approach is an exciting endeavor that can efficiently and accurately quantify the myofiber adaptation.

References

1. McCarthy JJ, Mula J, Miyazaki M, Erfani R, Garrison K, Farooqui AB, Srikuea R, Lawson BA, Grimes B, Keller C, van Zant G, Campbell KS, Esser KA, Dupont-Versteegden EE, Peterson CA. Effective Fiber Hypertrophy in Satellite Cell-Depleted Skeletal Muscle. *Development*. 2011; 138(17):3657–3666. [PubMed: 21828094]
2. Jackson JR, Mula J, Kirby TJ, Fry CS, Lee JD, Ubele MF, Campbell KS, McCarthy JJ, Peterson CA, Dupont-Versteegden EE. Satellite Cell Depletion Does Not Inhibit Adult Skeletal Muscle Regrowth Following Unloading-Induced Atrophy. *Am J Physiology - Cell Physiology*. 2012; 303(8):C854–C861. [PubMed: 22895262]
3. Malpica N, de Solorzano CO, Vaquero JJ, Santos A, Vallcorba I, Garcia-Sagredo JM, del Pozo F. Applying Watershed Algorithms to the Segmentation of Clustered Nuclei. *J Cytometry*. Aug; 1997 28(4):289–297.
4. Lin G, Adiga U, Olson K, Guzowski JF, Barnes CA, Roysam B. A Hybrid 3D Watershed Algorithm Incorporating Gradient Cues and Object Models for Automatic Segmentation of Nuclei in Confocal Image Stacks. *J Cytometry*. 2003; 56A(1):23–36.
5. Ancin H, Roysam B, Dufresne T, Chestnut M, Ridder G, Szarowski D, Turner J. Advances in Automated 3-D Image Analysis of Cell Populations Imaged by Confocal Microscopy. *J Cytometry*. 1996; 25(3):22–234.

6. Vincent L, Soille P. Watersheds in Digital Spaces: C on Immersion Simulations. *IEEE Trans Pattern Analysis and Machine Intelligence*. Jun; 1991 13(6):583–598.
7. Lin G, Chawla MK, Olson K, Barnes CA, Guzowski JF, Bjornsson C, Shain W, Roysam B. A Multi-Model Approach to Simultaneous Segmentation and Classification of Heterogeneous Populations of Cell Nuclei in 3D Confocal Microscope Images. *Cytometry*. 2007; 71A(9):724–736. [PubMed: 17654650]
8. Lin G, Adiga U, Olson K, Guzowski JF, Barnes CA, Roysam B. A Hybrid 3D Watershed Algorithm Incorporating Gradient Cues and Object Models for Automatic Segmentation of Nuclei in Confocal Image Stacks. *Cytometry*. 2003; 56A(1):23–36. [PubMed: 14566936]
9. Yu W, Lee HK, Hariharan S, Bu W, Ahmed S. Quantitative Neurite Outgrowth Measurement Based on Image Segmentation with Topological Dependence. *Cytometry*. 2009; 75A(4):289–297. [PubMed: 18951464]
10. Mao K, Zhao P, Tan P. Supervised Learning-Based Cell Image Segmentation for P53 Immunohistochemistry. *IEEE Trans Biomedical Eng*. Jun; 2006 53(6):1153–1163.
11. Yan P, Zhou X, Shah M, Wong S. Automatic Segmentation of High-Throughput RNAi Fluorescent Cellular Images. *IEEE Trans Information Technology in Biomedicine*. Jan; 2008 12(1):109–117. [PubMed: 18270043]
12. Grau V, Mewes AUJ, Alcaniz M, Kikinis R, Warfield S. Improved Watershed Transform for Medical Image Segmentation Using Prior Information. *IEEE Trans Medical Imaging*. Apr; 2004 23(4):447–458. [PubMed: 15084070]
13. Yang X, Li H, Zhou X. Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in Time-Lapse Microscopy. *IEEE Trans Circuits and Systems I: Regular Papers*. Nov; 2006 53(11):2405–2414.
14. Yang L, Tuzel O, Meer P, Foran D. Automatic Image Analysis of Histopathology Specimens Using Concave Vertex Graph. *Proc Int'l Conf Medical Image Computing and Computer-Assisted Intervention (MICCAI'08)*. 2008:833–841.
15. Schmitt O, Hasse M. Radial Symmetries Based Decomposition of Cell Clusters in Binary and Gray Level Images. *J Pattern Recognition*. Jun; 2008 41(6):1905–1923.
16. Jung C, Kim C. Segmenting Clustered Nuclei Using HMinima Transform-Based Marker Extraction and Contour Parameterization. *IEEE Trans Biomedical Eng*. Oct; 2010 57(10):2600–2604.
17. Cheng J, Rajapakse J. Segmentation of Clustered Nuclei with Shape Markers and Marking Function. *IEEE Trans Biomedical Eng*. Mar; 2009 56(3):741–748.
18. Kolmogorov V, Zabih R. What Energy Functions Can Be Minimized via Graph Cuts? *IEEE Trans Pattern Analysis and Machine Intelligence*. Feb; 2004 26(2):147–159.
19. Boykov Y, Funka-Lea G. Graph Cuts and Efficient N-D Image Segmentation. *Int'l J Computer Vision*. Nov; 2006 70(2):109–131.
20. Bernardis E, Yu S. Finding Dots: Segmentation as Popping Out Regions from Boundaries. *Proc IEEE Conf Computer Vision and Pattern Recognition (CVPR)*. Jun.2010 :199–206.
21. Shi J, Malik J. Normalized Cuts and Image Segmentation. *IEEE Trans Pattern Analysis and Machine Intelligence*. Aug; 2000 22(8):888–905.
22. Wen Q, Chang H, Parvin B. A Delaunay Triangulation Approach for Segmenting Clumps of Nuclei. *Proc IEEE Int'l Symp Biomedical Imaging: From Nano to Macro (ISBI)*. 2009:9–12.
23. Nath SK, Palaniappan K, Bunyak F. Cell Segmentation Using Coupled Level Sets and Graph-Vertex Coloring. *Proc Int'l Conf Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2006:101–108.
24. Faustino GM, Gattass M, Rehen S, de Lucena C. Automatic Embryonic Stem Cells Detection and Counting Method in Fluorescence Microscopy Images. *Proc IEEE Int'l Symp Biomedical Imaging: From Nano to Macro (ISBI)*. 2009:799–802.
25. Chen C, Li H, Zhou X, Wong S. Constraint Factor Graph Cut-Based Active Contour Method for Automated Cellular Image Segmentation in RNAi Screening. *J Microscopy*. 2008; 230(2):177–191.

26. Wu Z, Gurari D, Wong JY, Betke M. Hierarchical Partial Matching and Segmentation of Interacting Cells. *Proc Int'l Conf Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2012:389–396.
27. Yu W, Lee HK, Hariharan S, Bu W, Ahmed S. Evolving Generalized Voronoi Diagrams for Accurate Cellular Image Segmentation. *Cytometry Part A*. 2010; 77(4):379–386.
28. Janowczyk A, Chandran S, Singh R, Sasaroli D, Coukos G, Feldman MD, Madabhushi A. High-Throughput Biomarker Segmentation on Ovarian Cancer Tissue Microarrays via Hierarchical Normalized Cuts. *IEEE Trans Biomedical Eng*. May; 2012 59(5):1240–1252.
29. Sertel O, Dogdas B, Chiu CS, Gurcan MN. Microscopic Image Analysis for Quantitative Characterization of Muscle Fiber Type Composition. *J Computerized Medical Imaging and Graphics*. 2011; 35:616–628.
30. Parvin B, Yang Q, Han J, Chang H, Rydberg B, Barcellos-Hoff MH. Iterative Voting for Inference of Structural Saliency and Characterization of Subcellular Events. *IEEE Trans Image Processing*. Mar; 2007 16(3):615–623.
31. Hafiane A, Bunyak F, Palaniappan K. Fuzzy Clustering and Active Contours for Histopathology Image Segmentation and Nuclei Detection. *Advanced Concepts for Intelligent Vision Systems*. 2008; 5259:903–914.
32. Qi X, Xing F, Foran D, Yang L. Robust Segmentation of Overlapping Cells in Histopathology Specimens Using Parallel Seed Detection and Repulsive Level Set. *IEEE Trans Biomedical Eng*. Mar; 2012 59(3):754–765.
33. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images. *IEEE Trans Biomedical Eng*. Apr; 2010 57(4):841–852.
34. Akakin HC, Kong H, Elkins C, Hemminger J, Miller B, Ming J, Plocharczyk E, Roth R, Weinberg M, Ziegler R, Lozanski G, Gurcan MN. Automated Detection of Cells from Immunohistochemically-Stained Tissues: Application to Ki-67 Nuclei Staining. *Proc SPIE*. Feb. 2012 8315:article 831503.
35. Kong H, Gurcan M, Belkacem-Boussaid K. Partitioning Histopathological Images: An Integrated Framework for Supervised Color-Texture Segmentation and Cell Splitting. *IEEE Trans Medical Imaging*. Sep; 2011 30(9):1661–1677. [PubMed: 21486712]
36. Li G, Liu T, Nie J, Guo L, Malicki J, Mara A, Holley SA, Xia W, Wong ST. Detection of Blob Objects in Microscopic Zebrafish Images Based on Gradient Vector Diffusion. *Cytometry Part A*. 2007; 71(10):835–845.
37. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Systems, Man and Cybernetics*. Jan; 1979 9(1):62–66.
38. Hariharan H, Koschan A, Abidi M. Multifocus Image Fusion by Establishing Focal Connectivity. *Proc IEEE Int'l Conf Image Processing (ICIP)*. 2007; 3:III-321–III-324.
39. Kuhl FP, Giardina CR. Elliptic Fourier Features of a Closed Contour. *J Computer Graphics and Image Processing*. 1982; 18(3):236–258.
40. Kuhl FP. Elliptical Fourier Descriptor of a Closed Contour. *Computer Graphics and Image Processing*. 1982; 18:236–258.
41. Fitzgibbon AW, Pilu M, Fisher RB. Direct Least-Squares Fitting of Ellipses. *IEEE Trans Pattern Analysis and Machine Intelligence*. May; 1999 21(5):476–480.
42. Zhang Z. Parameter Estimation Techniques: A Tutorial With Application to Conic Fitting. *Image and Vision Computing*. 1997; 15(1):59–76.
43. Matei B, Meer P. Reduction of Bias in Maximum Likelihood Ellipse Fitting. *Proc Int'l Conf Pattern Recognition*. 2000; 3:794–798.
44. Taubin G. Estimation of Planar Curves, Surfaces, and Nonplanar Space Curves Defined by Implicit Equations with Applications to Edge and Range Image Segmentation. *IEEE Trans Pattern Analysis and Machine Intelligence*. Nov; 1991 13(11):1115–1138.
45. Matei B, Meer P. A General Method for Errors-in-Variables Problems in Computer Vision. *Proc IEEE Conf Computer Vision and Pattern Recognition (CVPR)*. 2000; 2:18–25.
46. Liu B, Huang J, Yang L, Kulikowsk C. Robust Tracking Using Local Sparse Appearance Model and K-Selection. *Proc IEEE Conf Computer Vision and Pattern Recognition (CVPR)*. 2011:1313–1320.

47. Wang J, Yang J, Yu K, Lv F, Huang T, Gong Y. Locality-Constrained Linear Coding for Image Classification. *Proc IEEE Conf Computer Vision and Pattern Recognition (CVPR)*. 2010:3360–3367.
48. Bien J, Xu Y, Mahoney MW. Robust Subspace Segmentation by Low-Rank Representation. *Proc Advances in Neural Information Processing Systems (NIPS)*. 2010:217–225.
49. Liu G, Lin Z, Yu Y. Robust Subspace Segmentation by Low-Rank Representation. *Proc Int'l Conf Machine Learning (ICML)*. 2010:663–670.
50. Liu, J.; Ji, S.; Ye, J. SLEP: Sparse Learning with Efficient Projections. Arizona State Univ; 2009. <http://www.public.asu.edu/~jye02/Software/SLEP>
51. Turk M, Pentland A. Eigenfaces for Recognition. *J Cognitive Neuroscience*. 1991; 3(1):71–86.
52. Tenenbaum JB, de Silva V, Langford JC. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*. 2000; 290(5500):2319–2323. [PubMed: 11125149]
53. Roweis ST, Saul LK. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*. 2000; 290(5500):2323–2326. [PubMed: 11125150]
54. Friedman J, Hastie T, Tibshirani R. Additive Logistic Regression: A Statistical View of Boosting (with Discussion and a Rejoinder by the Authors). *The Annals of Statistics*. 2000; 28(2):337–407.
55. Liang Y, Wang L, Liao S, Zou B. Feature Selection via Simultaneous Sparse Approximation for Person Specific Face Verification. *Proc IEEE Int'l Conf Image Processing (ICIP)*. 2011:789–792.
56. Mundra PA, Rajapakse JC. SVM-RFE with Relevancy and Redundancy Criteria for Gene Selection. *Proc Second IAPR Int'l Conf Pattern Recognition in Bioinformatics*. 2007:242–252.
57. He X, Cai D, Niyogi P. Laplacian Score for Feature Selection. *Proc Advances in Neural Information Processing Systems*. 2006; 18:507.
58. Nie F, Xiang S, Jia Y, Zhang C, Yan S. Trace Ratio Criterion for Feature Selection. *Proc 23rd Nat'l Conf Artificial Intelligence*. 2008; 2:671–676.
59. Robnik-Sikonja M, Kononenko I. Theoretical and Empirical Analysis of Relieff and Rrelieff. *Machine Learning*. 2003; 53(1/2):23–69.
60. Zhao Z, Liu H. Spectral Feature Selection for Supervised and Unsupervised Learning. *Proc Int'l Conf Machine Learning (ICML)*. 2007:1151–1157.
61. Song L, Smola A, Gretton A, Borgwardt KM, Bedo J. Supervised Feature Selection via Dependence Estimation. *Proc 24th Int'l Conf Machine Learning*. 2007:823–830.
62. Ling H, Jacobs D. Shape Classification Using the Inner-Distance. *IEEE Trans Pattern Analysis and Machine Intelligence*. Feb; 2007 29(2):286–299.
63. Brack A, Bildsoe H, Hughes S. Evidence that Satellite Cell Decrement Contributes to Preferential Decline in Nuclear Number from Large Fibres during Murine Age-Related Muscle Atrophy. *J Cell Science*. 2005; 118:4813–4821. [PubMed: 16219688]
64. Joachims, T. Making Large-Scale SVM Learning Practical. In: Scholkopf, B.; Burges, C.; Smola, A., editors. *Advances in Kernel Methods - Support Vector Learning*. MIT Press; 1999.
65. Fei-Fei L, Perona P. A Bayesian Hierarchical Model for Learning Natural Scene Categories. *Proc IEEE Conf Computer Vision and Pattern Recognition (CVPR)*. 2005; 2:524–531.
66. Leibe B, Leonardis A, Schiele B. Combined Object Categorization and Segmentation with an Implicit Shape Model. *Proc European Conf Computer Vision Workshop Statistical Learning in Computer Vision (ECCV'04)*. 2004:17–32.
67. Lazebnik S, Schmid C, Ponce J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. *Proc IEEE Conf Computer Vision and Pattern Recognition (CVPR)*. 2006; 2:2169–2178.
68. Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification Using Support Vector Machines. *Machine learning*. 2002; 46(1–3):389–422.
69. Duan KB, Rajapakse JC, Wang H, Azuaje F. Multiple SVM-RFE for Gene Selection in Cancer Classification with Expression Data. *IEEE Trans NanoBioscience*. Sep; 2005 4(3):228–234. [PubMed: 16220686]

Biographies

Hai Su received the BE and MS degrees from the University of Electronic Science and Technology of China, in 2003 and 2006, respectively. He is currently working toward the PhD degree in the Department of Computer Science, University of Kentucky. His research interests include imaging informatics and biomedical image analysis.

Fuyong Xing received the BS and MS degrees from the Department of Information and Communication Engineering at Xian Jiaotong University, China, in 2006 and 2009, respectively. He is currently working toward the PhD degree in the Department of Computer Science, University of Kentucky. His current research interests include medical image analysis, machine learning, and computer vision.

Jonah D. Lee received the BS degree from the University of Oregon and the PhD degree in the Department of Physical Education, Sport, and Exercise Science from Ball State University, in 2004 and 2010, respectively. He is currently a postdoctoral fellow with the Department of Rehabilitation Sciences, College of Health Sciences and the Department of Physiology, College of Medicine, University of Kentucky. His research interests include the

regulation of aged-skeletal muscle plasticity and the clinical, preventative and performance implications of muscle adaptation across the lifespan.

Charlotte A. Perterson received the BS degree from the University of Notre Dame and the PhD degree from the University of Virginia, followed by two postdoctoral fellowships, the first at the National Eye Institute in the National Institutes of Health (NIH) and the second at Stanford University School of Medicine. She is a Joseph Hamburg Endowed professor and the associate dean for research at the College of Health Science, University of Kentucky. Her research interests include elucidation of cellular and molecular mechanisms controlling skeletal muscle structure and function. She is currently funded by the NIH to study the role of adult stem cells in muscle maintenance and growth, mechanisms underlying muscle fatigue, the influence of obesity on muscle function, and inflammation in muscle response to exercise. Her work emphasizes on changes that occur with age with the long- term goal of preventing frailty and loss of functional independence.

Lin Yang received the BE and MS degrees from Xian Jiaotong University, in 1999 and 2002, respectively, and the PhD degree in the Department of Electrical and Computer Engineering from Rutgers, the State University of New Jersey in 2009. He is an assistant professor with the Division of Biomedical Informatics at the Department of Biostatistics, and the Department of Computer Science at the University of Kentucky. He was an assistant professor in the Department of Radiology in the University of Medicine and Dentistry of New Jersey and a graduate faculty in the Department of Biomedical Engineering in Rutgers University, from 2009 to 2011. His major research interests include biomedical image analysis, computer vision, machine learning, and imaging informatics. He is also working on high performance computing and computer-aided diagnostics.

Fig. 1.

The original z-stack images are shown on the left and the all-in-focus image obtained using the multi-focus image fusion is shown on the right.

Fig. 2.
An overview of the proposed algorithm.

Fig. 3.

(a) Original contour. (b) Concave point detection. The concave points with angle γ is removed because $\gamma > 165^\circ$. The concave points with angle β is removed because the midpoint of edge c' is inside the contour. The contour point with angel α is as the true concave points.

Fig. 4.
A demonstration of various ellipse fitting results.

Fig. 5.

Robust ellipse fitting using HEIV regression and an example to illustrate the Set 3 features.

(a) Three concave points are marked with 1, 2, and 3. (b) Three ellipses are generated by robust ellipse fitting using the contour segments s_{ij} , where $i, j = 1, 2, 3$, $i \neq j$ denote the indices of the concave points. (c) and (d) Three fitting ellipses using two out of three contour segments. (e) The ellipse generated using all three segments separated by three concave points.

Fig. 6.

Examples to illustrate the Set 3 features. (a) and (b) The fitting ellipses whose centers are close to the cell boundary, *i.e.*, d_{ecc} is small, and (c) and (d) represent good ellipse fittings with large d_{ecc} .

Fig. 7.

Examples to illustrate the Set 4 features (irregularity of the object boundaries). Both cells shown above are single cells. (b) is the ellipse fitting result of (a) that has high r_{irg} . (d) is the ellipse fitting result of (c) that has lower r_{irg} .

Fig. 8.

The demonstration of the effectiveness of mean-shift clustering based on inner distance. The ground-truth myonuclei detection should identify two touching nuclei on the top and one single nuclei in the bottom (in total three myonuclei). This is a challenging problem because of the bright dots irregularly distributed inside the my nuclei. This is due to the DAPI intercalating preferentially into heterochromatin. The red dots in (b), (c), and (d) denote the clustering results. The yellow lines in (f) and (g) denote a sample Euclidean distance between two seeds. The yellow line in (h) denotes a sample inner distance between two seeds. (a) The initial fitting ellipses' centers generated by the classifier. (b) The clustering results based on Euclidean distance with bandwidth $bw_{dc} = 10$. (c) The clustering based on Euclidean distance with bandwidth $bw_{dc} = 14$. As presented in (b) and (c), it is difficult to find a unified bandwidth that can produce accurate detection results for both cases. However, using inner distance, a unified bandwidth ($bw_{dc} = 14$) can be used and the accurate detection results for both cases are shown in (d).

Fig. 9.

(a)–(m) The original z-stack images. (n) The final all-in-focus image after multi-focus image fusion.

Fig. 10.

The first row is the initial ellipse fitting results. The second row are the contours. The third row represents the contours and the fitted ellipses. The fourth row lists the ellipses selected by the classifier. (a) The effect of overlapping ratios: Ellipses that poorly match a myonucleus are removed. (b) The effects of the axis ratio, perimeter and areas of the ellipses and myonuclei contours: Many suboptimal fitting candidates are removed. (c) The effect of concave point depth: Fitting ellipses that contain deep inside concave points are removed. (d) The effect of the distance from the ellipse center to the contour: Ellipses whose centers are near to the contour boundaries are removed. (e) The effect of the contour irregularity: Many suboptimal fitting candidates are removed. (f) The effect of the distance from ellipse centroids to the contour centroid and the inter distance between ellipse centroids: The fitting results whose centers are close to the object boundaries are removed.

Fig. 11.

The selected discriminative geometric features using sparse representation. For the meaning of the ellipse features corresponding to the feature index, please refer to Table 1.

Fig. 12.

The mean-shift clustering based on inner distance. The first row represents the original output of the classifier, each red dot denotes the center of one fitting ellipse. The second row represents the final meanshift clustering results based on geometric inner distance. The ground truth annotation for the myonucleus detection should be two touching myonuclei for column 1, one myonuclei for column 2, and two touching myonuclei for column 3. It is obvious that in the second row, the inner distance based clustering can produce accurate results for all three cases.

Fig. 13.
The automatic myonuclei detection results of nine randomly selected image patches.

Fig. 14.

The ROC curves using the sparse dictionary learned with Kselection and the BOW model, respectively. 15 percent of the training data is selected as K in K-selection, and as the number of visual words in BOW. A SVM classifier is learned and tested on 2,400 testing samples including 600 positive samples and 1,800 negative samples.

Fig. 15.

The comparison of four feature selection methods in terms of classification performance on a testing dataset used in Fig. 14.

Fig. 16. The average comparative AUCs using sparse representation, AdaBoost, SVM-RFE, and PCA.

Fig. 17.

The comparative seed detection results using three randomly cropped image patches. Row 1 denotes the original images, and rows 2, 3, 4, and 5 represent the detection results created by LoG [33], IRV [30], SPV [32], and the proposed algorithm, respectively.

TABLE 1

A Summary of the Geometric Features for Learning Based Fitting Refinement

Sets	Symbols	Feature Index in Figure 11	Definitions
1	r_{aoc}	4	The ratio of the overlapping area with respect to the contour area
	r_{aoe}	5	The ratio of the overlapping area with respect to the ellipse area
	r_{oaoc}	6	The ratio of the overlapping area with respect to the sum of the ellipse area and the contour area.
	$match_{qual1}$	8	The match-quality rank in iteration 1
	$match_{qual2}$	9	The match-quality rank in iteration 2
	$match_{qual3}$	10	The match-quality rank in iteration 3
2	A_e	2	The area of an ellipse.
	r_{axis}	3	The ratio of the major and minor axis of an ellipse.
	A_c	11	The area of a myonuclei contour.
	p_c	7	The perimeter of a myonuclei contour.
3	d_i	1	The concave point depth of the i -th ellipse generated for one contour.
	d_{ecc}	42	The distance from the center of an ellipse to its contour.
4	r_{irg}	12	The measure of the irregularity (smoothness) of the contour.
5	f_{1i}	13–16, 17–19	The group of features based on the features in Set 1–2 and feature d_i in Set 3. For a specific feature f_i , ($i = 1, 4, 5, 6, 8, 9, 10$) and a group of ellipses generated from a contour, calculate f_i for each ellipse by subtracting the mean of f_i of all the ellipses in this group from f_i
	f_{2i}	20–24, 25, 26	Similar to f_{1i} , for an ellipse and the group of ellipse generated from the same contour, subtract the median of f_i , ($i = 1, 4, 5, 6, 8, 9, 10$) of all the ellipses in this group from f_i of this ellipse.
	f_{3i}	34–40	Similar to f_{1i} , subtract the maximum value of the f_i , ($i = 1, 4, 5, 6, 8, 9, 10$) of all the ellipses in the same group.
	f_{4i}	27–33	Similar to f_{1i} , subtract the maximum value of the f_i , ($i = 1, 4, 5, 6, 8, 9, 10$) of all the ellipses in the same group.
6	d_{cc}	41	The distance from ellipse centroid to its contour centroid.
	d_{acc}	43	The average distance from one ellipse centroid to the centroids of the other ellipses of the same contour.

TABLE 2

The Comparison of Classification Performance of Four Feature Selection Methods

Methods	TPR	Sensitivity	Specificity	Accuracy
AdaBoost	0.76	0.81	0.82	0.82
SVM-RFE	0.74	0.80	0.82	0.82
PCA	0.72	0.79	0.80	0.80
Proposed	0.81	0.81	0.83	0.83

TABLE 3

The Pixel-wise Seed Detection Accuracy Compared with Ground Truth

	Mean	Standard Devision	Min	Max
LoG [33]	3.72	4.22	0.55	7.99
IRV [30]	4.28	3.86	0.24	7.93
SPV [32]	2.99	3.54	0	7.96
Proposed	2.6	2.90	0.17	7.93

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 4

The Comparative Performance Metric

	MR	OR	ER	P	R	F1
LoG [33]	0.27	0.003	1.73	0.86	0.94	0.90
IRV [30]	0.31	0.34	1.69	0.83	0.90	0.86
iSPV [32]	0.21	0	0.97	0.89	0.94	0.91
Proposed	0.12	0	1.02	0.93	0.96	0.94